

Independent Representation of Abstract Arguments and Relations

Derek Devnich (devnich@ucla.edu)

Greg T. Stevens (stevens@psych.ucla.edu)

John E. Hummel (jhummel@psych.ucla.edu)

Department of Psychology, 1285 Franz Hall

University of California, Los Angeles

Los Angeles, CA 90095

Abstract

Propositions specifying properties of, or relations among, one or more arguments form a central part of human mental representations. Representing a proposition entails binding each relational role to its argument. At the same time, computational considerations suggest that roles and arguments should be represented independently of one another in working memory (WM). We report an experiment using General Recognition Theory (Ashby & Townsend, 1986) to test the independence of relational roles from their arguments in WM. The results suggest that roles and arguments are independent in WM.

Computational Perspectives on Representation

Working memory (WM) tasks require a person to hold novel objects or relations in mind, rearranging them according to the demands of the task (Cowan, 2000; Jonides, 1995). One example of a common WM task is thinking about a proposition for the purpose of encoding it into memory or reasoning about it. In order to represent a proposition it is necessary to bind the arguments of the relation to their relational roles. For instance, the proposition *owns* (Bill, car), stating that Bill owns a car, requires binding Bill to *owner* and car to *owned*. Failing to bind the roles to their arguments would make it impossible to distinguish *owns* (Bill, car) from *owns* (car, Bill). A common approach to binding in the connectionist literature is *conjunctive coding*; that is, designating separate units for separate role bindings (e.g., Halford et al., 1994; O'Reilly, Busby, and Soto, in press; Smolensky, 1990). For example, one unit or set of units might represent *Bill as owner*, a second set would represent *car as owned*, a third would represent *car as owner*, and a fourth *Bill as owned*. The proposition *owns* (Bill, car) would be represented as activity on the first two sets of units, whereas *owns* (car, Bill) would be represented as activity on the latter two. Although conjunctive coding is adequate (even necessary; see Hummel & Holyoak, 1997; O'Reilly & Rudy, 2001) as a basis for representing relatively permanent bindings in long-term memory (LTM), it is a poor choice for representing temporary bindings of roles and arguments in WM. In particular, conjunctive coding violates role-argument independence, making it inadequate as a basis for relational generalization (Holyoak & Hummel, 2000; Hummel & Holyoak, 1992, 1997, 2003). Computational considerations suggest instead that roles and arguments must be

represented independently in WM and bound together dynamically.

One commonly proposed basis for dynamic binding in working memory is synchrony of neural firing (e.g., Hummel & Holyoak, 1992, 1997; Shastri & Ajjanagadde, 1993; Sougné and French, 1997). These proposals suggest that units (e.g., neurons) representing relational roles fire *in* synchrony with neurons representing their arguments, and *out* of synchrony with other role-argument bindings. As a representational mechanism, neural synchrony has several purported flaws (see, e.g., O'Reilly, et al., in press). First, synchrony is transient in the absence of maintenance. Second, the need for such maintenance makes synchrony appear too fragile to be widely used in the brain. Finally, any knowledge represented by synchronous firing would have to interact eventually with other (presumably conjunctive) representations in the brain. Given these considerations, and the clear need for conjunctive coding for storage in LTM, postulating synchrony as a binding mechanism in addition to conjunctive coding appears less than parsimonious. These critiques are well-founded with respect to the requirements of long-term memory, which must be permanent and relatively robust. Ironically, they are also an excellent description of the properties of working memory: transient, fragile, low-capacity, and needing to be integrated with long-term memory (Jonides, 1995). In short, the representational requirements of working memory appear to be dramatically different from those of long-term memory (Hummel & Holyoak, 1997).

These theoretical arguments are all well and good, but it remains an open empirical question as to whether roles are represented independently of their arguments in WM. We report an experiment using the complete identification paradigm (Ashby & Townsend, 1986) to investigate whether relational roles and arguments are represented independently in WM. If they are, then inasmuch as people are capable of knowing which roles are bound to which arguments (which they clearly are), this would strongly suggest that roles and arguments are bound together dynamically (which is not to say that the dynamic binding is necessarily done by synchrony; see Hummel & Holyoak, 1997). By contrast, if roles do not appear to be independent of their arguments in WM, this would suggest that conjunctive coding is adequate for both LTM and WM.

