EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2010, 22 (3), 391–421

Psychology Press
Taylor & Francis Group

Complexity of categorical syllogisms: An integration of two metrics

Tracey A. Zielinski
University of Queensland, Brisbane, St. Lucia, Australia

Geoffrey P. Goodwin
University of Pennsylvania, Philadelphia, PA, USA

Graeme S. Halford
Griffith University, Nathan, Australia

The complexity of categorical syllogisms was assessed using the relational complexity metric, which is based on the number of entities that are related in a single cognitive representation. This was compared with number of mental models in an experiment in which adult participants solved all 64 syllogisms. Both metrics accounted for similarly large proportions of the variance, showing that complexity depends on the number of categories that are related in a representation of the combined premises, whether represented in multiple mental models, or by a single model. This obviates the difficulty with mental models theory due to equivocal evidence for construction of more than one mental model. The "no valid conclusion" response was used for complex syllogisms that had valid conclusions. The results are interpreted as showing that the relational complexity metric can be applied to syllogistic reasoning, and can be integrated with mental models theory, which together account for a wide range of cognitive performances.

Keywords: Relational complexity; Mental models; Categorical syllogisms; Deductive reasoning; Reasoning strategies.

The purpose of this research was to assess how the relational complexity (RC) metric accounts for performance in syllogistic inference, and to attempt to integrate it with mental models theory (Johnson-Laird & Byrne, 1991). Relational complexity (RC; Andrews & Halford, 2002; Halford, Cowan, & Andrews, 2007; Halford, Wilson, & Phillips, 1998) is defined by the number of related variables in a cognitive representation. Higher RC values are

Correspondence should be addressed to Graeme Halford, School of Psychology, Griffith University, Mount Gravatt Campus, Nathan 4111, Australia. E-mail: g.halford@griffith.edu.au

http://www.psypress.com/ecp DOI: 10.1080/09541440902830509

associated with increased processing loads, and adults are limited to four variables (Halford, Baker, McCredden, & Bain, 2005). Concepts too complex to be processed in parallel are handled by *segmentation* (decomposition into smaller segments that are processed serially) and *conceptual chunking* (recoding representations into fewer variables).

RC has accounted for complexity effects in children's reasoning (Andrews & Halford, 2002; Andrews, Halford, Bunch, Bowden, & Jones, 2003; Halford & Andrews, 2004, 2006; Halford, Andrews, Dalton, Boag, & Zielinski, 2002) including the foundational principle of transitivity, and the longstanding enigma of children's difficulty with class inclusion (Halford, Andrews, & Jensen, 2002; Inhelder & Piaget, 1964). It has also been applied to sentence comprehension (Andrews, Birney, & Halford, 2006), the Tower of Hanoi (Halford et al., 1998), and knights and knaves problems (Birney & Halford, 2002). It has been linked to working memory (Halford et al., 2007; Oberauer, Sub, Wilhelm, & Wittmann, 2008) and it has been applied to mathematics education (English & Halford, 1995) and to air traffic control (Boag, Neal, Loft, & Halford, 2006).

We will focus specifically on categorical syllogisms, which are deductive arguments containing two premises and a conclusion, for example:

Some teams in the World Cup are African.
No African teams have ever won the World Cup.
∴ Some teams in the World Cup have never won it.

Mental models (MM) theory is the metric that accounts for the most variance in syllogistic reasoning (Espino, Santamaria, & Garcia-Madruga, 2000; Evans, Newstead, & Byrne, 1993; Johnson-Laird & Byrne, 1991). Problems that require only one mental model are reliably easier than those that require two or three models (Bucciarelli & Johnson-Laird, 1999; Johnson-Laird & Bara, 1984; Johnson-Laird & Byrne, 1991). However it is unclear whether people actually construct more than one model. Whereas Bucciarelli and Johnson-Laird (1999) found multiple models were constructed under certain facilitative conditions, Newstead, Handley, and Buck (1999) found that individuals constructed no more models for syllogisms that theoretically require multiple models than they do for single model syllogisms, nor was there any correlation between number of models constructed and overall accuracy. Polk and Newell (1995) and Markovits and Barrouillet (2002) proposed a theory based on progressive elaboration of a single representation. Although propensity to construct alternative models may be related to individual differences, and to various task variables (Bucciarelli & Johnson-Laird, 1999; Newstead, Thompson, & Handley, 2002), some uncertainty remains about number of mental models as a complexity metric.

RC theory is essentially an MM theory and it has the essential properties of mental models. The most important assumptions in this context are that a mental model is an analogue of the structure implied by the premises (it consists of an array of entities rather than an abstract rule), that it is constructed by the reasoner, and can be used to generate an inference (Goodwin & Johnson-Laird, 2005; Halford, 1993). We also assume, consistent with MM theory, that the premises will be encoded first, then an integrated representation of the premises will be formed. This "forward processing" hypothesis appears reasonably robust (Stupple & Ball, 2007). RC reflects the load entailed in processing the relations implied by the premises, and RC theory would not be applicable to reasoning based on soundness (logical validity and truth of premises), where conclusions are retrieved from semantic memory. RC theory assesses complexity by the number of distinct categories that are implied by the premises of a syllogism, and differs from the MM theory of Johnson-Laird and Byrne (1991) in that it does not assume construction of more than one mental model. If successful, this approach would obviate the uncertainty of the MM metric and the resulting integration of MM and RC theory would increase parsimony in the field. In assuming that participants attempt to construct a model that represents relations implied by the premises, and integrate the representations to determine the logically correct conclusion, mental models theory is quite distinct from the probability heuristic model (PHM; Oaksford & Chater, 2007). RC theory also differs from PHM in that the latter does not incorporate a complexity metric per se, and it needs to be supplemented by relational complexity, or a similar metric, in order to account for confidence in conclusions (Halford, 2009). Ultimately there may be scope for further integration there.

In categorical syllogisms both premises and conclusion are in one of four "moods":

All X are Y	(A: universal affirmative premise)
Some X are Y	(I: particular affirmative premise)
No X are Y	(E: universal negative premise)
Some X are not Y	(O: particular negative premise)

The premises follow one of the four figural arrangements noted in Table 1. Hence, there are 16 different versions for each figure (four moods for each premise), which leads to 64 syllogisms. Of these, 27 have valid conclusions. Whether individuals generate their own conclusions or evaluate given conclusions, performance is generally above chance, but it is far from perfect and there is large variance between syllogisms and between individuals (for a review, see Polk & Newell, 1995). We will assess the ability of the RC metric to account for between-syllogisms variance.

TABLE 1
Conclusions according to the figural heuristic

Figure	*Premise arrangement*	*Conclusion form suggested by heuristic*
1	XY–YZ	X-Z
2	YX–ZY	Z-X
3	XY–ZY	n/a
4	YX–YZ	n/a

RELATIONAL COMPLEXITY OF CATEGORICAL SYLLOGISMS

The essence of RC as applied to categorical syllogisms is illustrated in Figure 1. Figure 1A displays the eight distinct classes of entity that arise given the interrelation of three separate properties. Relational complexity is defined by the number of these classes that are required to represent the combined premises, taking account of strategies that participants use. Figure 1B illustrates how a subset of these classes is explicitly related for the syllogism:

Premises: *All X are Y*
All Y are Z
Conclusions: *All X are Z, and Some Z are X.*

Representations of the separate premises are shown in lines 1 and 2 of Figure 1B, and of the combined premises in line 3. The full representation of the combined premises entails three classes of entities: XYZ, $\neg XYZ$, $\neg X \neg YZ$ (where $\neg$ means negation). The full representation of the combined premises is ternary relational. However, chunking can be applied because the relation between $\neg XYZ$ and $\neg X \neg YZ$ does not need to be processed, and so these two categories can be chunked (fused) into one category, by Principle 1, defined later. This is shown in line 4 of Figure 1B, which represents a relation between XYZ and $\neg$X(Y, $\neg$Y)Z. Therefore effective RC is 2. This application of conceptual chunking is known as the *principle of neglect* (in this case the status of Y can be neglected). A more general account of the principles of RC analyses is given below. Chunking is part of the encoding process, and in our complexity assessments we have not found it necessary to assume that chunking imposes a processing load.

