Picture naming*

Wilhelm R. Glaser

Psychological Institute, University of Tübingen, Friedrichstrasse 21, D-7400 Tübingen, Germany

Abstract


Picture naming has become an important experimental paradigm in cognitive psychology. To name a picture can be considered an elementary process in the use of language. Thus, its chronometric analysis elucidates cognitive structures and processes that underlie speaking. Essentially, these analyses compare picture naming with reading, picture categorizing, and word categorizing. Furthermore, techniques of double stimulation such as the paradigms of priming and of Stroop-like interference are used. In this article, recent results obtained with these methods are reviewed and discussed with regard to five hypotheses about the cognitive structures that are involved in picture naming. Beside the older hypotheses of internal coding systems with only verbal or only pictorial format, the hypotheses of an internal dual code with a pictorial and a verbal component, of a common abstract code with logogen and pictogen subsystems, and the so-called lexical hypothesis are discussed. The latter postulates two main components: an abstract semantic memory which, nevertheless, also subserves picture processing, and a lexicon that carries out the huge amount of word processing without semantic interpretation that is necessary in hearing, reading, speaking and writing.

1. Basic data and early hypotheses

Concrete objects, natural or man-made, constitute the inventory of our everyday world. Only a small part of our actions consist of direct physical operations with these objects; most of them are symbolically mediated. There are two comprehen-

*I am grateful to Margrit O. Glaser for many helpful discussions and her assistance in formulating this text. The help of Hanne Bonnellycke in all questions concerning English usage is gratefully acknowledged. Furthermore, I thank Wido La Heij and an anonymous reviewer for many substantial suggestions to improve this article.
sive classes of symbols for concrete objects: pictures, that is, photographs, line
drawings or pictograms, and words, in particular nouns. Pictures become symbols
of their objects or object classes by physical similarity. Therefore, it is a very
plausible hypothesis that recognizing pictures comprises essentially the same
cognitive processes as perceiving the objects themselves (Potter, 1979). Recogniz-
ing pictures does not require particular steps of learning or development beyond
learning to know the represented objects. On the other hand, the relation
between a noun and the corresponding class of objects is determined, in a certain
sense arbitrarily, during the centuries of evolution of a language. Nouns are,
furthermore, often polysemous because an individual word can represent very
different objects. Likewise, any given object can be named with different words
depending on its context and the intentions of the speaker. Everybody has to
learn spoken words, their meaning, and their written form. Both kinds of
symbolism, picture and word, play a fundamental role in human culture.

The psychological analysis of symbol use postulates basic mapping processes
between objects and symbols as well as between different symbol systems and
modalities. Therefore, the early experimental psychologists had already begun to
investigate these processes. The most promising method was until now chronomet-
ric analysis.

Cattell (1885) found that reading aloud a printed list of 100 nouns in the
speaker's native tongue took about 25–35 s, whereas 50–60 s were required for
naming a comparable list of 1-cm wide line drawings or small colour dots. Because tachistoscopic experiments had shown that his pictures were recognized
slightly faster than his words, Cattell concluded that the naming time is prolonged
due to particular difficulties to retrieve the name of a recognized object. The high
speed of the reading response, on the other hand, is explained by a strong
association between a written word and its pronunciation caused by extensive
practice (Cattell, 1885, p. 650, 1886, p. 65). Eventually, Cattell conjectures that
loud reading responses became automatic because they are repeated very often in
everyday life, whereas naming requires a voluntary effort. Thus, Cattell antici-
pated the modern distinction of automatic and controlled cognitive processes.
However, we will neither discuss what exactly Cattell means by the word
automatic, nor to what degree overt reading is really practised by the average
adult.

This line of investigating picture (or colour) naming by comparing it with
reading was continued by Brown (1915). He predicted from the differential
practice hypothesis that sufficient additional training should bring naming times
down to the reading times, whereas reading times should not show considerable
improvement. The two predictions failed: although the naming responses were
accelerated by 25.8% during 12 days of extensive training, the time for naming
remained 41% longer than that for reading. The latter was still reduced by 16.8%
due to the same amount of training (Brown, 1915, Table i). With these results, it
appeared impossible to explain the difference in reading and naming times by a
different amount of practice. Later on, Lund (1927) and Ligon (1932) found a
constant reading–naming difference over age for all subjects who had learned to
read, although reading as well as naming times strongly decreased from the first to
the ninth grade among students.

Fraisse (1969) reported experiments which confirmed that the reading–naming
difference cannot be reduced by practice. Therefore he rejects the differential-
practice hypothesis. Furthermore, like Cattell (1885), he obtained almost the
same tachistoscopic recognition threshold for line drawings and for words. Event-
ually, Fraisse (1969) outlined a tentative explanation in terms of higher com-
patibility between written and spoken word, compared to the picture and its
name. For each word, there is only one reading response, which is uniquely
determined by the print. On the other hand, there are several ways to name a
common object. They cause a response uncertainty in the naming task which
increases the reaction time because the correct answer is to be determined by
additional restrictions given in the instruction. Beyond these theoretical considera-
tions, Fraisse (1967, 1969) presented a particular experimental highlight: the
identical symbol O was named as circle in 619 ms, as zero in 514 ms, and read as
oh in 453 ms.

Since then, this reading–naming difference was often replicated by several
authors: 260 ms (Potter & Faulconer, 1975); 348 ms (Smith & Magee, 1980,
Experiment 3); 173 ms (Irwin & Lupker, 1983, Experiment 3); 222 ms (Glaser &
Düngelhoff, 1984, Experiment 1, neutral condition at stimulus onset asynchrony
(SOA) = 0 ms); 257 ms (Potter, So, Von Eckardt, & Feldman, 1984); 133 ms
(Glaser & Glaser, 1989, Experiment 6, control condition, mean over SOAs);
145 ms (Theios & Amrhein, 1989); 178 ms (Bajo, 1988).

In these experiments, the number of response alternatives was not less than
four, and was often far greater. Gholson and Hohle (1968) compared reading and
naming times for different numbers of alternatives. The usual difference as found
for six alternatives decreased for four and disappeared for two alternatives. In a
picture-naming task, La Heij and Vermeij (1987) found response times in the
magnitude of the reading times for two alternatives. Similarly, Virzi and Egeth
(1985, Experiment 2, “vocal” condition) did not obtain a reading–naming differ-
ence in a spatial variant of a Stroop-like experiment with two alternatives. On the
other hand, this difference was fully preserved in the Stroop experiment with two
colour and colour–word alternatives by Simon and Sudalaimuthu (1979). How-
ever, I will not discuss why naming can become so efficient in the special case of
two alternatives.