Assessing Independence: General Recognition Theory and the Complete Identification Paradigm

Ashby and Townsend (1986) developed the General Recognition Theory (GRT), a multidimensional generalization of signal detection theory, to provide a framework for assessing environmental and processing dependencies in perception and action. GRT proposes a minimal processing model of representation, consisting of input channels, perceptual processes, and decisional processes. The authors assume that the input channels for different perceptual dimensions do not overlap; hence, when signals in the input channels covary, this indicates an environmental correlation between the relevant dimensions. Input channels thus faithfully mirror the statistical properties of the world; when dimensions are uncorrelated in the environment, the representational system shows *perceptual independence*. GRT assumes that perceptual processes map these inputs onto a multidimensional perceptual space, and that decisional processes associate different regions of this space with the appropriate response (Maddox & Ashby, 1996). If the perceptual representation of one dimension does not depend on the perceptual representation of the other, the representational system shows *perceptual separability*. Similarly, if the subject's decision about the level of one dimension does not depend on the level of the other, the representational system shows *decisional separability*. The two forms of separability cannot be disambiguated empirically; one can only demonstrate that two dimensions are fully separable, or that they are integral at some unknown point in processing.

The purpose of GRT is to provide the most general possible account of human performance with respect to perception and action. The theory makes the minimum number of assumptions about processing mechanisms necessary to justify its analyses of the empirical data. It does not claim to support any particular algorithmic theory of representation, and its assumptions under-specify any such theory. However, given empirical data that imply separability or integrality of the relevant dimensions, one can constrain the range of possible mechanisms that could produce such a pattern of behavior (complete integrality occurs when each level of the first dimension has a preferred level of the second dimension at which it is processed most efficiently. The respective levels are permanently bound and cannot be dissociated during performance, in spite of the subject's best efforts). Movellan and McClelland (2001) showed that separable processing implies independent representation of dimensions within the input and hidden units of any neural network architecture. Hence, a finding of separability provides strong support for independent representation. In contrast, a finding of complete integrality could either imply conjunctive representation or independent representation with cross-talk.

Ashby & Townsend (1986) describe mathematical tests for assessing independence (of the perceptual, rather than

representational sort) and separability based on confusion data from a *complete identification* task (where identification is the limit case of categorization). In this task, the subject makes unique responses to all possible combinations of two (or more) levels of two (or more) stimulus dimensions (typically four responses: two levels of each of dimensions). For example, stimuli might be red and blue circles and squares: A red square would get one response, a blue square a second response, a red circle a third response, and a blue circle a fourth. The dimensions are perceptually independent if their effects are statistically independent:

$$P(a_2b_2|A_iB_j) = [P(a_2b_1|A_iB_j) + P(a_2b_2|A_iB_j)] * [P(a_1b_1|A_iB_j) + P(a_1b_2|A_iB_j)],$$

where A_i and B_j are the values of dimensions A and B in the world, and a_i and b_j are the values identified by the subject. The dimensions are fully separable (both perceptually and decisionally) if responses for any given level of one dimension do not depend on the level of the other:

$$\text{for } i = 1, 2 \\ P(a_1b_1|A_iB_1) + P(a_1b_2|A_iB_1) = P(a_1b_1|A_iB_2) + P(a_1b_2|A_iB_2);$$

$$\text{for } j = 1, 2 \\ P(a_1b_1|A_1B_j) + P(a_2b_1|A_1B_j) = P(a_1b_1|A_2B_j) + P(a_2b_1|A_2B_j).$$

The graphical representation of this test for *marginal response invariance* can be seen in Figure 1. Each combination of levels of the dimensions is projected onto the representational space as a distribution of possible percepts (when stimulus dimensions are uncorrelated in the environment, these distributions are symmetrical; we will assume this for the sake of simplicity). When stimulus dimensions are completely separable, the perceptual distributions are positioned equidistant from one other along both coordinate axes. The subject's decision bounds are perpendicular to each other and to the coordinate axes. There are several ways in which dimensions can violate separability. These include the decision bounds not being perpendicular to the coordinate axes, the representational distributions being unevenly asymmetric, and the representational distributions not being equidistant in the representational space. All of these violations will result in some change to the decision bounds. For our purposes, the most interesting violation is the case in which the decision bounds are no longer perpendicular to each other or to the coordinate axes (complete integrality: see Figure 1(b)).