Effective RC estimates are based on the strategies used, including heuristics, which are determined by rational analysis, supported by empirical evidence (Halford et al., 2007). Previous applications have been predominantly to domains such as transitive inference, sentence comprehension, and knights and knaves puzzles, in which there is extensive empirical evidence of

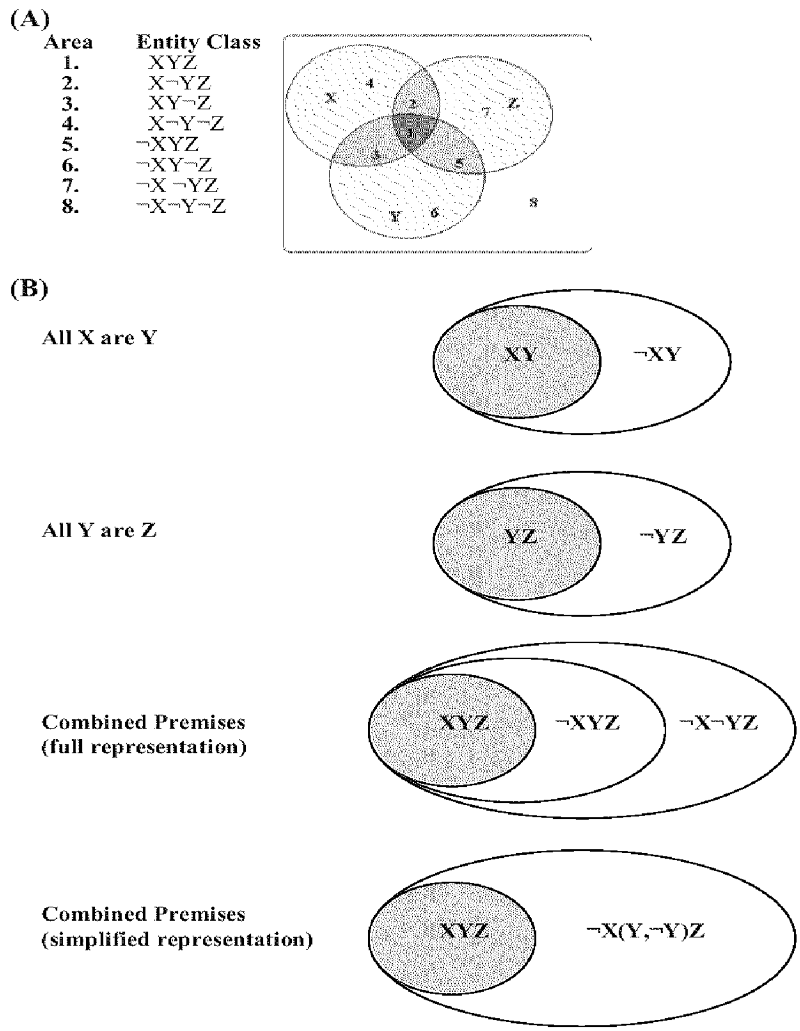

Figure 1. (A) The eight distinct classes of entities formed by combining three binary categories. (B) Interrelation for a subset of categories corresponding to full and simplified representations of the syllogism: *All X are Y* and *All Y are Z*.

cognitive processes used. Analyses are guided by domain-general principles (Andrews et al., 2006; Andrews & Halford, 2002; Birney, Halford, & Andrews 2006; Halford et al., 1998, 2007), the most important of which for our purposes are:

1. Variables can be chunked or segmented only if relations between them do not need to be processed.
2. Effective RC for a cognitive process is the least complex relation required to represent the process, using the least demanding strategy available to humans for that task.
3. RC of a task is based on the most complex step.

When applied to syllogistic reasoning, the guiding principle is that *RC of a syllogism is the lowest complexity value that yields a correct solution*. The RC principles are consistent with the singularity, relevance, and satisficing assumptions proposed by Evans (2006; see Table 1). We assume one mental model (singularity), that mental models are selected on the basis of plausibility (relevance) and models are accepted if they are consistent with current goals (satisficing). The satisficing assumption is particularly important, because it means that participants do not need to know the rationale for their strategies, but only need to be satisfied that the strategy is consistent with their goals. Simplifying heuristics that produce correct answers make some problem forms easier by increasing the range of strategies available, and they influence the number of correct responses. We assume that participants have a tendency to use simplifying strategies and heuristics when they are available, but we do not assume that participants know that the strategies are valid. There is very little evidence that participants know why (or even that) their mental models and heuristics work. We test the proposition that participants use heuristics even when they do not yield a correct answer.

However, individuals will sometimes construct a model that is less complex than needed to yield a valid conclusion. We assume a general tendency to rely on a range of simplifying strategies (which we outline shortly). On higher RC syllogisms, simpler strategies will often not fully represent the structure of the task, leading to error (Halford et al., 1998, Section 3.5), so higher error rates are predicted on higher RC problems.

The principle that RC of a syllogism is the lowest complexity value that yields a correct solution implies that RC analyses are based on computational level theory of cognitive processes, rather than the algorithmic level. Therefore we envisage that there might be more than one algorithm corresponding to a particular RC computational process. This "indeterminism" at the process level accords with recent work which has shown the high degree of variance in participants' syllogistic reasoning strategies (see Bucciarelli & Johnson-Laird, 1999). We are also agnostic about the specific content of a mental model, such as whether it consists of tokens (as in Johnson-Laird's mental model theory), images, or other entities, because they are intertranslatable without affecting RC. RC is determined by the number of distinct classes of entities that have to be related, rather than the content of the entities.

Because our analysis is framed at the computational level, it does not require that participants have any metacognitive knowledge of when various simplification strategies will lead to valid conclusions. In fact, the predictions of the model depend on a lack of such knowledge: Simplifying strategies are not always valid, especially for higher RC problems, and therefore lead to error. If participants had explicit knowledge of this, they could adjust their strategies, and the model would no longer predict more errors on the higher RC problems. As we will see, the RC account yields empirical predictions about performance across the full set of 64 syllogisms, including novel predictions that are not made by any rival theory.

RELATIONAL COMPLEXITY ANALYSIS OF SYLLOGISMS

The most complex step is premise combination, because this is where the most complex relations are represented. As illustrated in Figure 1B, 2, 3, and 4, combined premises represent relations between X, Y, Z, whereas representations of individual premises represent only relations between two variables (XY or YZ). There is also empirical evidence in transitive inference that premise integration imposes the highest processing load (Maybery, Bain, & Halford, 1986). The complexity of premise integration holds unless it is circumvented by atmosphere or figural heuristics, in which case complexity is based on the heuristics. The minimal effective RC for each syllogism, based on the RC of the most complex step in the task, is specified in Table 2.

Our analysis can be broken into the following steps: premise representation, premise combination, conceptual chunking and heuristics that reduce effective RC, and complexity assessment.

Premise representation

Representations required for categorical syllogisms can be defined initially by normative graphical models of the information contained in the premises (Stenning & Oberlander, 1995; Stenning & Yule, 1997). Effective RC is then determined by taking account of strategies and heuristics.

The fully represented and simplified models for the four premise moods are shown in Figure 2. Areas that are necessary (i.e., define cases that must exist) according to the premises are shaded. Areas that are not shaded are considered possible (i.e., define cases that may, but do not definitely, exist). Models are simplified primarily by representing the end term (X or Z) in a premise in full, while *neglecting* areas that are concerned solely with the middle term (Y) of a premise.

The positive universal premise (*All X are Y*) can be simplified by neglecting the possible area representing $\neg$XY, where $\neg X$ means "not X", so that XY is considered only as a single necessary entity. According to the principle of neglect, $\neg$XY is neglected because it does not represent any relation involving X.

The positive particular premise (*Some X are Y*) can be simplified by neglecting the possible predicate area ($\neg$XY). In this case XY is the

TABLE 2
Conclusions, percentage correct, relational complexity (RC), and number of mental models (MM) by figure

	All X are Y	*Some X are Y*	*No X are Y*	*Some X are not Y*
(A) Figure 1 syllogisms: XY–YZ				
All Y are Z	All X are Z 100% Some Z are X RC = 2 MM = 1	Some X are Z 89.5% Some Z are X RC = 2 MM = 1	Some Z are not X 5.3% RC = 4 MM = 3	NVC 4.9% RC = 5 MM = 2
Some Y are Z	NVC 15.8% RC = 4 MM = 2	NVC 17.8% RC = 4 MM = 2	Some Z are not X 31.6% RC = 4 MM = 3	NVC 17.8% RC = 5 MM = 2
No Y are Z	No X are Z 94.7% No Z are X RC = 2 MM = 1	Some X are not Z 52.6% RC = 3** MM = 3	NVC 25.1% RC = 4 MM = 2	NVC 20.2% RC = 4 MM = 2
Some Y are not Z	NVC 8.5% RC = 4 MM = 2	NVC 13.8% RC = 4 MM = 2	NVC 26.3% RC = 4 MM = 2	NVC 34.8% RC = 4 MM = 2
(B) Figure 2 syllogisms: YX–ZY				
All Z are Y	All Z are X 73.7% Some X are Z RC = 2 MM = 1	NVC 6.9% RC = 4 MM = 2	No X are Z 94.7% No Z are X RC = 2 MM = 1	NVC 12.1% RC = 4 MM = 2
Some Z are Y	Some X are Z 89.5% Some Z are X RC = 2 MM = 1	NVC 16.6% RC = 4 MM = 2	Some Z are not X 31.6% RC = 3** MM = 3	NVC 19.0% RC = 4 MM = 2
No Z are Y	Some X are not Z 10.5% RC = 4 MM = 3	Some X are not Z 5.3% RC = 4 MM = 3	NVC 24.3% RC = 4 MM = 2	NVC 29.6% RC = 4 MM = 2
Some Z are not Y	NVC 9.7% RC = 5 MM = 2	NVC 26.3% RC = 5 MM = 2	NVC 19.0% RC = 4 MM = 2	NVC 30.0% RC = 4 MM = 2