It is noteworthy that Potter et al. (1984) and Biederman and Tsao (1979)
obtained reading–naming differences of 305 ms and 266 ms, respectively, in the
Chinese language. This contradicts an explanation of the reading–naming differ-
ence due to a close grapheme–phoneme correspondence in Western languages.
Oldfield and Wingfield (1965) obtained further fundamental data on picture naming. They investigated the influence of word frequency (as counted by Thorndike & Lorge, 1944) on picture naming and found a negative linear relation between naming latency and the logarithm of word frequency. The slope was \(-254\) ms per ten-log frequency unit. Oldfield and Wingfield (1965) attributed this frequency effect to the process of name retrieval. It seemed plausible that frequent words are retrieved faster than rare ones, but an explanation for the log-linear shape of this relation is lacking. It is reasonable to compare this result with reading. Here, the log-linear relation between word frequency and reading latency holds as well, but with a much lesser slope of about \(-30\) ms per ten-log frequency unit (Frederiksen & Kroll, 1976; Huttenlocher & Kubicek, 1983). That argues again for a more automatic or more compatible access of the internal word code by printed words rather than by pictures.

Concerning the theoretical explanations, it was the cognitive revolution that provided the preconditions for greater progress, because now structures of internal processing stages together with their codes and procedures could be modelled.

2. Important models for picture naming

In the late 1960s, several models that are relevant for picture naming emerged from different starting points. The logogen model (Morton, 1970) was centred around internal word codes (logogens) that are necessary to recognize, produce and understand words. Seymour (1973) and Warren and Morton (1982) further elaborated the logogen model in order to cover naming and semantic comparisons of pictures and physical objects. Starting from an early computerized language comprehension system, Collins and Quillian (1969) developed a first network model of semantic memory which was essentially improved by Collins and Loftus (1975). Smith, Shoben, and Rips (1974) presented a featural model of semantic decisions, but it proved impossible to decide empirically between network and feature models (cf. Hollan, 1975).

Starting from phenomena of mental imagery, Paivio (1971) proposed his dual code hypothesis of semantic memory. It states that semantic knowledge is represented internally by a verbal as well as a pictorial code. Which one of them is retrieved depends on the demands of the task; generally, picture stimuli tend to activate pictorial codes, and word stimuli are first encoded as word codes. Of course, there are internal translations between both coding systems. This hypothesis was soon disputed with profound methodological and, to some extent, philosophical arguments by Pylyshyn (1973). The core of this critique is that the attributes verbal and pictorial characterize modes of experience in perception or imagery, but not explanatory theoretical constructs. Therefore, "the need to
postulate a more abstract representation—one which resembles neither pictures nor words and is not accessible to subjective experience—is unavoidable" (Pylyshyn, 1973, p. 5).

Thus, with the cognitive revolution, the central questions began to address the modality or the "format" of the internal codes and their transformations in naming, reading and different novel types of tasks. I hypothesize that five different, fundamental classes of hypotheses can be discerned, although all of their demarcations are already clear:

1. There is only one word-like internal code system. Its units represent clusters of associations among perceived objects, their properties and functions, experiences from actions with them, and the spoken and written modalities of their verbal labels (e.g., Deese, 1962). This hypothesis is a heritage from behaviorism and difficult to reconcile with a great deal of present-day empirical data.

2. There is only one internal semantic code of concepts but not of words. The concepts are represented pictorially; the traces in long-term memory represent past perceptions and experiences. Essentially, this hypothesis is also a historical reminiscence. Nevertheless, it can explain some phenomena of mental imagery as well as the superior memory performance for pictures rather than for words (cf. Paivio, 1969). However, it remains unresolved how processes of abstraction and generalization could operate on an internal picture code and how they are functionally connected with verbal symbols.

3. The dual-code hypothesis tries to escape these problems by combining hypotheses 1 and 2. Thus, the concrete, perceptual properties of concepts should be coded as internal pictures, the abstract and functional features as well as all linguistic attributes in an internal word system (Paivio, 1971, 1978). This hypothesis was the starting point of many experimental tests which are reported below. It implies some form of a simple architecture of human cognition. Thus, its fundamental structure is given in Figure 1(a). However, the dual-code hypothesis, too, does not fully explain higher cognitive processes like comparison, classification, abstraction and inference, and it is seriously disputed by many experimental data.

4. At present, the most accepted hypotheses postulate a central, abstract, amodal and propositional internal code for long-term storage. This code is generally unconscious and not accessible to introspection. It provides the informational basis for the more complex cognitive operations. The percepts of external objects and events are translated into this code if they are to be stored or used in cognitive tasks. The contents of this storage can be read out for conscious operations in working memory, among them mental imagery, and they can be used for perception and action. The connection of this abstract code system with perception and action on the one hand, and with language understanding and
Pictorial, verbal, and abstract components of the cognitive system according to different hypotheses: (a) dual-code hypothesis (after Pellegrino, Rosinski, Chiesi, & Siegel, 1977, Figure 1, p. 384); (b) hypothesis of a central, abstract code (after Seymour, 1973, Figure 2, p. 39); (c) lexical hypothesis (after Glaser & Glaser, 1989, Figure 5, p. 31).
production on the other, requires large recoding systems with extended processing capacity and inherent long-term storage (Potter, 1979; Seymour, 1976; Snodgrass, 1980, 1984; Theios & Amrhein, 1989). Their long-term components contain canonical or prototypical pictures of everyday objects, productions for skilled actions, and the morphemes of all words that an individual knows together with their syntax as well as their phonemic and orthographic properties (e.g., Seymour, 1976; Snodgrass, 1984). Figure 1(b) shows a configuration of systems according to this hypothesis that is close to Seymour’s (1973) modification of Morton’s logogen model. It is essentially based on Seymour’s (1973, 1976) ideas.