One can conduct the test for marginal response invariance using accuracy rates or response times (Ashby & Maddox, 1994; Maddox & Ashby, 1996; Thomas, 1996). Both of these are measures of confusability; less confusable stimuli are processed in less time with fewer errors. GRT asserts that percepts become less confusable as distance to the decision bounds increases; hence, both of these measures are proxies for the distance of the perceptual distribution

from the decision bounds, and can be used to derive the underlying representational space. For instance, in Figure 1(b), subjects would respond quickly and accurately to stimuli A_2B_1 and A_1B_2 , while making slow, error-prone responses to A_1B_1 and A_2B_2 . The choice of measure depends on experimental design: Accuracy is appropriate for highly-confusable stimuli and unlimited trial time, while reaction time is appropriate for less confusable stimuli in a speeded-classification paradigm. Our experiment is a speeded-classification task, and uses reaction time as the relevant measure of performance.

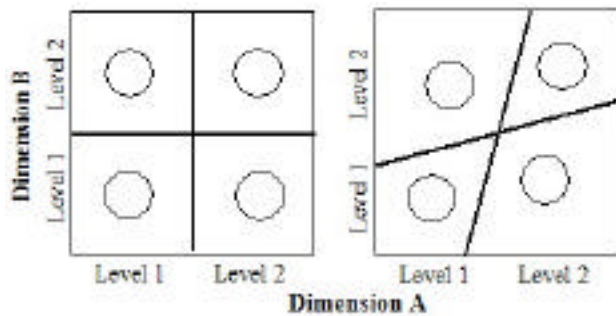


Figure 1: Hypothetical representation of two dimensions displaying either (a) complete separability, or (b) complete integrality.

Processing Abstract Arguments and Relations

Although Ashby and Townsend (1986) designed GRT and the complete identification paradigm as a tool for investigating the representation of perceptual dimensions, we are adapting it for the purpose of investigating the independent representation (or lack thereof) of abstract roles and arguments presented verbally. In the most general sense, the two dimensions were the role type and argument type of one of the arguments (the *target argument*) given in a sentence. More specifically, the relation was a *Power* relation with role values of *dominant* and *subordinate*. The target argument was a *Creature Type* with values of *animal* and *human*. For example, in the sentence “The man admired the elephant,” the target argument is the elephant (the target argument was always whichever argument was not “man”) and its value is “animal”; the relation is admired, and its value is “object dominant” (see Table 1). Each subject performed the complete categorization experiment, with unique responses for all possible combinations of the levels of each dimension. Responses were compared for response time differences. We also examined accuracy data to ensure that results were not a product of speed-accuracy trade-off.

Method and Materials

Stimuli Subjects responded to sentence stimuli in the form, “The <subject> <verb> the <object>.” One of the nouns was designated as the target, and the subjects’ goal was to classify that noun. Stimuli were constructed from 30 nouns (15 animals and 15 humans) and 30 verbs (15 subject-

dominant and 15 object-dominant), such that the total stimulus set contained an equal number of animal and human targets, all equally likely to be dominant or subordinate, appearing in the subject or object position with equal frequency (see Table 1). The total stimulus set contained 1800 sentences. All verbs were conjugated in the simple past tense.