Table 2 (*Continued*)

	All X are Y	*Some X are Y*	*No X are Y*	*Some X are not Y*
(C) Figure 3 syllogisms: XY–ZY				
All Z are Y	NVC 26.3% RC = 4 MM = 2	NVC 19.0% RC = 4 MM = 2	No X are Z 84.2% No Z are X RC = 2 MM = 1	Some X are not Z 15.8% RC = 3 MM = 2
Some Z are Y	NVC 15.4% RC = 4 MM = 2	NVC 27.1% RC = 4 MM = 2	Some Z are not X 21.1% RC = 4 MM = 3	NVC 26.3% RC = 5 MM = 2
No Z are Y	No X are Z 73.7% No Z are X RC = 2 MM = 1	Some X are not Z 21.1% RC = 4 MM = 3	NVC 41.3% RC = 4 MM = 2	NVC 24.3% RC = 4 MM = 2
Some Z are not Y	Some Z are not X 26.3% RC = 3 MM = 2	NVC 25.1% RC = 5 MM = 2	NVC 23.9% RC = 4 MM = 2	NVC 34.0% RC = 4 MM = 2
(D) Figure 4 syllogisms: YX–YZ				
All Y are Z	Some X are Z 31.6% Some Z are X RC = 4 MM = 3	Some X are Z 78.9% Some Z are X RC = 2 MM = 1	Some Z are not X 21.1% RC = 4 MM = 3	Some Z are not X 52.6% RC = 3 MM = 2
Some Y are Z	Some X are Z 68.4% Some Z are X RC = 2 MM = 1	NVC 39.7% RC = 4 MM = 2	Some Z are not X 15.8% RC = 4 MM = 3	NVC 34.8% RC = 4 MM = 2
No Y are Z	Some X are not Z 15.8% RC = 4 MM = 3	Some X are not Z 31.6% RC = 4 MM = 3	NVC 49.4% RC = 4 MM = 2	NVC 23.1% RC = 4 MM = 2
Some Y are not Z	Some X are not Z 47.4% RC = 3 MM = 2	NVC 30.8% RC = 4 MM = 2	NVC 29.6% RC = 5 MM = 2	NVC 44.1% RC = 4 MM = 2

**RC based on atmosphere and figural heuristics.

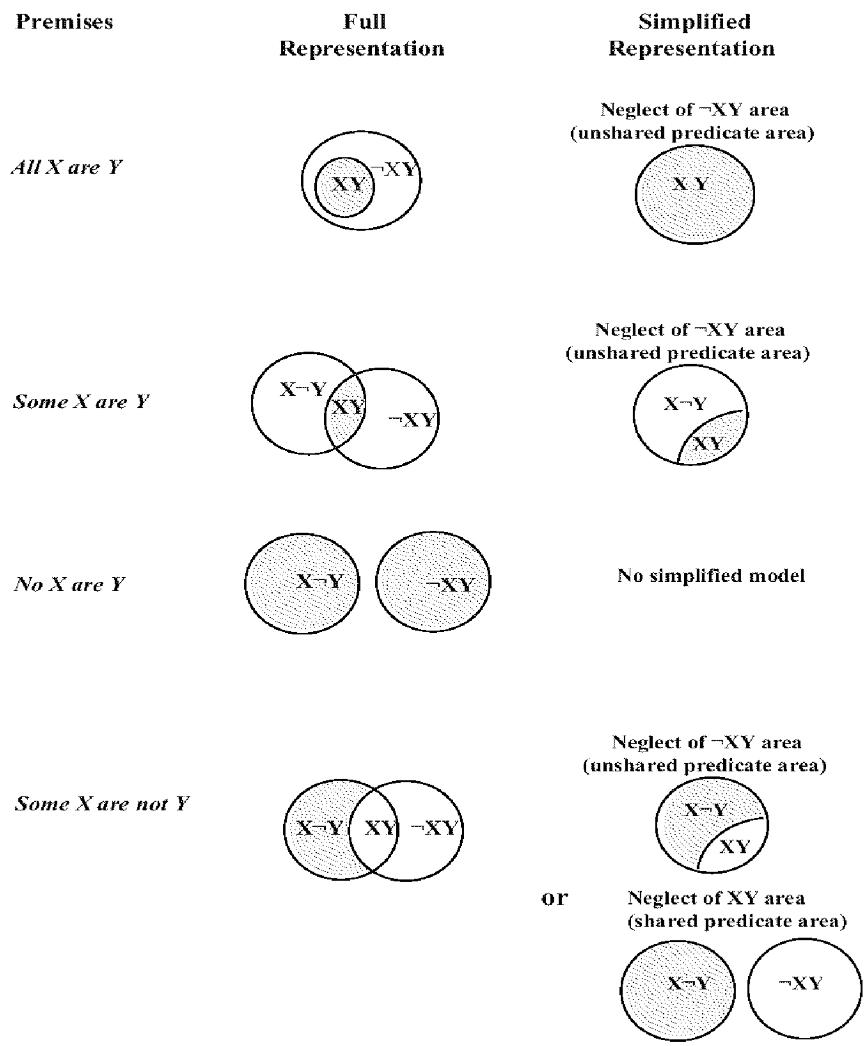

Figure 2. Full and simplified representations of the four premise moods.

necessary area, whereas X ¬ Y is retained as a possible area. Again, ¬ XY is neglected because it does not represent any relation to X.

The negative universal premise (*No X are Y*) cannot be simplified as both mutually exclusive areas must be portrayed. This reflects the standard interpretation of syllogisms in which no variable represents an empty set.

The negative particular premise (*Some X are not Y*) can be simplified so that again only the subject of the premise is represented in full, the remaining predicate area being neglected. For this premise, X ¬ Y is the necessary area, whereas XY is the remaining possible area.

We also suggest a second method of simplifying the negative particular premise (*Some X are not Y*) again reducing the model from three to two areas to be considered. In this case the shared region of the full premise model is neglected, and the resultant simplified model equates to the negative universal premise (*No X are Y*). Again, in this model, $X \neg Y$ is the necessary area, whereas $\neg X$ Y is the remaining possible area. Thus, both suggested simplifications of the negative particular premise (*Some X are not Y*) preserve the one necessary area, $X \neg Y$, but lend different priority to the two remaining possible areas.

Although RC analyses do not depend on participants knowing the rationale for their strategies (use of strategies without full understanding is common, and arguably typical) we expect participants would recognise that full representation of end terms of the premises has priority. A similar principle is incorporated in the Verbal Reasoning model (Polk & Newell, 1995). However, we assume that the high processing demands of premise integration will often lead participants to default to more simplified premise representations, which can sometimes lead to error. This is a form of conceptual chunking, because it reduces the dimensionality of the representation. For some problems, this process is efficient, and preserves logical validity, but for other problems—most notably the higher RC problems—it leads to error.

Premise combination

The premise combination principles we use can be applied to both simplified and full models. First, middle term (Y) areas are made to correspond (overlap completely), and then the end term (X, Z) areas are combined in a maximal way (representing all possible areas).

There is a valid conclusion if, and only if, there is a necessary area (Stenning & Oberlander, 1995). A necessary area in the premises is preserved as necessary in the integrated representation, if and only if it is not subdivided in the process of integration—it must remain intact (see Stenning & Oberlander, 1995). Only then can a valid conclusion can be drawn, and the nature of such a conclusion can be identified by exploring the relationship between the necessary area(s), and the other end term areas in the model. If there is no necessary area preserved in the integrated representation, then no valid conclusion can be drawn.

Complexity assessment

Once a graphical model of effective RC has been established, RC for solutions based on MM of the premises is calculated by simply counting the number of distinct spatial areas that are represented, which corresponds to the number of distinct classes of entities that are related in the representation. The results of this analysis across the 64 problems, and the comparison with MM theory is shown in Table 2. Complexity varies from two spatial areas (binary relation) to five spatial areas (quinary relation).

EXAMPLE PROBLEMS

The following problems illustrate some of the principles underlying our complexity analysis. Consider the following problem:

All X are Y
Some Z are not Y ⇒ Some Z are not X.

