Remembering connectives: Visual-auditory consonance influences representations in time

Bradley J. Morris (morrisb@gvsu.edu)
Department of Psychology, Grand Valley State University
Allendale, MI 49401 USA

Rebecca Neiman (niemanre@mail.gvsu.edu)
Department of Psychology, Grand Valley State University
Allendale, MI 49401 USA

Abstract

Reasoning errors often arise when information is represented incorrectly. One example is the "conjunction effect", a tendency to misremember OR statements as AND statements. We investigated the how the time course of consonance between visual and auditory information influenced memory for connectives. We measured memory for statements after manipulating the consonance between language and visual information. The results indicated that subjects correctly recalled more statements given language-visual consonance and were more likely to false alarm (e.g., incorrectly recalling OR statements as AND) given language-visual dissonance. We modeled performance using a Simple Recurrent Network. Our model, in which the training set was structured similarly to natural language input, provided a reasonable analog. Taken together, the results suggest that connective representations are influenced by the concordance between visual and language input and that the bias toward conjunctions arises, in part, from the relatively high frequency of conjunctions in visual and linguistic input.

Keywords: Representation, logical connectives, neural network, memory

Reasoning errors often arise from defective representations of natural language (Johnson-Laird, 2003). One influential theory of language comprehension is that people create semantic-situational models of information during comprehension (Johnson-Laird, 2003; Zwaan & Radavansky, 1998). Mental models are cognitive simulations of objects and relations that preserve the structure of their input, specifically, meaning and the situations in which information occurs (e.g., spatial relations between objects; Johnson-Laird, 2003). For example, during comprehension, readers often integrate images and texts into a single model (Gernsbacher, 1990) and represent descriptions of space visually (Spivey, Tyler, Eberhard, & Tanenhaus, 2000).

Inferences occur by searching and evaluating models within a bounded working memory space. There is considerable evidence that reasoners use models, rather than formal rules (Rips, 1994), for deductive reasoning. For example, subjects forget the form of arguments rapidly during problem solving (Johnson-Laird & Stevenson, 1970) yet maintain the "gist" of the information (i.e., retain meaning but forget connectives; Johnson-Laird, 2003). The tendency to retain the "gist" of information reduces working memory burden but has the potential to distort inferences derived from these representations.

One such a distortion is the tendency to misremember statements as conjunctions, or the *conjunction effect* (Rader & Sloutsky, 2001). For example, given the statement *There is a star or a pencil in the box* people often recall this statement as *There is a star and a pencil in the box*. This small change yields large effects on drawing valid or true conditions because the conditions for truth-falsity differ for AND and OR statements. These results are clearly incompatible with current accounts in which representation is based on syntactic form (e.g., Rips, 1994) because semantic information was recalled at significantly higher levels than form (i.e., connective).

Although the effect is clear, there is no consensus on why the conjunction effect occurs. Although formal rules cannot account for this effect, current accounts of mental model theory are also insufficient to account for these effects. One factor that likely plays a role is the complexity of statements, defined as the number of models required to produce a veridical representation. Johnson-Laird (2003) suggests that conjunctions are simpler because they require only one model while other forms require more than one. Although there is evidence to support this suggestion (Feldman, 2000; Morris & Hasson, 2010), it focuses on the complexity of evaluating models rather than the cause of creating defective representations (e.g., models). We examine two possibilities: the influence of visual and auditory information and the time course of model construction.

Integrating visual and auditory information in time

As discussed above, models are mental simulations of information derived from sensation and language. That is, models are constructed from information derived from multiple modalities (e.g., visual and auditory information). Evidence form previous research demonstrates that visual and auditory information influence the types of comprehension models created (Gernsbacher, 1990). Spivey et al. (2000) demonstrated that language comprehension influenced visual perception of complex scenes.

Specifically, that language constrained the search of visual space by focusing attention on specific targets (and ignoring non-relevant information). It has been well established that the dual coding of visual and auditory information increases recall of information (Paivio, 1969). Conversely, mismatches between visual and auditory information can impede recall, as in the stroop effect (Goldfarb & Treisman, 2010). Thus, the concordance between visual and auditory information may influence representations (Spivey et al., 2000). Information from visual and auditory sources must be integrated during comprehension and one question is how comprehension and memory is influenced by the time course of integrating this information.