Table 1. Relations and arguments

<i>humans</i>	<i>animals</i>	<i>subject dominant / object subordinate</i>	<i>subject subordinate / object dominant</i>
scientist	kangaroo	commanded	venerated
mechanic	elephant	defeated	respected
musician	iguana	protected	escaped
engineer	antelope	oppressed	adored
attorney	wildebeest	employed	admired
janitor	hyena	punished	worshiped
writer	raccoon	chastised	revered
student	giraffe	attacked	dreaded
plumber	parrot	judged	envied
doctor	rabbit	chased	obeyed
athlete	turtle	taught	heeded
artist	eagle	blocked	served
actor	lizard	pushed	feared
cook	bear	kicked	begged
thief	crow	hit	fled

Design Each subject received two practice blocks and ten experimental blocks of 62 trials each. Each subject saw 600 of the 1800 total possible stimuli; these stimuli were selected according to a latin square design, such that every subject was exposed to every verb and every noun, and every three subjects were exposed to the complete stimulus set. A subject saw each stimulus once within a given block, and at most twice in the course of the experiment. Each block contained an approximately equal proportion of all stimulus types. Stimuli were presented randomly without replacement within blocks. Subjects classified the target by pressing one of four keys (A, D, L, and ‘); keys were assigned to responses according to three successive latin squares (one latin square for every four subjects).

Participants The participants were 35 native English speakers enrolled in an introductory Psychology class at UCLA. They received course credit for their participation.

Procedure Prior to the experiment, subjects participated in a brief paper-and-pencil training exercise to ensure that they understood the instructions for the experiment. They were instructed to classify the target nouns in 16 sentences (taken from the experimental stimuli) as animal or human, and as

dominant or subordinate. “Dominant” was defined as “the creature which has more power in the situation described by the sentence,” while “subordinate” was defined as “the creature which has less power in the situation described by the sentence.” Experimental stimuli were presented on a PC computer using Superlab software. Subjects were instructed to classify stimuli by pressing one of four keys (A, D, L, and ‘), with each key assigned to a different combination of categories (dominant animal [DA], dominant human [DH], subordinate animal [SA], and subordinate human [SH]). Subjects were encouraged to respond as quickly as possible without making mistakes. Each stimulus presentation consisted of a fixation cross (0.5 seconds), the stimulus itself (3.5 seconds), and visual feedback (0.5 seconds). Subjects received one of three visual feedback messages: “Correct,” “Wrong,” or “Too Slow”.

Results

Response Time Practice block trials and the first two trials of each block were excluded from the response time (RT) analysis. We analyzed remaining trials for separability, following Ashby & Townsend’s (1986) definition of marginal response invariance (see Table 2 for a summary of reaction time results). For two dimensions of two levels each, the analysis is comprised of four independent tests of simple effects (it is worth noting that a conventional ANOVA is not appropriate for this analysis, as the definition of marginal response invariance systematically excludes subsets of the data that would be required to perform an ANOVA properly). Subjects were 99 ms faster to correctly classify a stimulus as animal when the argument was embedded in a Dominant relation, as compared to when it was embedded in a Subordinate relation ($t(34) = 6.562$, $p < 0.00001$), and 99 ms faster to correctly classify a stimulus as human when the argument was embedded in a Dominant relation, as compared to when it was embedded in a Subordinate relation ($t(34) = 6.307$, $p < 0.00001$). In addition, subjects were 89 ms faster to correctly classify a stimulus as dominant when the role was filled by an Animal, as compared to when it was filled by a Human ($t(34) = 4.922$, $p < 0.00002$), and 91 ms faster to correctly classify a stimulus as subordinate when the role was filled by an Animal, as compared to when it was filled by a Human ($t(34) = 5.000$, $p < 0.00002$).

We repeated these analyses after dividing trials into two groups: Those in which the target appeared as the subject of the sentence, and those in which it appeared as the object. The direction of the effect remained the same for each test, regardless of the grammatical category of the target. However, the effect was always stronger when the target appeared as the subject. Subjects’ classification of animals embedded within Dominant or Subordinate roles was affected by grammatical category ($F(1,34) = 28.187$, $p < 0.00001$). Subjects’ classification of humans embedded within Dominant or Subordinate roles was affected by grammatical category ($F(1,34) = 25.025$, $p < 0.00002$). Subjects’ classification of dominant roles filled by Animals

or Humans was affected by grammatical category ($F(1,34) = 9.061$, $p < 0.005$). However, subjects’ classification of subordinate roles filled by Animals or Humans was not affected by grammatical category ($F(1,34) = 1.677$, $p < 0.204$).