Full and simplified representations are shown in Figure 3. The first two rows show both full and simplified representations of the premises. For the purpose of this example we will only consider one of the possible simplified representations of the second premise. The third row shows full and simplified valid representations of combined premises (valid representations are those from which only the correct conclusion can be drawn). As before,

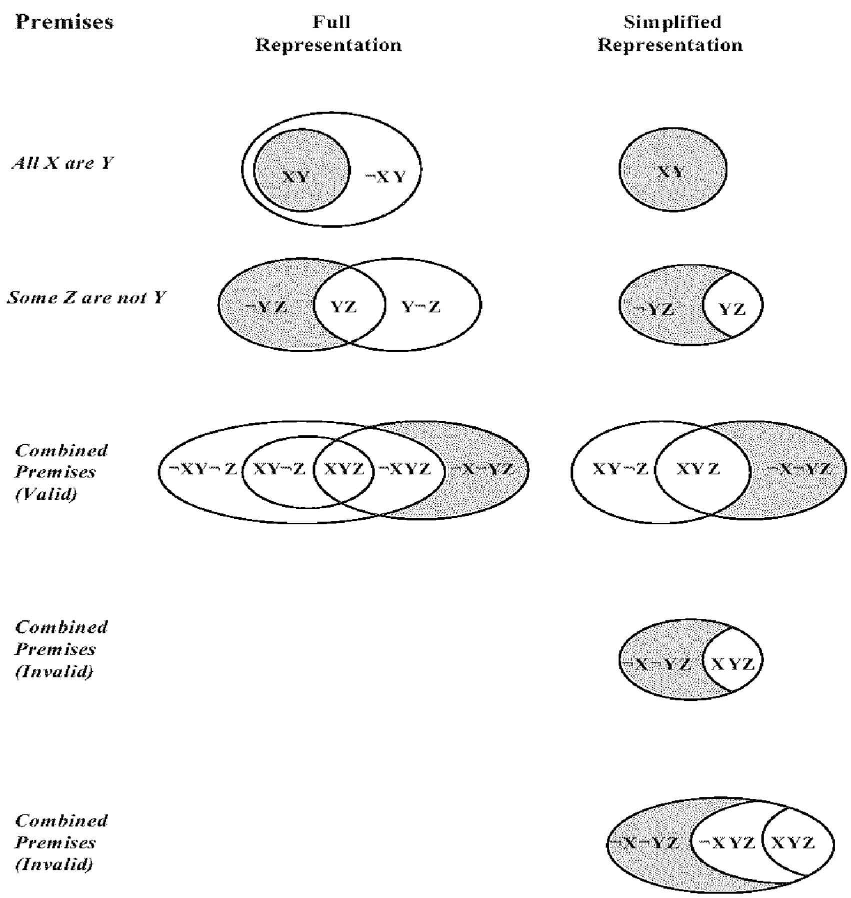

Figure 3. Full and simplified representations of the syllogism: *All X are Y* and *Some Z are not Y.*

areas that necessarily exist according to the premises are shaded. In this case we have three areas to consider. There is one necessary area:

1. ¬X ¬Y Z: there *must* be Zs that are not Ys and thus not Xs.

It is also possible that there *may* exist:

2. X Y Z: Zs that are Ys and also Xs; and
3. X Y ¬Z: Xs that are Ys but not Zs.

These three pieces of information together lead to the valid conclusion: Some Z are not X, and the RC of this syllogism is ternary. This conclusion is consistent with the one which would be yielded from the full representation of the combined premises shown on the left, where again the only conclusion is Some Z are not X.

Now let us consider the problem:

No X are Y
All Y are Z ⇒ *Some Z are not X.*

This problem does not lend itself to simplification. The representations are shown in Figure 4. Premise 1 cannot be simplified any further. Using the reduced representation of Premise 2 and combining it with the model of Premise 1, we would end up with a model of the combined premises as shown in Figure 4. This simplified representation of combined premises is invalid as it would lead to two erroneous conclusions: *No X are Z* and *No Z are X*. Therefore, the simplification heuristic would not yield correct answers

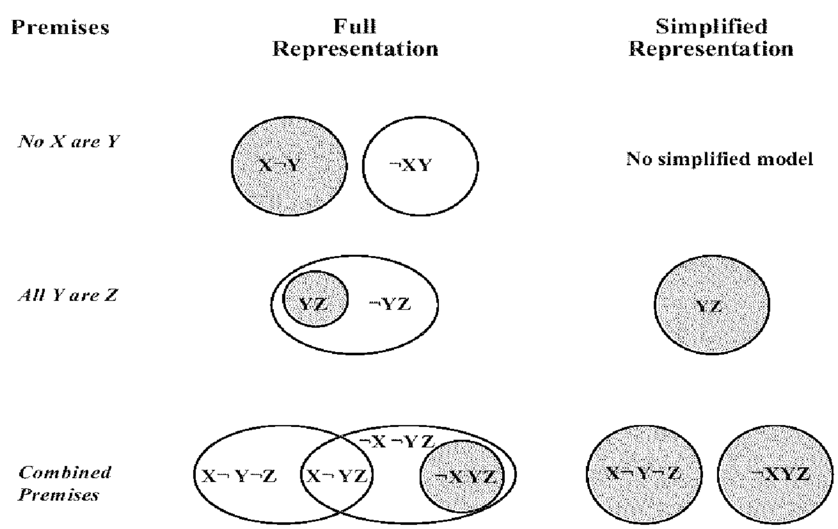

Figure 4. Full and simplified representations of the syllogism: *No X are Y* and *All Y are Z*.

in this case, in violation of Principle 2. In contrast, the full representation of combined premises generates only the valid conclusion: *Some Z are not X*.

Thus, the only valid model here combines full representations of both premises. There are four areas in this model, so the RC of this syllogism is quaternary.

REVERSIBILITY PRINCIPLE

For premises with moods, *Some* and *No*, subject and object roles can be switched without affecting the meaning of the premise. However, reversing the premises changes the *simplified representation* of the premise, and can reduce complexity.

For example, consider the syllogism:

Some Y are X
No Z are Y ⇒ *Some X are not Z*.

Complete representations of the combined premises would contain five distinct areas, and so the RC of this syllogism would be classed as quinary. By reversing the first premise and applying the principle of neglect, the representation is reduced to four areas (see Table 2).

ADJUSTMENT FOR "HEURISTIC" SOLUTIONS: FIGURAL AND ATMOSPHERE EFFECTS

Figural (Johnson-Laird & Byrne, 1991; Johnson-Laird & Steedman, 1978) and atmosphere (Begg & Denny, 1969; Woodworth & Sells, 1935) heuristics can reduce the complexity of the representation that is used. The figural heuristic implies that if the figure is XY–YZ, the conclusion should be in the form X-Z, while if the figure is YX–ZY, the conclusion (after swapping the order of the premises) should be in the form Z-X (see Table 1). The simplest application of the figural heuristic entails matching the terms of the conclusion to the end terms in the premises, so is a mapping of a binary relation. Thus this strategy is binary relational. The heuristic is likely to be cued by the end terms, so XY–YZ figures cue a conclusion beginning with X, and so on, as suggested by Evans (2006) for other reasoning tasks.

The atmosphere heuristic incorporates attributes of quality (positive, negative) and quantity (universal, particular). The principle of quality asserts that where one or both premises are negative, the conclusion will be negative, otherwise it will be positive. This corresponds to the rule premise(negative) which is a unary relation. This is mapped to conclusion(negative) which is also a unary relation. The mapping of one unary relation to another is a unary relation (Halford et al., 1998, Section 6.2.2). The principle of quantity asserts that where one or both premises are particular, the conclusion will be particular, otherwise it will be universal. This corresponds to the mapping of premise(particular) to conclusion(particular) and is a unary relation. The combination of the principle of quality and the principle of quantity is a mapping of two variables into a new variable, and is a ternary relation: premise(negative) + premise(particular) $\Rightarrow$ conclusion(negative, particular).

For eight Figure 1 or 2 problems, valid conclusions can be generated by applying both the figural and atmosphere heuristics. For a further two Figure 4 problems, the atmosphere heuristic can be used in conjunction with either order of end terms (X-Z or Z-X) to generate a valid conclusion.

Figural and atmosphere heuristics can be segmented into separate steps, because they are independent: in principle (the figural heuristic can be applied without the atmosphere heuristic, and vice versa). However, only the atmosphere heuristic can yield a valid conclusion by itself (for the two Figure 4 problems, just mentioned, which have symmetrical conclusions). The figural heuristic can never be employed alone to yield a valid conclusion. Our complexity analyses accord with this observation: The RC of the syllogism is equal to the RC of the least complex heuristic, or the least complex combination of heuristics, that will yield a valid conclusion. When application of these heuristics offers a less complex means of determining the correct solution to a problem than can be generated by the construction of a mental model of the combined premises, the problem is assigned a complexity appropriate to these heuristics, by Principle 2 (noting that Principle 2 defines complexity by the least complex relation required to perform the process). It should be noted that for only two problems is the application of the figural and atmosphere heuristics used to define the level of RC (see Table 2). Although the figural and atmosphere heuristics can be used to generate a correct solution on a number of problems as outlined previously, construction of a mental model is generally found to be no less parsimonious than use of these heuristics.