(5) Recently, psycholinguists have shown that the production of speech embodies large and complicated operations with morphemes or words outside the central propositional system (for a comprehensive review see Levelt, 1989). On the other hand, Glaser and Glaser (1989) tried to integrate the known data of studies on Stroop and picture-word interference into a unique cognitive model. In particular, they found that the Stroop-like word-word interference does not exhibit semantic components. This provides further evidence for an extensive internal word processing without semantic components. Thus, they suggested a modification of Seymour’s model, which is presented in Figure 1(c). It rests on two ideas. First, the amount of autonomous, semantically uninterpreted word processing that is necessary in language use is underestimated in Seymour’s model. Thus, the logogen component of Seymour’s model is upgraded to the linguistic one of two central parts of the long-term memory. It contains the complete thesaurus of morphemes that an individual knows, perhaps in several different languages. Second, whereas a word as a perceptual object has its graphemic and phonemic features which are internally connected with representations of morphemes and lexemes, the features of physical objects are essential components of the meaning of their concepts. Thus, an iconogen system in addition to the semantic memory seems to be unnecessary. Although the concepts of the semantic memory are abstract nodes in a network, they essentially comprise physical features. Thus, we suggest eliminating the iconogen system from Seymour’s model and implementing its functions, the semantic analysis of perceived pictures and objects, into the semantic system. This leads again to a relatively simple architecture with only one fundamental distinction between a large concrete and abstract, semantic and non-verbal system on the one side and an equally large verbal and linguistic system on the other. For lack of an established name, we will call this the lexical hypothesis. At first glance, it seems to be similar to the dual-code hypothesis (3). However, there are at least two clear demarcations. First, in the lexical hypothesis, picture stimuli are functionally tied with the abstract module. Second, the word module provides no semantic processing, not even of abstract meanings or relations. Of course, it is still to be demonstrated that the essentials of the lexical hypothesis, the extension of the logogen system and the deletion of a separate iconogen system besides the semantic memory are empirically justified.
3. Experimental methods

These hypotheses were developed in close connection with a growing repertory of experimental methods. At the beginning, there were only naming and reading latencies and their comparison. Later on, a great methodological progress arose from the techniques of double stimulation (Kantowitz, 1974). Thereby, two pictures, two words, or a picture and a word are presented simultaneously or in close temporal succession. There are three main types of double-stimulation experiments:

(1) In the matching paradigm, both stimuli are to be compared according to a given rule. This technique descends from the same–different matching task as initiated by Posner and Mitchell (1967). It became a very productive tool for tackling hypothetical internal stages and processes.

(2) In the priming paradigm (Beller, 1971), the facilitating influence of the voluntary or involuntary processing of one stimulus (prime) on the latency of the response to the other stimulus (target) is investigated.

(3) In the Stroop paradigm any reaction to the irrelevant stimulus (distractor) is to be suppressed. Nevertheless, the distractor often interferes strongly with responding to the target. The amount and time course of this interference provide information on internal structures and processes (e.g., Glaser & Glaser, 1989). Although the logic of priming and Stroop experiments are very similar, both techniques were developed independently and were only recently systematically compared with each other (see La Heij, Dirkx, & Kramer, 1990; La Heij, Van der Heijden, & Schreuder, 1985).

However, the study of picture processing was not only extended by these recent experimental techniques but also by an extension of picture naming itself. If presented with a picture of a chair and asked “what is that?”, a subject can well respond furniture, chair, or kitchen chair. All three responses are labels for categories of objects which form a taxonomy. Apparently, these three labels represent different levels of categorization. In particular Rosch (1975) and Rosch, Mervis, Gray, Johnson and Boyes-Braem (1976) elaborated the different psychological functions of these levels. The response chair belongs to the basic level. This level contains the fundamental classifications of everyday life. The labels of this level represent classes of objects that share a great number of perceptual and functional features. Basic level objects are easiest to be drawn. In contrast, superordinate categories as, for example, furniture, comprise many visually and functionally different objects that share only a few properties. They are defined by rather abstract rules, and they can be visualized only by prototypes at basic level. Finally, the subordinate level results from additional distinctions within basic level categories. They are less relevant for everyday life but become far more important in professional work. To picture at subordinate level requires again drawing a
basic object but with emphasis on special features that underlie a subordinate
delineation.

This level of designation was introduced as an independent variable into the
naming task towards the end of the 1960s. Thereby, the term picture naming was
usually restricted to the verbal responses at basic level, whereas naming responses
at superordinate level were called categorizing. We will adopt this terminology in
the following, but it must be emphasized that for a given picture both tasks consist
of retrieving a verbal label in the mental lexicon.

Now, we will present the most important recent results concerning picture
processing that are based on the aforementioned experimental methods. We will
begin with the categorizing studies, then proceed to the priming investigations,
and end with the Stroop-like experiments.

4. Pictures and their superordinate labels

If pictures of everyday objects like chair, hat, or saw are presented and subjects
are asked "what is that?", then they respond almost solely at basic level, that is,
they answer "chair", "hat", and "saw" (Potter & Faulconer, 1975; Rosch et al.,
1976, Experiment 10; Segui & Fraisse, 1968). Nevertheless, instructed to do so,
they can also easily respond with the superordinate labels, for example
"furniture", "clothes", or "tool", but now the reaction times are usually longer
than in the naming task at basic level. This categorizing–naming difference was
often reported in the literature. Its amount was 162 ms (Segui & Fraisse, 1968),
71 ms (Smith, Balzano, & Walker, 1978), 152 ms (Irwin & Lupker, 1983,
Experiment 3), 266 ms (Smith & Magee, 1980, Experiment 4) and 102 ms (Glaser
& Düngelhoff, 1984, Experiments 1 and 2, neutral condition at SOA = 0 ms).
However, small amounts of this difference (47 ms; Glaser & Glaser, 1989,
Experiment 6, control condition, mean over SOAs) were also obtained.

There is another important variant of the categorizing task. Now, the concept
at basic level is not presented as a picture but as a word. Therefore, the subject
has to name the superordinate category of a concept that is activated by reading
silently the stimulus word. Of course, the reaction time in this word-categorizing
task is longer than that in reading, but it is also prolonged compared to picture
categorizing. This word–picture difference for categorizing was already reported
by Segui and Fraisse (1968). For the categories furniture, clothes, musical
instrument, vegetable, and weapon, these authors obtained a word–picture differ-
ence of 75 ms. Other researchers found 66 ms (Irwin & Lupker, 1983, Experiment
3), 126 ms (Glaser & Düngelhoff, 1984, Experiment 2, neutral condition at
SOA = 0 ms), 95 ms (Glaser & Glaser, 1989, Experiment 6, control condition,
mean over SOAs).