Models are created online as information becomes available (Zwaan & Radavansky, 1998). The time course of availability may influence comprehension in that information from one modality may be present before information from another modality. In natural language comprehension information is presented sequentially (i.e., one word at a time), not simultaneously. Visual information may be simultaneously available (during the entire language presentation), may be sequentially available (may become available as the language information becomes available), or may be absent. Take the following example illustrated in Figure 1:

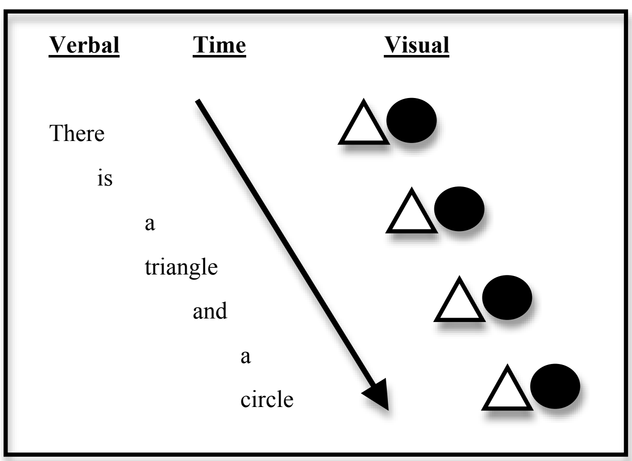

Figure 1. An example of the time course of verbal and visual information

a parent is verbally describing a visual scene to a young child. In this scene, the child can see a triangle and a circle. The parent describes this scene with the statement "There is a triangle and a circle". The visual information, two objects, form a visual conjunction that is immediately available in working memory. The verbal description is added word-by-word in a sequential fashion but available later in working memory. Importantly, the visual conjunction is in concordance with the verbal conjunction, that is, the verbal description fits the visual information. Thus, regardless whether each modality creates a separate model or whether they contribute to multiple components of the same model, each modality contributes consistent information.

These different sources of information may be available at different times during model construction. Because visual information can be presented simultaneously, it is likely that, in static scenes, this information is present before verbal information that is presented sequentially. The timing of information may be related to the likelihood of creating specific types of models. For example, seeing a visual conjunction may be a cue that the verbal statement will be phrased as a conjunction. Thus, it seems likely that time course is involved with model creation on two levels: (1) Visual information is likely to be available before information in language, providing an advantage for visual information in the sequence of model creation and (2) the presentation of visual information cues expectations about the accompanying language description, specifically, that this description should be consistent with the visual scene.

Visual information may cue likely comprehension models. If so, then the time course for visual and auditory information influences model creation. For example, an AND statement might be associated with the simultaneous presentation of the objects being named. If this is the case, then the type of visual simulation created might influence how connectives are comprehended and remembered. For example, if a person sees two objects at the same time (e.g., a ball and a box sitting on a table), a veridical model would contain a "visual conjunction" (i.e., both objects present in the visual set) that is consistent with the conjunction in language.

A disjunction that names the same visual objects would yield a "visual conjunction" which would be different from the verbal disjunction. In this case, such a mismatch might influence representation in that the statement might be misremembered as a conjunction (consistent with the visual model) rather than a disjunction (consistent with the verbal model). If, however, the objects were not presented at the same time, e.g., shown sequentially such that the objects were never visible together, this may result in a different visual model than the visual conjunction suggested earlier. Such a "visual disjunction" might be aligned with a verbal disjunction but may interfere with a verbal conjunction.

In this way, the mapping between the visual and language information may influence the types of models created. For example, the simultaneous presentation of visual information may suggest a verbal description of a conjunction. Such information may explain the tendency for people to misremember disjunctions as conjunctions (Rader & Sloutsky, 2001). If a verbal disjunction is paired with visual conjunction, it may result in interference leading to erroneously recalling a disjunction as a conjunction. Thus, semantic-situational accounts would predict differences between conditions in that simultaneous visual presentation should be associated with an increased tendency to recall as a verbal conjunction while a sequential presentation should be associated with a decreased tendency to recall as a conjunction. If the difference is complexity only (i.e., the number of models), then performance should be influenced only by connective and differences in visual information should not influence performance.

Method

Subjects. Ninety-eight undergraduates were given course credit for participating in the experiment. Each subject was randomly assigned to one of five conditions.

