Not guppies, nor goldfish, but tumble dryers, Noriega, Jesse Jackson, panties, car crashes, bird books, and Stevie Wonder

GERT STORMS, PAUL DE BOECK, IVEN VAN MECHELEN, and WIM RUTS
Katholieke Universiteit, Leuven, Belgium

This paper focuses on the *guppy effect* (Osherson & Smith, 1981), that is, on the existence of examples of conjunctive concepts that are more typical of the conjunction than of both constituents. The most frequently given examples of this effect, *guppy* and *goldfish*, are shown not to be more typical of the conjunction *pet fish* than of *fish* in two between-subjects and one within-subjects experiment. The frequency of the effect in a large empirical study is investigated, and better examples of the effect are suggested.

In an influential article, Osherson and Smith (1981) demonstrated that prototype theory (see, e.g., Posner & Keele, 1968; Rosch, Simpson, & Miller, 1976) formalized according to the principles of fuzzy-set theory (Zadeh, 1965) fails to explain membership in complex concepts. Numerous publications published in the seventies have shown that entities fall neither in nor sharply out of a concept's extension. Rather, the boundary between membership and nonmembership in a concept's extension is graded, or *fuzzy*. Fuzzy-set theory provides a framework within which to formalize such graded membership in that it expands classical set theory (where membership in a set is dichotomous—i.e., either 1 or 0) by allowing membership in a set to take any value between 0 and 1. The more typical an entity is for a category, the closer its membership value is to 1; the less an entity is related to the category, the closer its membership value is to 0.

Osherson and Smith (1981) argue that prototype theory, when formalized in terms of fuzzy-set theory, contradicts strong intuitions people have about conjunctive concepts. Within fuzzy-set theory, conjunctive concepts like *red house* and *square field* are most naturally represented as fuzzy intersections, which are defined by a minimum rule: graded membership in the conjunctive concept is assumed to equal the minimum of the graded membership value in the two constituent concepts.

Osherson and Smith then deduce that prototype theory, when formalized in terms of fuzzy-set theory, "will lead to a contradiction whenever an object is more prototypical of

a conjunction than of its constituents" (Osherson & Smith, 1981, p. 45), because it not only breaks the minimum rule, but even a maximum rule. The authors further express the intuition that numerous familiar conjunctions provide counterexamples, and, thus, that prototype theory, at least when formalized in terms of fuzzy-set theory, cannot account for graded membership in conjunctive concepts. They then give *guppy* as an intuitive example, which, they argue, is more prototypical of the conjunctive concept *pet fish* than it is of either *pet* or *fish*.

The *guppy* example has since been referred to in many articles on conjunctive concepts by leading authors and in important journals and is sometimes called the *guppy effect* (see, e.g., Cohen & Murphy, 1984; Fodor & Lepore, 1996; Hampton, 1987, 1988, 1991, 1996; Huttenlocher & Hedges, 1994; Jones, 1982; Medin & Smith, 1984; Murphy, 1988; Osherson & Smith, 1982; Smith & Osherson, 1984). Several authors also referred to *goldfish* as an example of the guppy effect (see, e.g., Fodor & Lepore, 1996; Hampton, 1987, 1988). However, none of these authors verified this guppy effect empirically. In this paper, we describe three experiments where *guppies* and *goldfish* are shown to be bad examples of the guppy effect, and we propose better examples of the effect, derived from the data of a large empirical study (Storms, De Boeck, Van Mechelen, & Ruts, 1996).

In an attempt to verify empirically whether the example *guppies* is more typical of the conjunction *pet fish* than it is of both *pets* and *fish*, we adopted the same procedure as that described in Hampton (1988) and in Storms, De Boeck, et al. (1996). In short, 30 exemplars (10 exemplars of *pet fish*, 10 exemplars of *pets*, and 10 exemplars of *fish*), including both *guppy* and *goldfish*, were selected on the basis of production frequencies in an exemplar-generation task (thus selecting the 10 most frequently generated exemplars in each of the three lists). Next, three groups of 10 English subjects (all students at the British School of Brussels) each rated the typicality of all 30 exemplars for one of three categories—*pets*, *fish*, or *pet fish*—on a 7-point rating scale, ranging from −3 (not related) to +3

— Accepted by previous editor, Geoffrey R. Loftus

(highly typical). Estimated reliability of the ratings (using the split-half method) for the categories varied from .98 to .99. Both *guppy* and *goldfish* failed to show the effect referred to by Osherson and Smith (1981). More specifically, the mean typicalities of *guppy* were 1.7, 1.8, and 1.6 for *fish*, *pets*, and *pet fish*, respectively. For *goldfish*, mean typicalities of 3.0, 2.6, and 3.0 were found.