The word–picture difference for categorizing is theoretically at least as im-
portant as the naming-reading difference. Assuming that sensory encoding of the stimulus takes about the same time for pictures and for words (Fraisse, 1969; Potter & Faulconer, 1975; Potter et al. 1984; Theios & Amrhein, 1989), then the following predictions can be derived. If there is only one internal code with verbal format, then picture categorizing should take longer than word categorizing, because picture stimuli must be translated into the word code which is directly accessed by word stimuli. This assumption is at odds with the data. If the single internal code has pictorial format, then picture categorizing should be faster than word categorizing, as is indeed the case, because pictures should have more direct access to the internal code. However, as mentioned above, this hypothesis has other problems: it cannot explain abstract cognitive operations as, for example, the build-up of type-token relations of mentally stored to perceived objects (cf. Pylyshyn, 1973, p. 7). Also the categorizing–naming difference of pictures is difficult to reconcile with this hypothesis. If the internal code had a visual format, recognizing, for example, a chair as an instance of furniture should need less feature detection than its identification at basic level.

The dual-code hypothesis, too, cannot predict the word–picture difference in categorizing. Because the demarcations at the superordinate level are rather abstract, they should be stored in the verbal system that is accessed faster by word than by picture stimuli. Thus, contrary to the data, word categorizing should require less time than picture categorizing.

According to the hypothesis of a single abstract code, picture as well as word stimuli are translated into an abstract format that allows their superordinate category membership to be determined. Thus, equal reaction times for categorizing pictures and words are predicted as long as two auxiliary hypotheses hold. Firstly, not only perceptual encoding, but also access to the abstract semantic code take the same time for picture and word stimuli, and secondly, the times for selection and articulation of the response are equal in both cases. Therefore, the abstract-code hypothesis remains only in line with the observed word–picture difference in categorizing, if at least one of the auxiliary hypotheses is rejected. Because the verbal responses are the same in picture and word categorizing, there is no reason to doubt that the times to select and articulate them are equal in both tasks. On the other hand, the abstract-code hypothesis can only be reconciled with the word–picture difference in categorizing, if the first auxiliary hypothesis is rejected, that is, if faster access of the abstract system for pictures than for words is assumed. This leads over to the lexical hypothesis.

The lexical hypothesis states close functional connections of object and picture perception to the abstract semantic system, whereas word processing contains extended linguistic functions that are carried out in the lexical system without semantic evaluation (Glaser & Glaser, 1989; Levelt, 1989). Considered as a speech act, the categorizing task requires only little linguistic processing. Nevertheless, the processing of word stimuli should pass a linguistic stage before the
abstract code is reached. Thus, pictures have privileged access to the semantic system. The lexical hypothesis predicts the observed word–picture difference in categorizing.

As these theoretical considerations have shown, the picture–word difference in categorizing latency is one of the most important data. Thus, many experiments addressed its reliability. We will give a short survey. In addition to vocalizing the category name (Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Segui & Fraisse, 1968; Smith et al., 1978; Smith & Magee, 1980, Experiment 4), often a semantic decision was demanded from the subjects whether two objects belong to the same superordinate category or whether one object belongs to a certain category. For example, Potter and Faulconer's (1975) subjects had to say yes if a stimulus object belonged to a pre-specified superordinate category. Otherwise, they had to say no. The objects were represented by line drawings or their printed basic-level names, respectively. Altogether, 96 concepts from 18 different categories were used. The “yes” responses were on the average 57 ms faster for pictures than for words, the “no” responses by 44 ms. The mean of both measures was 51 ms. This is a clear picture–word difference that argues for a privileged access of the picture to the semantic code (see also Hogaboam & Pellegrino, 1978). Potter et al. (1984) replicated these results with Chinese native speakers who later lived in English-speaking countries. They were still more fluent in Chinese than in English, and on average they categorized pictures 34 ms faster than Chinese word symbols.

Rosch (1975) investigated the categorization task with the same–different matching technique. In each trial, two line drawings or two words that represented basic-level objects were given. The subjects had to indicate by pressing a “yes” or “no” key whether or not both objects belonged to the same superordinate category. Altogether, ten different categories were used. The author obtained a general picture advantage of about 260 ms (my calculation from Rosch, 1975, Figure 1, “yes” responses). Friedman and Bourne (1976) used a technique they called speeded-inference task. It is closely related to same–different matching: in each trial, a pair of words or pictures was presented in close succession. Both signified objects at basic level that could match either in a physical property such as size, or in a conceptual property, like membership in a common superordinate category. The subjects had to utter aloud the value of that property that was equal for both stimuli. As an illustration, suppose the item pool would contain only the four concepts hippo, mouse, bus, and car. Hippo and mouse are both animals, bus and car are both vehicles; besides that, hippo and bus are large, mouse and car are small, compared to the other animal or vehicle, respectively. Now the subject had to respond, for example, “large” to the stimulus pair bus–hippo or “vehicle” to the pair bus–car. In Experiment 1, these responses were given to picture pairs on an average of 96 ms and 128 ms faster than to word pairs.
Pellegrino, Rosinski, Chiesi, and Siegel (1977) extended the same-different categorical judgment task. They presented not only word-word and picture-picture pairs, but also mixed pairs of one word and one picture. The mixed pairs were called *picture-word* if the left element was the picture and the right element was the word. *Word-picture* designated the opposite case. Thus, the four modality conditions picture-picture, picture-word, word-picture, and word-word resulted. Both stimuli were simultaneously exposed and the subjects had to press either a "yes" key if both stimuli belonged to the same superordinate category or a "no" key if they did not. The two categories *animals* and *articles of clothing* were used.

The authors formulated models of additive stages for three theoretical alternatives: (1) dual code with category information in the pictorial and in the verbal system; (2) dual code with category information only in the verbal system; and (3) common abstract code for semantic information including category membership. Hypothesis 1 was extended by the auxiliary hypothesis that the subjects process the stimulus pair from left to right, according to reading habits. Thus, the category codes of picture-word stimuli should be matched in the word system, and those of word-picture pairs in the pictorial system. A set of four linear equations made it possible to estimate the internal transfer times between both systems under hypothesis 1 from the mean reaction times of the four stimulus conditions.