Table 2. Summary of Reaction Time Data

Test	Overall RT Differences	Subject/Object Differences
respond “animal” SA – respond “animal” DA	99 ms $p < 0.00001$	179 ms $p < 0.00001$
		16 ms $p < 0.51$
respond “human” SH – respond “human” DH	99 ms $p < 0.00001$	165 ms $p < 0.00001$
		31 ms $p < 0.11$
respond “dominant” DH – respond “dominant” DA	89 ms $p < 0.00002$	120 ms $p < 0.00001$
		55 ms $p < 0.017$
respond “subordinate” SH – respond “subordinate” SA	91 ms $p < 0.00002$	107 ms $p < 0.00001$
		75 ms $p < 0.0029$

Accuracy After discarding practice blocks and the first two trials of each experimental block, the results showed that subjects were 0.47 % more likely to correctly classify a stimulus as animal when the argument was embedded in a Dominant relation, as compared to when it was embedded in a Subordinate relation ($t(34) = 1.351$, $p < 0.19$), and 0.74 % more likely to correctly classify a stimulus as human when the argument was embedded in a Dominant relation, as compared to when it was embedded in a Subordinate relation ($t(34) = 2.253$, $p < 0.031$). In addition, subjects were 0.84 % more likely to correctly classify a stimulus as dominant when the role was filled by an Animal, as compared to when it was filled by a Human ($t(34) = 1.464$, $p < 0.15$), and 1.57 % more likely to correctly classify a stimulus as subordinate when the role was filled by an Animal, as compared to when it was filled by a Human ($t(34) = 3.011$, $p < 0.005$). Although not all of these results are reliable, they do show that the reaction time data are not the result of a speed-accuracy trade-off (see Table 3 for a summary of accuracy results).

The total proportion of errors was equivalent across dominant and subordinate stimulus conditions. This suggests that dominant and subordinate stimuli were of equal difficulty. However, the total proportion of errors in response to human stimuli (9.5 %) was slightly greater than in response to animal stimuli (8.5 %) ($F(1,34) = 5.666$, $p < 0.023$), suggesting that the human stimuli were more confusable with one another than were the animal stimuli.

Table 3. Summary of Accuracy Data

Test	Overall Accuracy Differences	Subject/Object Differences
respond “animal” DA – respond “animal” SA	0.47 % p < 0.19	0.74 % p < 0.041
		0.20 % p < 0.73
respond “human” DH – respond “human” SH	0.74 % p < 0.031	0.59 % p < 0.18
		0.90 % p < 0.063
respond “dominant” DA – respond “dominant” DH	0.84 % p < 0.15	1.99 % p < 0.0016
		- 0.28 % p < 0.78
respond “subordinate” SA – respond “subordinate” SH	1.57 % p < 0.005	1.69 % p < 0.036
		1.47 % p < 0.048

Discussion

Deriving the Representational Space The overall pattern of results is this: It is easier to classify an argument that fills a Dominant role, regardless of what that argument is. It is easier to classify a role that is filled by an Animal, regardless of what that role is. We characterize these results as “separability plus response bias.” The subjects do not display a preference for binding particular roles to particular fillers; rather, their decision bounds are shifted so that Dominant stimuli and Animal stimuli encompass more of the representational space (see Figure 2(a) for a depiction of this representational space).

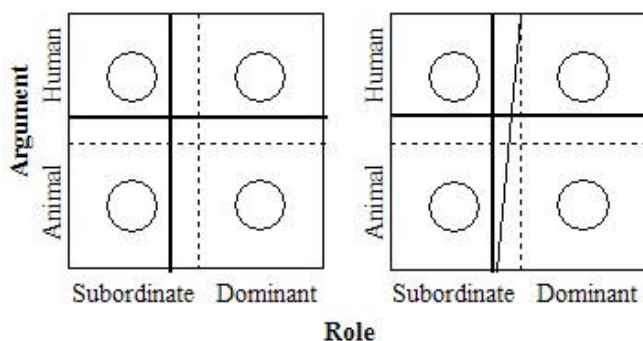


Figure 2: Representational space for the experimental dimensions. 2(a) shows separable processing plus a response bias. 2(b) shows a possible non-separable decision bound mapped onto the same space.