Design and Materials. Participants were presented with statements phrased with either AND, OR, or IF. There were five presentation conditions: control (no visual information; N = 17), ALL SIMULTANEOUS (all visual information was presented together; N = 22), ALL SEQUENTIAL (visual information was presented separately for an equal amount of time; N = 23) AND SIMULTANEOUS (AND statements were presented with simultaneous visual information and OR statements were presented with sequential visual information; N = 19), and AND SEQUENTIAL (AND statements were presented with sequential visual information and OR statements were presented with simultaneous visual information; N = 17). The materials were 45 statements: 15 conjunctions, 15 disjunctions, and 15 conditionals taken from Rader & Sloutsky (2001; Experiment 2). The statements described a hypothetical person (phrased as “This person…”). Pictures were chosen to illustrate each proposition within the statement. For example, the statement, “This person trains dolphins and bakes bread” was associated with a picture of a dolphin and a picture of a loaf of bread being taken out of an oven. As in Rader & Sloutsky (2001), the recognition materials were 225 descriptions consisting of five types: (1) actual statements, (2) different-connective 1 and 2 (e.g., AND statement presented as an OR statement; AND statement presented as an IF statement), (3) different-noun (e.g., *trains dolphins* presented as *trains seals*), and (4) non-logical connectives e.g., AND statement presented as a BUT statement). These statements were presented individually using Superlab presentation software.

Procedure. Participants were seated at a computer. Following Rader & Sloutsky (2001), subjects were given instructions to remember the statements exactly as they were presented because they would be tested on their memory for these items immediately following the learning phase. The learning phase consisted of presentation of information in a series of PowerPoint slides with voice over narration. In the simultaneous condition, both items were presented on screen for 6 seconds. In the sequential condition, the first item was presented for 6 seconds then removed and the second item was presented for 6 seconds and then removed (i.e., items were never shown on screen at the same time). In a pilot study there was no difference in the sequential condition whether each item was presented for 6 (same item presentation time) or 3 seconds each (same total presentation time).

Recall Test: Once the presentation trials were completed, participants were given a series of statements (in Superlab) and asked to determine whether they had seen the statements or not. Subjects also saw 4 instruction slides (e.g., “press the BLUE button”) in order to control for random responding. Subjects began the recognition portion immediately following the presentation of the statements. Students were instructed to press the BLUE button (the “L” key with a BLUE sticker) if they had seen the exact statement in part 1 and to press the ORANGE button (the “A” key with an ORANGE sticker) if they had no seen the exact statement in part 1. Subjects saw a fixation slide (+ in the center of the screen) for 500 MS statement before each slide. Each subject received a randomized order of the 225 recognition statements. A hit was defined as correctly identifying a statement presented in the learning phase. A false alarm was defined as incorrectly identifying a statement as presented in the learning phase.

Predictions. If statements are represented via syntax, then none of the visual conditions should significantly influence recall. If statement complexity is the only factor, then conjunctions should be recalled correctly more frequently than disjunctions and no differences should arise between visual conditions. If representations contain information from different modalities, then different visual presentation should influence recall. If visual information influences recall, then seeing visual conjunctions should increase correct recall for AND statements and increase False Alarms for OR statements. If visual information is presented sequentially, then this should increase correct recall for OR statements and increase False Alarms for AND statements.

Results and Discussion

Recall that subjects were presented 45 statements in the training session. We will discuss two sets of results: the proportion of statements that were correctly recognized (hits) and the number of incorrect recognition responses (false alarms). *IF* recognition rates, non-logical connectives, and different-noun statements were at correctly identified at ceiling across conditions (.95-.97) and will not be included in subsequent analyses. Hit rates for AND and OR statements differed significantly across conditions ($F(4, 94) = 3.7$, $p = .008$; see Figure 2). Bonferroni adjusted ($p < .05$) post hoc tests indicated that AND statements were correctly recalled more frequently than when presented with simultaneous visual information and OR statements were correctly recalled more frequently when presented with sequential visual information.