The results showed mean latencies of 715 ms for picture-picture pairs, 810 ms for picture-word pairs, 851 ms for word-picture pairs, and 900 ms for word-word pairs. The estimators for the internal transfer times were not significantly different from zero, so that the hypothesis of dual code with dual category information (1) was rejected. The difference of 185 ms between word-word and picture-picture pairs argues furthermore against a dual code with category information in the word system (2). Thus, the authors accept the hypothesis of a common semantic code (3). However, this hypothesis is only tenable if it is assumed that pictures have faster access to the semantic code than words. The mathematical model yields $185/2 = 92.5$ ms as the advantage of the pictures in retrieving semantic information.

There have been some similar experiments with essentially the same results (e.g., Klimesch, 1982; Rosinski, Pellegrino, & Siegel, 1977). Harris, Morris, and Bassett's (1977) Experiment 1 was almost identical to Experiment 1 by Pellegrino et al. (1977), but showed one fundamentally different result: the shortest latencies for "same category" responses were observed in the word-word condition. This is contrary to almost all other experimental evidence, and could be a serious challenge for the common-code hypothesis with privileged access for pictures. To my knowledge, the reason for this divergence is not yet known. However, I suppose there is an artifact in the work by Harris et al. (1977). As stimulus materials, they used only the two categories *animals* and *vehicles* with four
instances each. The four animals had very similar and in part rhyming three-letter names: pig, dog, cat, and rat. The four vehicles had names with a different number of letters: car, bicycle, train, and boat. That means that—with the exception of car—the subjects had to respond “yes” for every word pair in which no word or both words had three letters, and “no” if one word had three and the other word had more than three letters. It is likely that the subjects utilized these redundancies and matched length and similarity of the words. That could well have accelerated the desired semantic decision. Thus, the first-glance evidence against the common-code hypothesis is questionable. Nevertheless, it demonstrates the possibility that the subjects take advantage of linguistic surface similarity among the verbal labels of the instances of the categories.

The latter argument became very important with regard to picture stimuli. Snodgrass and McCullough (1986) try to explain the picture superiority in categorizing tasks with the visual similarity hypothesis. It states that the pictures are classified or matched at superordinate level not through the internal semantic code but by use of the visual surface similarity that is higher among pictures of the same category than among pictures of different categories. If this hypothesis is true, then “picture categorization is a poor task to use in evaluating hypotheses about the form of representation of semantic knowledge” (Snodgrass & McCullough, 1986, p. 153).

The authors provided a first experimental test of this hypothesis. In their Experiment 1, word–word or picture–picture stimuli were to be classified into two categories by key-pressing responses (cf. Hogaboam & Pellegrino, 1978). In the visually dissimilar condition the categories fruits and animals, and in the visually similar condition fruits and vegetables, were used. The instances of the latter categories were selected for high visual similarity between categories, for example apple, orange as fruits and tomato, potato as vegetables. Indeed, the picture superiority was replicated only in the visually dissimilar condition: pictures were classified as fruits or animals 40 ms faster than words. In the visually similar condition, the results were reversed: to discriminate between fruits and vegetables required 96 ms more for pictures than for words.

There is no doubt that the authors could reverse the word–picture difference in the categorizing task. But is their conclusion justified that the usual categorizing task does not tap semantic access by pictures and by words because picture categorizing is “artifactually” accelerated by the higher intra-category than inter-category similarity of the pictures? To answer this question by “yes” implies the assumption that Snodgrass and McCullough’s (1986) visual similarity condition is the standard, the ecologically most valid case, the control condition, to which the usual results are to be compared. However, that is unlikely. Everyday objects that are different at basic level as defined by Rosch (1975) are visually different among (e.g., guitar, chair) as well as within (e.g., guitar, trumpet) superordinate categories, even if there are more visually similar object pairs within (e.g., guitar,
than among (e.g., racket, violin; cf. Flores d'Arcais & Schreuder, 1987, Fig. 1c) superordinate categories. Thus, the adequate control experiment should contain basic objects with a low or medium visual similarity that is kept constant among and within categories. Many experimenters carefully controlled this variable (e.g., Flores d'Arcais & Schreuder, 1987; La Heij, 1988; La Heij et al., 1990).

However, the argument can be extended. According to Rosch's definition, the basic level is "the most inclusive (abstract) level at which the categories can mirror the structure of attributes perceived in the world" (Rosch, 1978, p. 30; cf. Rosch et al., 1976). Basic and superordinate levels of classification are essentially delineated with respect to visual features: "A working assumption...is that (1) in the perceived world, information-rich bundles of perceptual and functional attributes occur that form natural discontinuities, and that (2) basic cuts in categorization are made at these discontinuities." Thus, concepts at basic level are defined by "attributes in common, motor movements in common, objective similarity in shape, and identifiability of averaged shapes" (Rosch, 1978, p. 31). In contrast, superordinate categories are characterized by a very reduced set of common features, different motor movements, and the impossibility of drawing or imaging an average instance. Thus, it is questionable whether fruits and vegetables are really different superordinate categories at all, although they were explicitly considered as such by Rosch (1975). On the contrary, according to Rosch's definitions, at least the instances that were used by Snodgrass and McCullough (1986) share too many visual features for belonging to different superordinate categories, and thus they are delineated by rather abstract properties (cf. the subsequent discussion of the experiments by Hoffmann and by Flores d'Arcais).

Consequently, the results of Snodgrass and McCullough (1986) demonstrate that an extreme visual similarity among all basic objects that are to be discriminated according to an abstract criterion reverses the word-picture difference in categorizing. However, the usual picture superiority in categorization tasks is not disputed by these data.

In the German language, the problem of visual similarity in categorization tasks was tackled by Hoffmann and Klimesch (1984). In preliminary studies (e.g., Hoffmann & Ziessler, 1982), the subjects had to name characteristic features for the elements of an item pool which consisted of common nouns that belonged to subordinate, basic, or superordinate level in Rosch's (1975) terms. Two main results were obtained: (1) about 30% of the responses were concrete properties (e.g., tree: green, large, or has leaves), the rest were abstract (e.g., tree: solid, of wood, or shelter from rain); and (2) the proportion of concrete to abstract features decreased with an increasing level of a concept. Furthermore, the authors introduced a terminology which is different from that of Rosch (1975) to characterize three-level hierarchies: the lowest level is called sublevel, the highest superlevel; the medium level lies between them. Thus the superlevel of categoriz-
ing experiments usually contains categories like *musical instrument*, *vehicle*, or *weapon*. These examples include *string* and *wind instrument*, *street*, *rail*, *air*, and *water vehicle*, or *stabbing* and *firearm* as medium-level categories. *Violin*, *trumpet*, *car*, and *dagger* are examples of corresponding sublevel nouns.