PREDICTIONS AND EXPERIMENT

First, it is predicted that the difficulty of syllogisms should increase with their relational complexity, and that our metric will capture the variance between syllogisms as well as MM theory, across the full range of 64 syllogisms. This would demonstrate that the RC metric, which has been shown to be effective elsewhere, can be validly applied to categorical syllogisms. Because the field has been heavily worked we doubt it is possible to exceed the proportion of variance accounted for by other models, which might already be at asymptotic level.

Second, we predict that the no valid conclusion (NVC) response will be used for complex problems which do have valid conclusions. The reason is that NVC can be a default, or a surrogate "don't know", where processing demands are excessive.

Third, because our analysis categorises the NVC problems as being among the most complex (Levels 4 and 5), and we predict that the NVC response will be used for hard problems, correct answers to NVC problems will be artificially inflated. Neglect of this factor could distort the complexities of categorical syllogisms.

Fourth, the tendency to default to NVC responses on complex VC problems will be greater when combined figural and atmosphere heuristics cannot be used (i.e., on Figure 3 and 4 problems).

Fifth, availability of figural and atmosphere heuristics influences responses even where the heuristics cannot yield the correct answer (as on NVC problems).

METHOD

Participants

Twenty undergraduate students from the University of Queensland were recruited by means of advertisement throughout the campus. Twelve male subjects aged 17–38 years ($M = 22$, $SD = 6.34$) had one to four years of tertiary study ($M = 2.5$, $SD = 0.90$). Eight female subjects aged 18–21 years ($M = 19.5$, $SD = 1.07$) had one to four years of tertiary study ($M = 2.63$, $SD = 0.92$). Participants were paid \$40.00 for participating and each syllogism answered correctly earned a bonus of AU\$0.50. No participant had previously studied categorical syllogisms.

Materials

Tasks were presented on IBM-compatible computers with 15-inch colour monitors, and subjects were not permitted the use of pencil and paper in this study. Participants were presented with a set of two premises and asked to select the most appropriate conclusion for those premises from a set of nine alternative conclusions, as shown later. Selection was made by clicking the left button of the mouse. If two answers were deemed equally valid, they were to select either alternative. This procedure was adopted because it yields percentages correct that are comparable across the entire set of problem forms, and it enabled us to determine which set of responses each participant makes, both of which are essential to testing our hypotheses. There did not appear to be any archived data that were suitable for our purposes in these respects.

Conclusions, presented in random order, comprised the following alternatives:

- All X are Z
- All Z are X
- Some X are Z
- Some Z are X
- No X are Z
- No Z are X
- Some X are not Z
- Some Z are not X
- No valid conclusion.

Payment and provision of the nine alternatives were both intended to promote deep processing of the premises. However, forward processing, from premises to conclusion, was still expected. A backward strategy would be unlikely here (unlike methodology which requires verification of a single conclusion) because at least some processing of premises would be required to choose one of the exhaustive set of conclusions offered.

Test procedure

Each participant completed two practice items on the computer. Both items had the relatively simple form:

All X are Y
All Y are Z $\Rightarrow$ *All X are Z.*

The initial practice item was based on occupation (Doctors, Accountants, Lawyers); the second introduced the participants to the idea of representing categories by letters (A, B, C). Feedback was given. If the correct response was not given for the second practice item, the experimenter reviewed the practice problems with the participant to ensure understanding.

Test items

Each participant completed 64 categorical syllogisms in random order. All test problems used abstract contents, the letters X, Y, Z, as the premise variables. No feedback was given. Responses and response times were recorded automatically.

RESULTS

One participant was eliminated due to consistently rapid, poor performance that was outside the range of the rest of the group.[1] Table 3 shows the percentages of correct responses and latencies over each level of complexity. Only problems ranked as quaternary consist of both valid conclusion (VC) and (NVC) problems, thus we separate out these two types in Table 3. For all NVC syllogisms (RC 4 and 5 only), Table 3 also presents accuracy data that corrects for default NVC responding (see later).

As Table 3 indicates, prediction one was strongly confirmed—as RC increased, accuracy decreased monotonically. A one-way ANOVA with four levels of complexity, counting all problem forms, uncorrected, in each condition, yielded $F(3, 60) = 32.75$, $p < .001$.

NVC default responses

Supporting predictions two and three, there is evidence that the NVC response is given as a default when the complexity of the syllogism is high. Figure 5 indicates a higher level of accuracy on NVC problems when response time is limited to 10 s, as in Johnson-Laird and Bara's (1984) Experiment 1 (in which participants generated conclusions), than when there is no time limit, as in their third experiment (again, conclusion generation) and the current experiment (in which participants chose a conclusion from a set of options). Moreover, as Figure 5 shows, there was no accuracy difference between quaternary and quinary problems in Johnson-Laird and Bara's time-limited condition, whereas there is some observable decline for quinary problems in both of the conditions without a time limit. NVC problems are difficult, and there was no "don't know" option in any of these experiments, so it appears that when participants are required to respond quickly on difficult problems, they tend to use NVC as the default response.

This interpretation is further supported by the findings from the current experiment for VC problems, shown in Figure 6. Increased complexity produces longer latency for correct responses, and shorter latency for incorrect responses. Although a univariate ANOVA of latencies showed no significant main effects of RC, $F(2, 507) = 0.12$, $p = .89$, or accuracy,

[1] The participant produced 20% correct responses on the easiest syllogisms compared with the group mean of 85%, 70% NVC responses compared with 7% for the group, and a response latency of 8.49 s compared with 23.77 s for the group.

TABLE 3
Percentages of correct responses and mean latencies (with standard deviations) over each level of complexity

		Mean latency	
Relational complexity (RC)	*Percentage correct*	*Correct*	*Incorrect*
2 (VC)	84.7 (36.1)	31.4 (23.5)	42.1 (43.2)
3 (VC)	37.7 (48.7)	42.5 (27.8)	39.7 (29.3)
4 (All)	33.3 (47.1)	38.8 (38.3)	33.0 (31.4)
4 (VC)	19.1 (39.4)	48.0 (33.1)	32.6 (27.3)
4 (NVC)	38.4 (48.7)	37.1 (39.0)	33.2 (33.2)
5 (NVC)	30.8 (46.4)	39.0 (35.3)	35.9 (34.4)
4 (NVC) corrected (*)	25.1 (35.1)	—	—
5 (NVC) corrected (*)	20.0 (32.7)	—	—

Corrected (*) = corrected for NVC bias by percentage of NVC responses for quaternary VC problems.

$F(1, 507) = 0.11$, $p = .77$, there is a significant interaction, $F(2, 507) = 6.027$, $p = .003$. When problems were solved correctly, planned pairwise comparisons (Bonferroni-corrected, two comparisons) indicated that it took significantly less time to solve binary (31.56 s) than ternary problems (42.54 s, $p = .04$) and quaternary problems (48.04 s, $p = .001$). No significant differences were found between individual levels of complexity when problems were not solved correctly. However, a separate ANOVA indicated that incorrect responses on quaternary problems took less time than did incorrect responses on combined binary/ternary problems, $F(1, 268) = 4.273$,

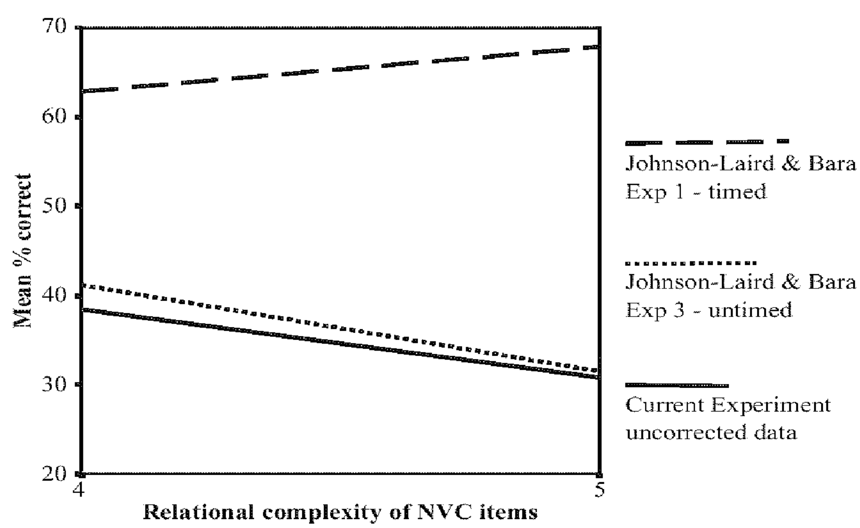

Figure 5. Mean accuracy for NVC problems over three experiments.