Since different fish may be popular as pets in the United States from those common in Europe, a replication experiment was conducted, with subjects from North America only. In this second experiment, the same set of exemplars was rated by three groups of 8 subjects (20 American citizens and 4 Canadians), all of whom were students in English postgraduate programs at the University of Leuven. Estimated reliability of the ratings (using the split-half method) for the categories varied from .94 to .97. Again, both *guppy* and *goldfish* failed to show the effect. The mean typicalities of *guppy* were 2.75, 1.5, and 2.13 for *fish*, *pets*, and *pet fish*, respectively. For *goldfish*, mean typicalities of, respectively, 3.0, 2.6, and 2.87 were found. The exemplar *sea horse*, however, did show the effect (with mean ratings of, respectively, 1.0, −0.5, and 1.88).

The two experiments described above may be criticized because we averaged data from different subjects in a between-subjects design. Therefore, a third experiment was conducted in which a within-subjects design was used. Twelve (Dutch-speaking) subjects rated the typicality of 36 Dutch words (6 exemplars of *pet fish* and 15 exemplars of both *pets* and *fish*) for both constituent categories (on consecutive days, in counterbalanced order) and for the conjunction (one week later) on a 7-point rating scale. Estimated reliabilities of the ratings for the categories varied from .98 to .99. Again, the results showed that, averaged over subjects, neither *guppy* nor *goldfish* shows the guppy effect. The mean typicality of *guppy* was 2.0, 1.7, and 1.5 for *fish*, *pets*, and *pet fish*, respectively. For *goldfish*, the mean typicalities were 3.0, 2.7, and 2.8, respectively. Thus, the typicality for the conjunction is lower than the typicality for *fish* in both cases. Moreover, none of the 12 subjects rated the typicality of either *guppy* or *goldfish* higher for the conjunction than they did for the *fish* constituent, showing that, even in a within-subjects design, *guppies* and *goldfish* are not examples of the guppy effect. None of the other 34 exemplars used in this

Table 1
Items From Storms et al. (1996) With a Significant Effect (*p* < .05)
on the Randomization Tests

Item	Combination	Difference Between Mean Conjunction Ratings and Maximum of the Mean Constituent Ratings
AIDS	causes of death–diseases	0.30
Pestilence	causes of death–diseases	0.75
Paraffin	energy sources–liquids	0.55
Tennis	field sports–outdoor sports	0.60
Denzel Washington	Americans–black people	0.75
Diana Ross	Americans–black people	0.50
Jesse Jackson	Americans–black people	1.15
Stevie Wonder	Americans–black people	1.00
Parrot	birds–pets	0.40
Parakeet	birds–pets	0.50
Tractor	machines–vehicles	0.65
Billiards	sports–games	0.60
Alcohol-felt-tip	writing implements–office equipment	0.70
Felt-tip	writing implements–office equipment	0.55
Fountain pen	writing implements–office equipment	0.20
Basketball shoes	footwear–sports equipment	0.75
Tumble dryer	furniture–household appliances	1.40
Musician	professions–hobbies	0.45
Sportsman	professions–hobbies	0.75
Soccer player	professions–hobbies	0.60
Fidel Castro	politicians–military men	0.55
Noriega	politicians–military men	1.35
Saddam Hoessein	politicians–military men	0.70
Panty	pants–underwear	1.15
Claus	given names–surnames	2.35
Cruise	trips (voyages)–recreational activities	0.50
Atlas	books–reference sources	0.55
Manual	books–reference sources	0.75
Bird book	books–reference sources	1.05
Hair brush	toiletries–brushes	0.40
Tranquilizer	drugs–medicines	0.45
Bonsai tree	decorative plants–trees	0.40
Car crash	murdering methods–suicide methods	1.15
Jumping out of a plane	murdering methods–suicide methods	0.95
Drowning	murdering methods–suicide methods	0.85
Flashing light	sources of light–signals	0.50

study showed the effect. The correlations (corrected for unreliability) between the typicality ratings given by the British and by the North American subjects were .93, .99, and .73 for *pets, fish*, and *pet fish*, respectively.

In a large research project, set up with a different purpose and described elsewhere in detail (Storms, De Boeck, et al., 1996), we studied membership ratings in 50 concept conjunctions described as relative clauses (such as *sports that are also games*), by means of 30 exemplars per combination (the 10 most frequently generated exemplars in an exemplar-generation task, both of the constituents and of the conjunction). In 36 out of 50 conjunctions, at least one exemplar showed a higher rating for the conjunction than for both constituent concepts.[1] In some conjunctions (e.g., *professions that are also hobbies*), more than one out of three exemplars showed the effect. A total of 119 words (out of the 1,500 words used) showed the guppy effect.