In their experiments, Hoffmann and Klimesch (1984) used three-step hierarchies with sublevel, medium-level and superlevel concepts. The preliminary data (Hoffmann & Ziessler, 1982) gave strong evidence that three types of hierarchies were to be distinguished:

1. In the *sensory superlevel* hierarchies, concrete features dominated from sublevel up to superlevel concepts. For example, the superlevel concept *tree* shares with the medium-level concepts *deciduous tree* and *coniferous tree* and with sublevel concepts like *oak*, *birch*, *pine*, and *fir* the characteristic visual features of trees.

2. In the *sensory medium-level* hierarchies, salient visual features comprise sublevel and medium level as, for example, *dagger* and *sword* as *stabbing arms*, and *colt* and *gun* as *firearms*, whereas the superlevel concept *weapon* is characterized by abstract properties.

3. In the *sensory sublevel* hierarchies, eventually, concrete, visual features only discriminate among sublevel objects, whereas the medium level as well as the superlevel are characterized by abstract features. It is noteworthy that Hoffmann and Ziessler (1982) found *food* to be a sensory sublevel hierarchy with the abstract medium-level categories *vegetables*, *fruits*, and *baked goods*.

Based on this preliminary work, Hoffmann and Klimesch (1984) replicated the category-matching experiment of Pellegrino et al. (1977) with picture–picture, picture–word, and word–word stimulus pairs that were administered to three different groups of subjects. The stimuli were selected at sublevel and were to be categorized, as usual, at superlevel. However, the type of hierarchy was strongly controlled as an independent variable. On the average, picture–picture pairs had a latency of 886 ms; picture–word pairs exhibited longer responses by 285 ms, and word–word pairs by 624 ms. That replicates the usual order of the differences. The relatively long latencies are due to seven different categories (instead of two in other experiments; e.g., Harris et al., 1977; Pellegrino et al., 1977; Snodgrass & McCullough, 1986), and the fact that the subject did not know the stimuli in advance and that each concept was exposed only once. The “yes” responses for the picture–picture pairs exhibited clear effects of type of hierarchy. The pairs of pictures that belonged to the *sensory superlevel* hierarchies were classified with a mean latency of 729 ms. The instances of the *sensory sublevel* hierarchies, on the other hand, required an average of 1048 ms. Most interesting was the *sensory medium-level* hierarchy. If both stimuli were instances of the same medium-level category, then the mean latency was 728 ms; if they belonged to different medium-level categories, they required 999 ms. Essentially the same results were
obtained with picture-word stimuli, whereas the word-word stimuli did not show any significant effect of hierarchy type.

The authors interpret this result within the framework of the common code hypothesis with privileged access for pictures: the internal concept code contains concrete and abstract properties. If the access goes through pictures, then the concrete properties are activated first. Thus, if the concrete properties of a pair of pictures are sufficient for categorizing, then a fast response occurs. On the other hand, if the later activated abstract properties are necessary for categorizing, then the response is delayed. Picture-word stimuli exhibit the same pattern of results for the same reason, that is, because one picture stimulus is sufficient to start the internal categorizing with the concrete properties. In contrast, word-word stimuli activate concrete and abstract properties about simultaneously so that categorizing responses due to concrete properties are not faster than those due to abstract ones. However, there are counterexamples in the literature to the latter conclusion, although it is clearly justified by Hoffmann and Klimesch's data. In a priming study with reading and lexical-decision tasks, Flores d'Arcais, Schreuder, and Glazenborg (1985) obtained data which suggest that also verbal stimuli activate the perceptual properties of a concept faster than the abstract and functional ones.

However, from the work of Hoffmann (e.g., Hoffmann & Klimesch, 1984) and of Flores d'Arcais (e.g., Flores d'Arcais & Schreuder, 1987) it is obvious that perceptual and functional properties of objects play different roles in categorization. Further research is required to clear the discrepancies. It is noteworthy that the results and conclusions of Hoffmann and Klimesch (1984) do not contradict those of Rosch (1975). In Rosch's terms, Hoffmann and Klimesch's types of hierarchies are characterized by the position of the basic-level concepts: they can be placed at superlevel (e.g., tree in the hierarchy of oak, deciduous tree, tree), medium level (e.g., stabbing arm in the hierarchy of sword, stabbing arm, weapon), or sublevel (e.g., apple in the hierarchy of apple, fruit, food).

So far, we are inclined to take for granted that the mean latencies increase monotonically from word reading over picture naming and picture categorizing up to word categorizing. The first of these latency differences, that between reading and naming, is beyond dispute. The second one, the categorizing—naming difference for pictures, was clearly demonstrated, but by a lesser number of experi-

---

1 An anonymous reviewer emphasized that these comparisons of response times for verbal and pictorial stimuli are valid only if the two sets of materials require about the same time for the early visual-recognition stage. Of course, this time depends essentially on the kind of drawing of the pictures and on the type font of the words. Thus, careful experimenters match their verbal and pictorial materials for tachistoscopic threshold (e.g., Potter & Faulconer, 1975) or for recognition time (e.g., Potter, Kroll, Yachzel, Carpenter, & Sherman, 1986). On the other hand, there are no hints that this condition could be seriously violated if clear outline drawings in the style of those by Snodgrass and Vanderwart (1980) and functional fonts (e.g., Courier, Helvetica) are used in sufficient contrast and size.
ments (e.g., Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Irwin & Lupker, 1983; Segui & Fraisse, 1968; Smith et al., 1978). It decisively supports an important supplement of the common abstract code hypothesis: the basic level provides the entry points of the semantic system for pictures (Glaser & Glaser, 1989, pp. 30, 32), and categorization as well as other semantic tasks are mediated by internal concept nodes at basic level. In other words, there is no semantic processing of pictures that circumvents this level, although that could be an attractive theoretical alternative: in categorizing, it could well be that the pictures are only partially analysed up to the point at which the category, for example furniture with the picture of a chair, is recognized. If that were the case, then the categorizing–naming difference should show the reverse direction: pictures should be categorized faster than named.