Recall that stimulus dimensions are separable if the decision bounds are perpendicular to each other and to the coordinate axes. To determine whether the shifted decision bounds remain perpendicular, we can test whether the deviation of the Role decision bound is equal for Animals

and Humans, and whether the deviation of the Argument decision bound is equal for Dominant and Subordinate roles. If the deviation of (for instance) the Role decision bound is not equal at the two levels of Argument, it might result in a skewed bound similar to that seen in Figure 2(b). To test whether the Role decision bound was shifted equally at the two levels of Argument, we compared the reaction time advantage for identifying animals in a Dominant role to the reaction time advantage for identifying humans in a Dominant role. There was no reliable difference between the two ($t(34) = 0.012$, $p < 0.99$). To test whether the Argument decision bound was shifted equally at the two levels of Role, we compared the reaction time advantage for identifying dominant roles filled by an Animal to the reaction time advantage for identifying subordinate roles filled by an Animal. There was no reliable difference between the two ($t(34) = 0.062$, $p < 0.95$). Thus, “separability plus response bias,” appears to be the most reasonable interpretation of our results.

Conclusions and Future Directions

Our results support the hypothesis that arguments and relational roles are represented independently in Working Memory. However, the complete categorization task is not a “pure” measure of Working Memory. The text comprehension and analysis required to perform the task certainly requires Working Memory; however, inasmuch as subjects are using their knowledge of the world in order to perform the classifications, they are also accessing long-term memory. It is quite possible that the response biases observed in our study are the result of expectations about the world encoded in long-term memory. Future research will attempt to disambiguate the contribution of the two memory systems.

Acknowledgments

The authors would like to thank Tom Wickens for discerning comments on the appropriate use of statistical tests and the members of the LISA lab for their invaluable feedback and support.

References

- Ashby, F.G. & Maddox, W.T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, 38, 423-466.
- Ashby, F.G. & Townsend, J.T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154-179.
- Browne, A. & Sun, R. (2001). Connectionist inference models. *Neural Networks*, 14, 1331-1355.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-114.
- Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist

- implications for processing capacity limitations in analogies. In K. J. Holyoak & J. A. Barnden (Eds.), *Advances in connectionist and neural computation theory, Vol. 2: Analogical connections*. Norwood, NJ: Ablex.
- Holyoak, K.J. & Hummel, J.E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich and A.B. Markman (Eds), *Cognitive dynamics: Conceptual and representational change in humans and machines*. Mahwah, NJ: Erlbaum.
- Hummel, J.E., & Holyoak, K. J. (1992). Indirect analogical mapping. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (516 - 521). Hillsdale, NJ: Erlbaum.
- Hummel, J.E. & Holyoak, K.J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104, 427-466.
- Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220-264.
- Jonides, J. (1995). Working memory and thinking. In E.E. Smith and D.N. Osherson, (Eds.), *Thinking*. Cambridge, MA: MIT Press.
- Maddox, W.T. & Ashby, F.G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 795-817.
- Movellan, J.R. & McClelland, J.L. (2001). The Morton-Massaro law of information integration: Implications for models of perception. *Psychological Review*, 108, 113-148.
- O'Reilly, R.C. & Busby, R.S. (2001). Generalizable relational binding from coarse-coded distributed representations. *Neural Information Processing Systems (NIPS)*. Conference presentation.
- O'Reilly, R.C., Busby, R. S. & Soto, R. (in press). Three forms of binding and their neural substrates: Alternatives to temporal synchrony. In A. Cleeremans (Ed.), *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford: Oxford University Press.
- O'Reilly, R.C. & Rudy, J.W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, 108, 311-345.
- Shastri, L. & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral & Brain Sciences*, 16, 417-494.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46, 159-216.
- Sougné, J. and French, R. (1997). A Neurobiologically Inspired Model of Working Memory Based on Neuronal Synchrony and Rythmicity. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.), *Proceedings of the Fourth Neural Computation and Psychology Workshop: Connectionist Representations*. London: Springer-Verlag.
- Thomas, R.D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology*, 40, 318-341.