One may wonder whether the higher ratings in the conjunction than in both constituents in these 119 words can be attributed to chance fluctuations in mean ratings (averaged over 10 subjects in every condition). To test the statistical significance of the effect, parametric tests are not suitable, because the distribution of the ratings within the different conditions may deviate considerably from normality, especially for words that have mean values close to +3 on the 7-point rating scale. Therefore, randomization tests (Edgington, 1995) were carried out for all words that showed the effect.[2] Out of the 119 cases that had higher conjunction ratings than they had constituent ratings, 36 words showed a significant effect at the 5% level (of which 9 are significant at the $\alpha = .01$ level). Table 1 shows these words, the conjunctions from which they were taken, and the difference in mean membership rating between the conjunction ratings and the maximum of the mean values of both constituents. The number of significant cases is less than what can be expected by chance with an α level of .05. To verify whether the effect is reliable and consistent over subjects, the subject groups in all conditions were randomly divided into two subgroups and the effect was checked again in both subgroups. In 61 of the 72 comparisons (2 × 36), the effect showed up again. Some of these cases—such as the exemplar *Claus* for the conjunction of *given names* and *surnames*—may be restricted to the Flemish situation. Others, however—such as *Noriega* for the conjunction of *politicians* and *military men*—can be considered general examples of the effect at stake.

In conclusion, our data support Osherson and Smith's (1981) intuition that conjunction exemplars do exist that are more typical of the conjunction than they are of both constituents. Our data thus support their argument that prototype theory, as formalized by fuzzy-set theory, cannot account for these data. However, the data also show that *guppy* and *goldfish* are not instances of this typicality phenomenon. For other conjunctions—such as *Americans that are also blacks, military men that are also politicians, books that are also reference sources*, and *professions that are also hobbies*—several excellent examples of the phenomenon described by Osherson and Smith (1981) occur.

REFERENCES

COHEN, B., & MURPHY, G. L. (1984). Models of concepts. *Cognitive Science, 8*, 27-58.

EDGINGTON, E. S. (1995). *Randomization tests* (3rd ed.). New York: Dekker.

FODOR, J., & LEPORE, E. (1996). The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition, 58*, 253-270.

HAMPTON, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15*, 55-71.

HAMPTON, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 12-32.

HAMPTON, J. A. (1991). The combination of prototype concepts. In P. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 91-116). Hillsdale, NJ: Erlbaum.

HAMPTON, J. A. (1996). Conjunctions of visually based categories: Overextension and compensation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 378-396.

HUTTENLOCHER, J., & HEDGES, L. V. (1994). Combining graded categories: Membership and typicality. *Psychological Review, 101*, 157-165.

JONES, G. V. (1982). Stacks not fuzzy sets: An ordinal basis for prototype theory of concepts. *Cognition, 12*, 281-290.

MEDIN, D. L., & SMITH, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35*, 113-138.

MURPHY, G. L. (1988). Comprehending complex concepts. *Cognitive Science, 12*, 529-562.

OSHERSON, D. N., & SMITH, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 9*, 35-58.

OSHERSON, D. N., & SMITH, E. E. (1982). Gradedness and conceptual combination. *Cognition, 12*, 299-318.

POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

ROSCH, E. H., SIMPSON, C., & MILLER, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception & Performance, 2*, 491-502.

SMITH, E. E., & OSHERSON, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8*, 357-361.

STORMS, G., DE BOECK, P., VAN MECHELEN, I., & RUTS, W. (1996). The dominance effect in concept conjunctions: Generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 1266-1280.

STORMS, G., RUTS, W., & VANDENBROUCKE, A. (1996). *Syntactic rephrasing of conceptual conjunctions*. Unpublished manuscript, University of Leuven.

ZADEH, L. A. (1965). Fuzzy sets. *Information & Computation, 8*, 338-353.

NOTES

1. In this study, the rating scales were labeled as *membership ratings*, whereas the rating scales in the three experiments above were labeled *typicality ratings*. A study by Storms, Ruts, and Vandenbroucke (1996), in which both typicality ratings and memberships were compared, showed that both rating scales correlate almost perfectly and thus can be considered equivalent.

2. In a randomization test, a large number of random permutations (e.g., 1,000) of all observations are used to redistribute the observed values over the different conditions of an experiment. For each randomization, a statistic is calculated—in this case, the difference between the mean conjunction membership rating and the mean rating of whichever constituent had the largest mean rating. The distribution of these calculated statistics over all random permutations is formed and the empirically found value is evaluated. If values less than α times 100 in the distribution have a more extreme value than the empirically obtained statistic, then the test is considered significant at the α significance level.