The third difference, that between word and picture categorizing, was more often demonstrated — and disputed. It provides the most conclusive evidence that category membership is stored in an internal abstract code which is accessed faster by pictures than by words. Thus, we have to discuss the recent criticism by Theios and Amrhein (1989). These authors also adhere to the hypothesis of one abstract semantic code as a basis for semantic processing, but they deny a privileged access of pictures to that code. They argue that the often observed picture superiority is an artifact due to the greater viewing angle for pictures than for words in these experiments. Surely, Theios and Amrhein (1989) are right that stimuli with greater size or energy are processed faster. But does this really justify the conclusion that the word–picture difference in categorizing “disappears when the pictorial stimuli are controlled for size, area, and featural line width” (Theios & Amrhein, 1989, p. 22)?

In their Experiment 1, these authors obtained a mean reading–naming difference of 160 ms for word and picture stimuli that were equalled for averaged visual width on the three levels of 1.5, 3.0, and 6.0 degrees. The means for reading were 640 ms, 591 ms, and 596 ms, and those for naming were 819 ms, 750 ms, and 738 ms, respectively, for the three size conditions. As concepts, five geometrical figures such as circle, square, and so on were used. The authors conclude that their reading–naming difference of 160 ms is an estimator for this difference if words and pictures are equal in size. They argue that the usually observed reading–naming difference is reduced with regard to its true value because cognitive picture-processing times are confounded with an artifactual acceleration that results from picture stimuli being far greater than word stimuli. For example, if their reading time for the small words is compared with their naming time for the large pictures, the reading–naming difference is reduced to 98 ms. In other words, the artifactual component amounts to 62 ms.

After that, Theios and Amrhein (1989, p. 11) take a reading–naming difference from Fraisse (1969) that amounts to 90 ms and argue that this value would approach the “true” 160 ms if it were corrected for an artifactual loss of 62 ms. Of
course, this is a fascinating result with these numbers; however, it cannot be generalized. Most reading–naming differences as reported in the literature are greater than 90 ms; moreover, they often exceed these authors' "true" value of 160 ms (cf. the values quoted in the first paragraph of this article). Fraisse (1960) himself reports 219 ms as reading–naming difference for a set of four geometric figures like those of the present authors. Furthermore, if the word width exceeds 3 degrees, then the artifact practically disappears in the data of Theios and Amrhein (1989). Thus, there remains the fact that the reading–naming difference may be artifactually reduced in empirical data if pictures and words are not equalled in size, but only if the word stimuli were less than 3 degrees wide.

It is our main question whether or not these results of Theios and Amrhein (1989) invalidate the naming–categorizing difference for pictures and the word–picture difference in categorizing. Obviously, the naming–categorizing difference is completely independent from these data because under both conditions pictures of the same size are used. The word–picture difference in categorizing, on the other hand, would be jeopardized. Because word categorizing is longer than picture categorizing, this difference would be artifactually increased, or, in the worst case, brought about, if picture processing were accelerated by an artifact. This is far more critical than the attenuated reading–naming difference.

However, according to present knowledge, we can argue as follows: Theios and Amrhein (1989) could not establish the artifact if the word stimuli were more than 3 degrees wide, disregarding picture size. Therefore, the observed word–picture differences in categorizing tasks can be taken for granted, if the word components of the stimuli fulfilled this condition. That is undoubtedly the case in the studies of Glaser and Düngelhoff (1984; width of the five-letter word 4 degrees), Glaser and Glaser (1989; 5 degrees), Pellegrino et al. (1977; 14 degrees; my calculation, W G.), Rosch (1975; 7 degrees for word pairs).

The conclusion is that the hypothesis of the faster access to the abstract semantic system for pictures rather than for words is not affected by the criticism of Theios and Amrhein (1989). The conviction of these authors that the often observed faster semantic processing of pictures than words results only from the usually different size of picture and word stimuli in those experiments is not justified by the available data. Of course, their claim for a control of this variable in further experiments is worth following (e.g., Snodgrass & McCullough, 1986; Bajo, 1988).

So far, the essential results were discussed that follow from a comparison between naming pictures at basic and at superordinate level. Converging evidence for the theoretical conclusions was sought from the comparison with word naming at both these levels and with matching tasks. We will now turn to the priming studies.
5. Picture naming in priming studies

The term *priming* designates two matters: an experimental paradigm and an effect that is often observed in these experiments. In principle, the target stimulus or stimulus pair in a reading, naming, categorizing, or matching task is preceded by another stimulus, the prime. The main independent variables are the instruction concerning the prime and the semantic or associative relatedness between prime and target. The instructions go from ignoring the prime over using its information for a more efficient response to the target up to responding in a similar way as to the target. The usual results show that the response latencies to the target are gradually reduced depending on prime–target similarity and on depth of processing of prime and target (Irwin & Lupker, 1983). This effect is called *priming*. Theoretically, it is explained as follows. The voluntary or involuntary processing of the prime activates internal nodes or links so that they are more readily available for the processing of a related target. Thus, the network along which internal codes activate one another should be mapped onto a pattern of observable priming effects.

Indeed, a large number of priming studies were undertaken to investigate the structures of long-term memory. We will give a short review of only the most important ones that concern picture naming. As in the last paragraph, we include studies of reading as well as of categorizing and matching pictures and words. On the other hand, we leave aside the numerous investigations that used the lexical decision task.

One of the first priming studies in picture naming was published by Sperber, McCauley, Ragain, and Weil (1979). In a picture–picture condition, the subjects had to name prime and target at basic level. The response to the prime triggered an interstimulus interval of 1000 ms followed by the target. If both pictures were of the same superordinate category, the response latency to the target was 51 ms shorter than otherwise. In one additional condition, the target pictures were blurred by defocusing the projection lens. Now, they were primed by 112 ms. That indicates that semantic priming has a perceptual component, at least if the targets are visually degraded (cf. Reinitz, Wright, & Loftus, 1989). In the word–word condition, when subjects had to read aloud prime and target, a priming by 19 ms was obtained if the targets were in focus, and by 56 ms if they were degraded. This is in accordance with the often quoted results by Meyer, Schvaneveldt, and Ruddy (1975). Following the Sternberg (1969) logic, that means that both the visual quality of the stimulus and the semantic relatedness of the prime influence at least one common stage in the reading or picture-naming pathway.