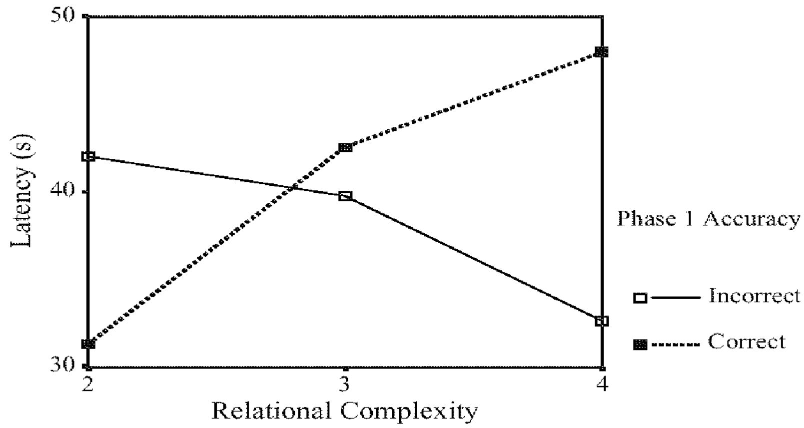

Figure 6. Estimated means of response time (s) for VC problems.

$p = .04$. Increasing complexity leads to an increasing likelihood of a rapid, incorrect response.

The default interpretation of NVC responses is corroborated by two further findings. The number of NVC responses for VC problems, shown in Figure 7A, is greater for complex problems (RC3 and RC4) than for simple problems (RC2). Supporting Prediction 4, number of NVC responses is greater for Figures 3 and 4 problems, for which combined figural and atmosphere heuristics do not apply. The figural heuristic cannot be employed for Figures 3 and 4 problems, because their premise structure (X-Y, Z-Y; Y-X, Y-Z, respectively) does not yield an ordering of the X and Z elements. (Although see earlier discussion—there are two Figure 4 problems that the atmosphere heuristic alone can yield a valid conclusion, because the valid conclusion is symmetrical). A 2 (Figures 1 and 2/Figures 3 and 4) $\times$ 3 (binary, ternary, quaternary complexity) ANOVA yielded significant main effects of figure, $F(1, 21) = 11.98$, $p = .002$, and of complexity, $F(2, 21) = 6.12$, $p = .008$, with no interaction. Binary problems produced significantly fewer NVC responses than ternary ($p = .023$) and quaternary problems ($p = .003$), which did not differ significantly. This suggests that NVC responses are used as a default for problems that are complex, and for which heuristics based on a combination of figure and atmosphere are unavailable, due to criteria not being met, as explained earlier. The finding of no increase in NVC responses from RC3 to RC4 suggests RC3 problems were sufficiently complex to elicit the NVC response in those participants who were inclined to adopt that default. The same effect of figure is observed on the number of NVC responses for NVC problems, shown in Figure 7B. A 2 (Figures 1 and 2/Figures 3 and 4) $\times$ 2 (quaternary/quinary complexity) ANOVA yielded a main effect of figure, $F(1, 33) = 10.39$, $p = .003$, but no effect of complexity and no interaction. The lack of complexity effects is probably due to the

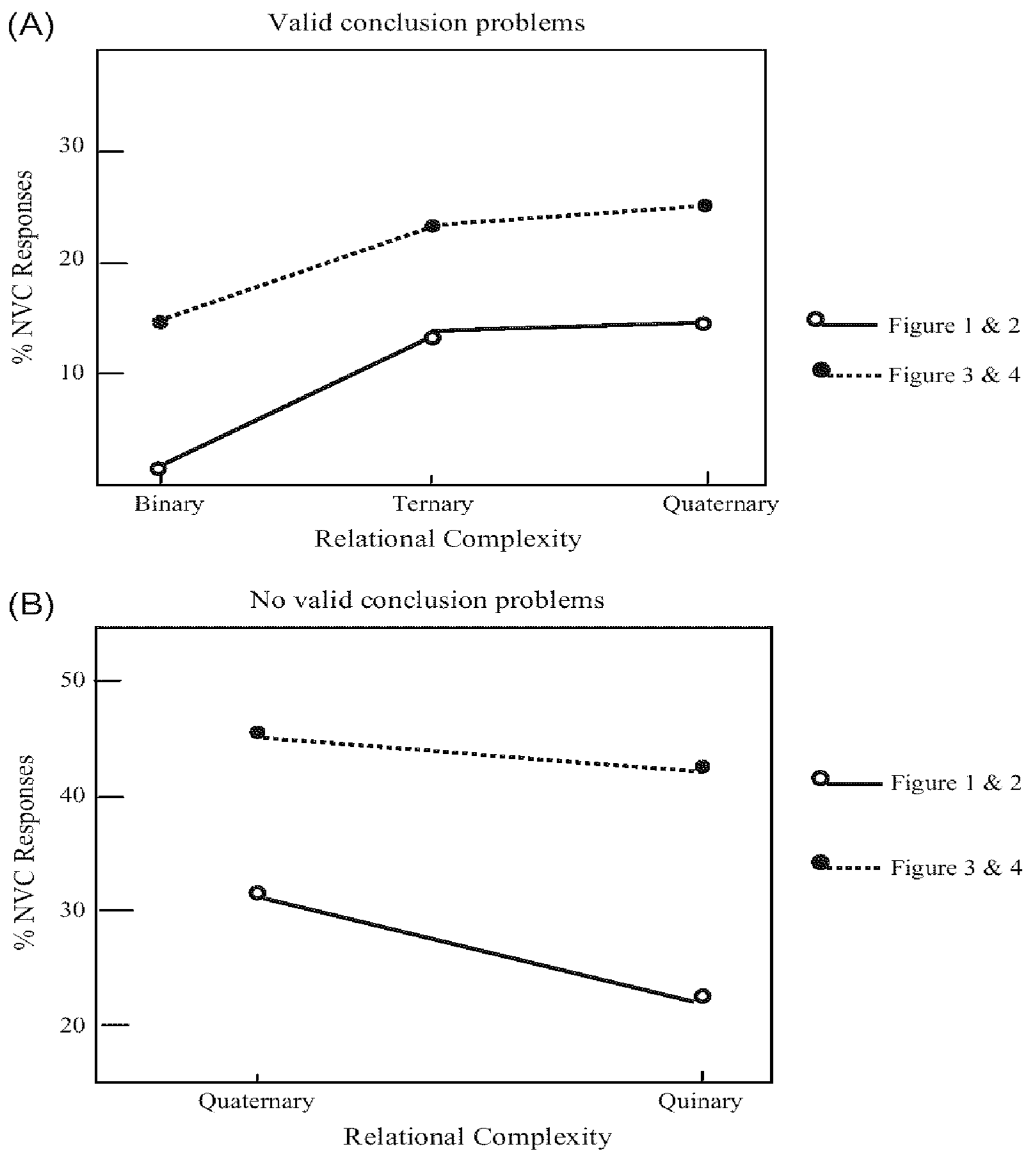

Figure 7. Percentage of NVC responses for Figures 1 and 2 versus Figures 3 and 4 on (A) VC and (B) NVC problems.

small range of complexities, both being close to the limit of capacity (four variables).

However, the data indicate that some ability to discriminate validity of inferences does exist, because there are more NVC responses for invalid RC4 syllogisms (38.4%) than for valid RC4 syllogisms (21.05%), $F(1, 778) = 21.14$, $p < .001$.

Supporting prediction five, participants were sensitive to the availability of the figural-atmosphere heuristic even where it can never yield correct answers, as in NVC problems. Table 4 shows types of errors on NVC

TABLE 4
Response type for syllogisms with no valid conclusion

	Figure 1 and 2 problems		*Figure 3 and 4 problems*	
*Response type**	*Number*	*Percentage*	*Number*	*Percentage*
1	113	29.7	147	45.5
2	122	32.1	N/A	N/A
3	28	7.4	85	26.3
4	38	10.0	29	9.0
5	53	13.95	35	10.8
6	26	6.85	27	8.4
Total	380	100	323	100

*Response type:
1. No valid conclusion (correct response)
2. Response based on combined figural and atmosphere heuristics
3. Response based on atmosphere heuristic
4. Response reflects mood of one premise only
5. Response mood is "Some" when premise mood is "Some not"
6. Other response.

problems, expressed as frequencies and percentages. Data for Figure 1/2 and Figure 3/4 problems are given separately. It is clear that for Figure 1/2 problems the most common response is that generated using the combined figural-atmosphere heuristic, closely followed by the NVC response. Together these two answers account for 61.8% of participants' responses. If we include subjects whose responses suggested use of the atmosphere heuristic only, we can account for 69.2% of responses on Figure 1/2 problems. A similar percentage of responses on Figure 3/4 problems is accounted for by considering NVC responses and responses generated by using the atmosphere heuristic (noting that the figural heuristic is not available for these problems). These two sets of responses account for 71.8% of answers on Figure 3/4 problems with NVC.