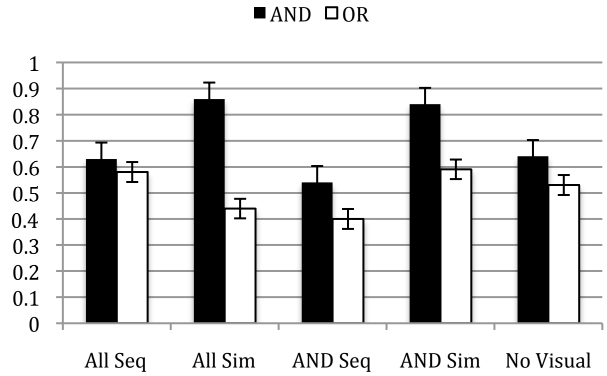

Figure 2: Correct recognition rates for AND and OR statements

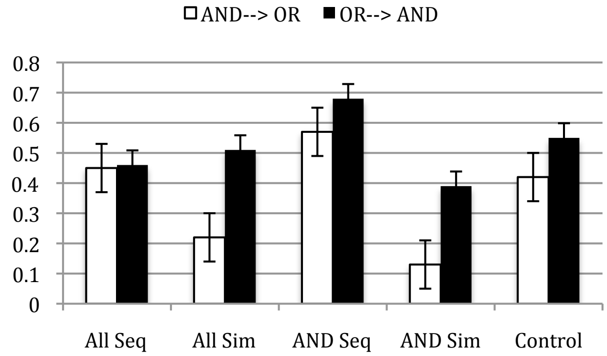

Figure 3: False alarm rates for AND and OR statements

False alarm rates will be discussed by connective. Recall that false alarms occurred when subjects indicated incorrectly that they had seen a statement during the training session. False alarm rates between conditions differed significantly for statements initially presented as AND ($F (4, 94) = 5.3$, $p = .001$; see Figure 3). Bonferroni adjusted ($p < .05$) post hoc tests indicated that false alarms were more frequent when AND statements were presented with sequential visual information. When compared to the control condition in which there was no visual information, subjects were more likely to false alarm statements when presented with sequential visual information and less likely to false alarm when given simultaneous visual information.

False alarm rates between conditions differed significantly for statements initially presented as OR ($F (4, 94) = 3.8$, $p = .007$; see Figure 3) Bonferroni adjusted ($p < .05$) post hoc tests indicated that false alarms were more frequent when OR statements were presented with simultaneous visual information. When compared to the control condition, false alarms were more likely with simultaneous visual information and less likely with sequential information.

The results demonstrate that auditory and visual information influence connective recall. The results provide strong evidence against syntactic and complexity accounts. The findings suggest that subjects created representations using both visual and auditory information. Correct recall rates were higher when visual and auditory were consonant. False alarm rates were higher when visual and auditory information was dissonant. Performance for each connective was related to the type of information with which it was consistent. For example, simultaneous visual information improved recall for descriptions that were conjunctions and increased false alarms for descriptions that were disjunctions (and vice versa). Although these data clearly demonstrate that auditory and visual information influence representations of connectives, the data do not provide information about the origins of a "default" representational format cues by visual information.

Computational Model

The experimental data demonstrate that visual information influences memory for connectives. In the introduction, we discussed possibilities for how visual information might cue model creation. We suggested two ways in which visual information cues likely models: (1) visual information is available before auditory information and (2) a preference for consistency between visual and auditory information cued by the visual information. But how might these cues arise? One possibility is that simply having access to visual information first during model creation will result in a heavier "weighting" for visual cues relative to auditory information. A second, related possibility is that these cues arise from the time sequence and from probabilities of consonant visual mappings with descriptions in natural language (a suggestion made by Rader & Sloutsky, 2001).

We created a neural network in order to investigate factors that may contribute to this bias. Because of the emphasis of time-based information, we created a Simple Recurrent Network (SRN; Elman, 1990) in MatLab in order to model the human data. A SRN is divided into four layers, an input layer, a hidden layer, a context layer, and an output layer. SRNs are unique in two ways. Architecturally, SRNs are unique in that every node in the context layer is connected to every unit in the input and hidden layers. In terms of processing, SRNs are unique in that input occurs sequentially. For example, when the first unit is entered this information is processed in the hidden and context layers. The second input step is processed in the hidden layers and the information from the previous step in the context layer is also provided to the hidden layers. This continues so that each new input is processed in concert with information from previous input (via the context layer).

Our SRN contained an input layer (24 nodes each) to correspond to visual and auditory information (separate semantic and connective coding), 36 hidden and 36 context nodes, and 24 output nodes. Because we modeled statements with connectives, the input strings were divided into the presence/absence of five different objects and two different connectives. For example, [100000101000] would represent semantic object 1 [100000], semantic object 2 [010000], and the connective AND [1 seventh place]. Visual information was either absent (all zeros), sequential (1111111000000), or simultaneous (10101010101010). Inputs were entered sequentially such that the information from the second input unit is entered into the hidden layer with the information from the first unit via the context layer. The model used a gradient decent method as a learning rule and mean squared error between target and predicted values as the minimization criterion.

Model 1. The first model was trained on using a set of 40 statements, 20 conjunctions 20 disjunctions, for each half of the statements were given consonant and half dissonant visual input. The network was trained for 5000 epochs on the training set. The test set consisted of 30 items in which the visual information was not included: 10 items from in the training set and 20 not in the training set (10 different-connective and 10 different-semantic). The best training performance occurred at epoch 280. The network correctly identified 80% of both AND and OR items in the test set. This result was very different from the experimental results, specifically, the error rates were lower than the experimental results and there were no differences between connectives. One possible explanation for the result is that because the network extracted the structure from the training set, the training set may be dissimilar to the “training set” in natural language.