Other notable errors (see Table 4, response types 4 and 5), accounting for around 20–24% of responses, included a tendency to base the mood of the response on the mood of one premise only where that strategy does not match the atmosphere heuristic (response type 4). For example, consider the syllogism:

All X are Y
Some Z are not Y $\Rightarrow$ *Some Z are not X.*

For this syllogism, approximately 10% of subjects responded "*All X are Z*". A second set of error responses (response type 5) found subjects giving a response in the form of "Some" when "Some not" was the mood of at least one of the initial premises. These particular response errors suggested a percentage of subjects were having difficulty understanding the logical difference between the premise moods, “Some” and “Some not”.

However, although analysis of error type has its own value and interest, for the sake of this paper, the important conclusions to be drawn from the data displayed in Table 4, are that for complex syllogisms for which figural and atmosphere heuristics cannot provide a correct answer (i.e., the NVC syllogisms) subjects were most likely to default to the combined figural and atmosphere heuristics. For those syllogisms that do not lend themselves to this set of heuristics, the default solution was NVC (note that 45.5% of Figure 3/4 problems were answered NVC vs. 29.7% of Figure 1/2 problems). The third most common response strategy for these complex syllogisms was use of the atmosphere heuristic without the figural heuristic.

Because RC analyses are based on cognitive processes employed, we have adjusted our accuracy data for NVC problems to estimate for (and remove) default responses. For each participant we calculated the percentage of “NVC” responses to the quaternary VC problems (i.e., where complexity is high). We then subtracted this percentage from the overall percentage of “NVC” responses. This avoids inflation of correct responses to with false positive NVC responses. However, NVC is not like heuristics such as atmosphere and figural heuristics, in that it is not based on recognised properties of the premises (such as attributes of quality, quantity, or structural features of the premises) but is a default response to complexity.

Comparison of mental models and relational complexity

In this section we estimate how well each metric predicted performance, using a number of variations of the MM metric. Recent formulations of the mental model theory emphasise the distinction between one and multiple model problems (Bucciarelli & Johnson-Laird, 1999; Johnson-Laird & Byrne, 1991, 1996); hence, we collapse their two- and three-model problems. We refer to this measure as MM_reduced.

We make separate analyses for all syllogisms, and for syllogisms with valid conclusions. Linear multiple regressions were performed with RC and MM_reduced entered as independent predictors, and percentage correct as the dependent variable. We also consider linear models of the data.

Linear multiple regression—all syllogisms. RC and MM_reduced account for similarly large proportions of the variance. Table 5 shows the results of the regression analyses run using the version of our data corrected by individual subject “NVC” responses. Across all 64 syllogisms (not differentiating between VC and NVC problems) the strength of the regression model was

TABLE 5
Complexity predictions from multiple regression analyses of three data sets

Variables	*Correlation*	*Strength (R^2)*	*Significance (F-value)*	*Nature of regression model* β	*t*	*Sig*	*sr^2*
All syllogisms, our data corrected for NVC response bias							
1. RC	.852	.810	$F(2, 61) = 129.675$,	1. −.299	−2.80	<.001	.025
2. MM_reduced			$p < .001$	2. −.631	−5.91	<.001	.109
VC syllogisms, our data corrected for NVC response bias							
1. RC	.903	.880	$F(2, 24) = 88.25$,	1. −.522	−3.18	.004	.051
2. MM_reduced			$p < .001$	2. −.439	−2.68	.013	.036
Johnson-Laird & Bara (1984) Exp. 1 (timed)							
1. RC	.903	.801	$F(2, 24) = 48.445$,	1. −.523	−2.47	.021	.051
2. MM_reduced			$p < .001$	2. −.395	−1.87	.074	.029
Johnson-Laird & Bara (1984) Exp. 3 (untimed)							
1. RC	.903	.906	$F(2, 24) = 115.15$,	1. −.538	−3.695	.001	.054
2. MM_reduced			$p < .001$	2. −.437	−3.00	.006	.035

relatively high, $R^2 = .810$, when both RC and the MM_reduced were included as predictors. The relatively high correlation of the predictors (.852) influenced the relative amounts of shared variance (.676) and unique variance (.134) explained by this model. RC and MM_reduced jointly accounted for a high proportion of the variance and both contributed significant unique variance, though this was slightly greater for MM_reduced.

This analysis was repeated, but with MM_reduced expanded to three levels as done by Bucciarelli and Johnson-Laird (1999). This showed that the maximum variance explained (.79) was close to, but slightly less than that explained by the analysis including the binary MM_reduced variable. A moderate correlation between the predictors (.456) influenced the relative amounts of shared variance (.34) and unique variance (.45) explained by this model. Although both predictors contributed significantly to the model, RC outperformed this version of MM_reduced explaining 82% of the unique variance contributed by the two predictors.

Curve estimation (linear and quadratic models)—all syllogisms. MM_reduced is a binary variable, so only a linear model of fit is possible. This model accounted for 78.5% of the variance, $F(1, 62) = 144.96$, $p < .001$, across all syllogisms. In contrast, the relationship between RC and accuracy data is best described by a quadratic model, $F(1, 61) = 130.31$, $p < .001$, accounting for 81% of variance (compared with a linear model accounting for 70% of variance), $F(1, 62) = 144.96$, $p < .001$. This appears to reflect a floor effect on quaternary and quinary relational problems (Table 3) suggesting that quinary relational problems are beyond participants' processing capacity (Halford et al., 1998), so normal strategies are no longer used.

Linear multiple regression—VC syllogisms. Considering the VC syllogisms in isolation (see Table 5), the strength of the regression model was quite high, $R^2 = .880$, when both RC and MM_reduced were included as predictors. High correlation between the predictors (.903) meant that shared variance (.793) was greater than the unique variance (.087) explained by this model. Although both predictors contributed significantly to the model, RC (58.6% of unique variance) slightly outperformed MM_reduced (41.4% of unique variance). A similar analysis using the full version of MM, analysed by the method of Johnson-Laird and Byrne (1991), was carried out. The overall strength of the regression model was high ($R^2 = .848$), as was the correlation between the predictors (.958). Although only RC contributed significant unique variance (91% of .046), most of the variance explained by the model was shared variance (.802).

Table 5 also shows regression analyses of data by Johnson-Laird and Bara (1984), both timed (Exp. 1) and untimed (Exp. 3). Given evidence presented earlier for inflated NVC responses, and Johnson-Laird and his colleagues' recent distinction between VC and NVC problems (Bucciarelli & Johnson-Laird, 1999), we examine VC syllogisms only.

Looking first at the timed condition (Exp. 1), the overall strength of the regression model was high at .880. High correlation between the predictors (.903) meant that shared variance (.721) was greater than the unique variance (.080) explained by this model. RC (63.75% of unique variance) was the only significant predictor in the model.

In the untimed condition (Exp. 3), the overall strength of the regression model was high at .906. High correlation between the predictors (.903) meant that shared variance (.817) was greater than the unique variance (.089) explained by this model. Although both predictors contributed significantly to the model, RC (60.7%) contributed more unique variance to the model than did MM_reduced (39.3%).

Curve estimation (linear and quadratic models)—VC syllogisms. For VC syllogisms only, the relationship between RC and accuracy data, across all data sets, is best described by a quadratic rather than a linear model (see Tables 3 and 6). We interpret the quadratic function as indicating a floor effect for quaternary problems. In all cases, both the linear and the quadratic models of RC explain more variance in subject performance than does the mental model variable.

DISCUSSION

This paper is to the first to show how RC analyses can be applied to syllogistic reasoning, and the research has demonstrated that it accounts for a useful proportion of variance in problem form difficulty. Thus, we have demonstrated that complexity analysis of syllogistic reasoning is possible, and we have also shown how people can simplify syllogistic reasoning to comply with capacity limitations. The RC metric has proven to be approximately as effective as MM in accounting for performance on categorical syllogisms. Although each metric has a small advantage in specific analyses, the dominant finding is that they account for similarly large proportions of the variance in performance, approximately 80%, regardless of how complexity is analysed in each model. RC and MM share common conceptions of the reasoning process, but differ in the explanations they imply for the effect of complexity. MM theory links the effect of complexity to construction of alternate mental models of the premises, which is influenced by recognition that one model might not be the only possible representation, which is an important factor in reasoning. RC theory accounts for complexity effects by the number of distinct categories of entities that are related in a representation of the premises. RC is directly

TABLE 6
Curve estimation for mental models (binary variable) and relational complexity across three data sets for VC syllogisms

	Variance explained (R^2)					
	Linear model				*Quadratic model*	
Data set	*MM_reduced*	*F-value*	*RC*	*F-value*	*RC*	*F-value*
Current	.83	$F(1, 25) = 121.83, p < .001$	.845	$F(1, 25) = 135.83, p < .001$	.88	$F(1, 24) = 88.25, p < .001$
Exp. 1* timed	.75	$F(1, 25) = 75.35, p < .001$	.77	$F(1, 25) = 84.95, p < .001$	.80	$F(1, 24) = 48.45, p < .001$
Exp. 3* untimed	.85	$F(1, 25) = 143.84, p < .001$	.87	$F(1, 25) = 167.69, p < .001$	.91	$F(1, 24) = 115.15, p < .001$

*Johnson-Laird and Bara (1984).

related to working memory demands, because measures of ability to bind elements into integrated representations in working memory are good predictors of reasoning (Oberauer et al., 2008).

Expansion of the 64 syllogism set to a 576 syllogism set (Roberts, 2005) offers potential for further discrimination between theories, but more information about strategies used with the expanded set is required before the RC metric can be applied to it. The 64 syllogism set is restricted, but the extensive data base on premise interpretations and strategies is valuable, indeed essential, for application of the RC metric. Furthermore, the finding that RC accounts for a similar proportion of variance to MM is consistent with findings by Andrews et al. (2006) that RC accounts for a similar proportion of variance in sentence complexity as the dependency locality theory metric (Gibson, 2000), again approximately 80%. In both categorical syllogisms and sentence comprehension, RC performs as well as a metric that has been specifically designed for that domain. It may be that this is close to the maximum proportion of variance that can be accounted for by any one model, at least in the present state of knowledge.

RC is also an effective metric in reasoning and cognitive development, mathematics education, and industrial complexity, as reviewed in the Introduction. RC is similar to metrics used successfully in implicit learning, including serial reaction time (Stadtler, 1992) and artificial grammar (van den Bos & Poletiek, 2008). A common metric applicable to many domains adds to the parsimony of cognitive theories. First, it enables equivalences between tasks to be defined independently of procedure or content, as Andrews and Halford (2002) have shown for six domains in cognitive development. It also provides a rational basis for rank ordering tasks according to their cognitive complexity. The ability to determine complexities independently of content has yielded many predictions that go well beyond intuition (Halford, 1993; Halford, Andrews, Dalton, et al., 2002; Halford, Andrews, & Jensen, 2002). Historically, it has proven difficult to find a basis for assessing complexity in reasoning with both flexibility and reliability, but RC appears to do so across a wide range of domains.

Relational complexity theory is essentially a theory of mental models (Goodwin & Johnson-Laird, 2005; Halford, 1993), but it conceptualises models in a different way than have previous mental model theories of reasoning. According to the current formulation, number of MMs is not required to assess complexity. Instead, complexity can be accounted for by number of categories that are related in a representation of the combined premises, where the representation is progressively elaborated by verbal reasoning or memory retrieval processes. This means that the issue as to whether people really construct more than one mental model, though it has some intrinsic significance, is not critical to assessment of complexity.

The result of this analysis is that RC theory is integrated with MM theory in the context of categorical syllogisms, and produced complexity estimates that are robust. This adds to integration by Goodwin and Johnson-Laird (2005), who adopted RC in their MM account of relational reasoning, and to the formulation by Markovits and Barrouillet (2002), who showed that elaboration of MM by children was influenced by RC. Our findings also corroborate the decision to collapse two- and three-model problems (Bucciarelli & Johnson-Laird, 1999; Johnson-Laird & Byrne, 1991, 1996) because this yields better prediction. RC theory yielded five predictions about categorical syllogisms, all of which were supported.

Another finding of the study is that "no valid conclusion" appears to be used as a default response for complex problems, as a substitute for "don't know". If no allowance is made for this it can distort assessments of performance, because it leads to a large number of false positive responses on NVC problems. Future studies could investigate factors that influence this tendency, such as mental ability. The default to NVC might be associated with more modest intellectual ability, because it obviates the need for the complex reasoning required to determine that there is no valid conclusion. Nevertheless, given that 37 of the 64 syllogisms have no valid conclusion, this tendency is a serious distortion. Our data also show that default to NVC occurs more with Figures 3 and 4 problems for which combined atmosphere and figural heuristics cannot be used. The finding that the same effect occurs for NVC problems where the heuristics cannot yield correct answers supports Principle 2 that participants tend to use simplifying heuristics without knowledge of whether they yield correct answers.

Our conclusion is that complexity as measured by the RC metric is a factor in syllogistic reasoning. Reasoning is subject to capacity limitations, but we have been able to identify highly effective strategies for reducing complexity, thereby avoiding overload. This study integrates the RC metric with MM theory, which potentially increases parsimony in the field. When combined with RC findings in other domains, this suggests RC is potentially a general cognitive complexity metric.

Original manuscript received April 2008
Manuscript accepted December 2008
First published online September 2009

REFERENCES

Andrews, G., Birney, D. P., & Halford, G. S. (2006). Relational processing and working memory in the comprehension of complex relative clause sentences. *Memory and Cognition*, *34*, 1325–1340.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology*, *45*, 153–219.

Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development*, *74*, 1476–1499.

Begg, I., & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology*, *81*, 351–354.

Birney, D. P., & Halford, G. S. (2002). Cognitive complexity of suppositional reasoning: An application of the relational complexity metric to the knight-knave task. *Thinking and Reasoning*, *8*, 109–134.

Birney, D. P., Halford, G. S., & Andrews, G. (2006). The influence of relational complexity on reasoning: The development of the Latin Square Task. *Educational and Psychological Measurement*, *66*, 146–171.

Boag, C., Neal, A., Loft, S., & Halford, G. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics*, *14*, 1508–1526.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, *23*, 247–303.

English, L. D., & Halford, G. S. (1995). *Mathematics education: Models and processes.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Espino, O., Santamaria, C., & Garcia-Madruga, J. A. (2000). Activation of end-terms in syllogistic reasoning. *Thinking and Reasoning*, *6*, 67–89.

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review*, *13*(3), 378–395.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates, Inc.

Gibson, E. J. (2000). Commentary on perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*, 43–48.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*, 468–493.

Halford, G. S. (1993). *Children's understanding: The development of mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Halford, G. S. (2009) Complexity provides a better explanation than probability for confidence in syllogistic inferences. *Behavioral and Brain Sciences*, *32*, 91–91.

Halford, G. S., & Andrews, G. (2004). The development of deductive reasoning: How important is complexity? *Thinking and Reasoning*, *10*, 123–145.

Halford, G. S., & Andrews, G. (2006). Reasoning and problem solving. In D. Kuhn & R. S. Siegler (Eds.),, *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 557–608). Hoboken, NJ: Wiley.

Halford, G. S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002). Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology*, *81*, 417–445.

Halford, G. S., Andrews, G., & Jensen, I. (2002). Integration of category induction and hierarchical classification: One paradigm at two levels of complexity. *Journal of Cognition and Development*, *3*, 143–177.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science*, *16*(1), 70–76.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, *11*, 236–242.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*, 803–831.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child.* London: Routledge & Kegan Paul.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Byrne, R. M. J. (1996). Authors' response: Mental models and syllogisms. *Behavioral and Brain Sciences*, *19*, 543–546.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, *10*, 64–99.

Markovits, H., & Barrouillet, P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review*, *22*, 5–36.

Maybery, M. T., Bain, J. D., & Halford, G. S. (1986). Information processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 600–613.

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

Newstead, S. E., Thompson, V. A., & Handley, S. J. (2002). Generating alternatives: A key component in human reasoning? *Memory and Cognition*, *30*, 129–137.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Oberauer, K., Sub, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641–652.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, *102*, 533–566.

Roberts, M. J. (2005). Expanding the universe of categorical syllogisms: A challenge for reasoning researchers. *Behavior Research Methods*, *37*(4), 560–580.

Stadtler, M. A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 318–327.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, *19*, 97–140.

Stenning, K., & Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology*, *34*, 109–159.

Stupple, E. J. N., & Ball, L. J. (2007). Figural effects in a syllogistic evaluation paradigm: An inspection time analysis. *Experimental Psychology*, *54*(2), 120–127.

Van den Bos, E., & Poletiek, F. H. (2008). Effects of grammar complexity on artificial grammar learning. *Memory and Cognition*, *36*(6), 1122–1131.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, *18*, 451–460.