Model 2. A new training set was created for model 2. This training set was based on the relative frequencies of each connective and consonant visual information. Morris (2008) reported that parents use AND statements approximately 12 time more frequently than OR statements to children between ages 2 and 5. Data from the British National Corpus (2011) contained 7.4 times as many AND statements than OR statements. Based on these data, we set a conservative estimate of relative frequency in which AND statement occurred 6 times more frequently than OR statements. The frequency of visual and auditory concurrence is less well defined, however, Harris, Jones & Grant (1982) found that children were not attending to visual objects during approximately 50% of labeling events. Based on these data, we set a conservative estimate of visual-auditory consonance for the training set: 50% of items without any visual information, 40% visual, simultaneous and 10% visual, sequential.

Using the new training set, Model 2 was trained for 5000 epochs and tested using the same test set used in Model 1. The results were quite different from the previous model. Best training performance occurred at epoch 2817. The network correctly identified 70 % of AND but only 30% of OR items. The network false alarmed on 50% of OR items, and 80% when presented with simultaneous visual information. The network false alarmed on only 25% of AND items, all but one false alarm was associated with sequential visual input. The results are much more similar to the experimental results than the first model and suggest that natural language frequency plays a role in the origin of the conjunction bias.

General Discussion

These results suggest that language and visual information are integrated over time into connective representations. Our experiment demonstrated that subjects were more likely to recall OR statements as AND statements when they saw simultaneous visual information (i.e., visual conjunctions). This tendency was reduced when subjects saw sequential visual information with verbal disjunctions (and vice versa). In general, false alarm rates increased when visual and auditory information was not consonant. More specifically, when an OR statement was presented with simultaneous visual information or when an AND statement was presented with sequential visual information. This result is similar to Spivey et al. (2000) in that the integration of language and visual information in comprehension is bi-directional. We investigated one possibility for why visual information cues likely models: the relative frequency of the connectives in natural language. The results of Model 1 demonstrated that, given roughly equal input, the model produced no conjunction effect. Once the training set was changed to reflect baseline occurrences of both connectives in natural language and their occurrence with visual information, the model produced effects that were much closer to the human data.

These results are consistent with a model account in which connectives are “simulations” that are structured like the information in the environment. In the case of visual conjunctions (i.e., two objects occurring simultaneously), when simulated in memory, this structure is likely to be recalled as a language-based conjunction. Because adults tend to encode the “gist” of information (rather than verbatim information), the specific connective is likely to be lost in the representation. Finally, AND statements are more likely to occur in natural language, weighting AND as a more likely connective. Thus, given visual conjunction and no specific connective, recalling as AND may be more likely. The results also suggest that reasoning from models derived from consonant visual-verbal information may result in different conclusions than reasoning from models derived from dissonant visual-verbal information.

References

British National Corpus. (2011).

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14* (2), 179-211.

Feldman, J. (2000) Minimization of Boolean complexity in human concept learning. *Nature, 407*, 630–633.

Gernsbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Earlbaum.

Goldfarb, L. & Treisman, A. (2010). Are some features easier to bind than others? The congruency effect. *Psychological Science, 21*(5), 676-681.

Harris, M., Jones, D., & Grant, J. (1983). The nonverbal content of mothers' speech to infants. *First Language, 4,* 21-31.

Johnson-Laird, P.N. (2003). Mental models and reasoning. In Leighton, J.P., & Sternberg, R.J. (Eds.) *The Nature of Reasoning.* Cambridge: Cambridge University Press. Pp.169-204.

Johnson-Laird, P.N. & Stevenson, R. (1970). Memory for syntax. *Nature, 227*, 412.

Morris, B. J. (2008). Logically speaking: Evidence for Item-based acquisition of the connectives AND & OR. *Journal of Cognition and Development, 9*, 67–88.

Morris, B. J., & Hasson, U. (2010). Multiple sources of competence underlying the comprehension of inconsistencies: A developmental investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36* (2), 277-287.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76*(3), 241-263.

Rader, A., & Sloutsky, V. (2001). Conjunctive bias in memory representations of logical connectives. *Memory and Cognition, 29* (6), 838-849.

Rips, L. (1994). *The psychology of proof: Deduction in human thinking.* Cambridge, MA: MIT Press.

Zwaan, R.A. & Radavansky, G.A. (1998). Situation models in language and memory. *Psychological Bulletin, 123*, 162-185.