In Experiment 3 by Sperber et al. (1979), the four stimulus conditions
word-word, word-picture, picture-word, and picture-picture were administered in mixed blocks. Compared to the blocked conditions of word-word and picture-picture pairs, the priming effects were considerably reduced. Reading was primed 10 ms by word and 8 ms by picture stimuli, picture naming 31 ms by picture and 13 ms by word stimuli. These data show that there was within-modality as well as, though to a lesser degree, between-modality priming. The picture was more effective as prime and more susceptible to priming as target.

McCauley, Parmelee, Sperber, and Carr (1980) and Carr, McCauley, Sperber, and Parmelee (1982) conducted similar experiments. As additional independent variables, (1) exposure duration of the prime with the three levels below threshold, at threshold, and suprathreshold, and (2) interstimulus interval from prime offset to target onset with two levels of 90 ms and 490 ms were introduced. The four modality conditions picture-picture, picture-word, word-picture, and word-word were blocked. For the target, a speeded response was required, whereas the prime was to be named after the target. Averaged over threshold conditions and interstimulus intervals, the pattern of results was very similar to that from Sperber et al. (1979), but the priming effects were about twice as large: 23 ms (word-word), 15 ms (picture-word), 36 ms (word-picture) and 54 ms (picture-picture). Apparently, the blocked variation of the modality condition and, perhaps, naming the prime after the target, raised the priming effect. Again, picture naming was particularly prone to be primed and pictures were more effective as primes than words. Thereby, the primes that were presented below threshold exhibited essentially the same effects. The authors conclude that the semantic priming does not depend on conscious identification of the primes by the subject. However, that was disputed by Merikle (1982) and by Purcell, Stewart, and Stanovich (1983). We will not pursue this discussion (see, for example, Hines, Czerwinski, Sawyer, & Dwyer, 1986). Generally, the authors favour the common-code hypothesis and assume a functional closeness of pictures to that code: “Stimuli that are more similar to semantic representations will prime more effectively because they are able to activate semantic representations more rapidly. Such stimuli will also be more primeable as targets because activated semantic representations will make greater or more efficient contact with target stimuli that are more similar to them” (Carr et al., 1982, p. 772).

A similar study, which used the repetition-priming technique and included categorizing responses, was conducted by Durso and Johnson (1979). Ninety concepts at basic level, together with additional filler items, were presented in random order. In every trial, one concept was given. In mixed blocks, one half of the stimuli were words, the other half were line drawings. The reading-naming group of subjects had to name the pictorial and to read the verbal stimuli. The categorizing group had to categorize the stimuli as natural or man-made by pressing keys. The priming conditions were created by repeating some of the stimulus concepts with a controlled distance to their first presentation. This
repetition was factorially crossed with stimulus modality, so that picture–picture, picture–word, word–picture, and word–word presentations of the same concept emerged. In the picture–picture condition, two different pictures of the same object were used in order to avoid "physically same" matches.

In the reading–naming group, the same rank order of priming effects was found as in the studies by Sperber et al. (1979) and by Carr et al. (1982): 32 ms (word–word), 15 ms (picture–word), 109 ms (word–picture), and 175 ms (picture–picture). Thus, the high primeability of the picture-naming response was demonstrated again. On the other hand, the high efficiency of the picture prime is almost completely restricted to picture naming and does not work in reading. Thus, the authors conjecture "that a word trace contains information capable of aiding subsequent processing of both pictures and words, whereas picture traces are more specific in that they aid the processing of pictures to a large extent while aiding the processing of subsequent words to a small extent" (Durso & Johnson, 1979, p. 457). Unfortunately, that is scarcely more than a paraphrase of the data. Similar data which also showed that using a word as a response in a naming task does not prime later recognition of this word were also found by Scarborough, Gerard, & Cortese (1979).

The categorizing task yielded priming effects of 90 ms (word–word), 45 ms (picture–word), 27 ms (word–picture), and 51 ms (picture–picture). They were generally lower than in the reading–naming task, and now picture processing was markedly less primed. It is difficult to integrate these data theoretically. The authors sketch a "generic-specific" hypothesis. According to that hypothesis, a word stimulus should activate rather general features of its concept, whereas a picture should activate more specific aspects (see also Lupker, 1988). Furthermore, semantic excitation should spread rather from the general to the specific than in the opposite direction. That would explain that picture primes had practically no effect on word reading. The categorizing task, on the other hand, should force the activation of these specific properties that are important for the decision between natural and man-made objects. Because that, in particular, is difficult for word stimuli, priming should become effective. Durso and O'Sullivan (1983) tried to support this hypothesis by further experiments, but essentially it remains ad hoc. A superior hypothesis states that naming a picture can only prime a later reading of the picture's label if the subject's strategy connects reading of the word stimulus with semantic processing (Bajo, 1988).

Irwin and Lupker (1983) aimed at a systematic comparison of the effects of prime and target modality (word or picture), the task concerning the prime (categorizing, reading–naming, or naming the colour in which the prime is printed), and the task concerning the target (reading–naming, categorizing). As usual, prime and target modality were factorially crossed (picture–picture, picture–word, word–picture, and word–word). These four conditions were presented in mixed blocks, whereas the task concerning prime and target was
blocked. As materials, 72 concepts at basic level from six superordinate categories were used.

The results were as follows. When the target had to be read or named, there was no significant facilitation by a picture or word prime regardless of whether the prime had to be categorized, named or its colour had to be named. When the target had to be categorized, words and pictures were primed to an equal degree and the amount of facilitation depended on the task concerning the prime. Related primes facilitated by 231 ms, if they were to be categorized; by 119 ms, if they were to be named; and by 57 ms, if their colour was to be named. Thus, the degree of semantic processing of the prime was the only independent variable that exhibited graded facilitatory effects on the latency in target categorizing. The authors interpret their results within the framework of the semantic network model by Collins and Loftus (1975) and the model of a common code with logogen and iconogen subsystems by Seymour (1973). Thus, the observed priming in the categorizing task results from deeper target processing in the common code system. The dependency of this effect from the priming task is explained in the same way by different levels of prime processing. However, it remains unexplained why the naming of target pictures was not primed, not even by picture primes. This result is contrary to almost all other data. One hypothetical reason may be that every stimulus concept was presented four times to each subject. However, if the fourfold repeated use of the same stimulus concepts would "wash out" the priming effects, then at least the first presentation of each stimulus should exhibit priming, but this, too, was not the case in the data of Irwin and Lupker (1983).

Bajo (1988) conducted a comprehensive study in order to clarify the inconsistencies of the results of Sperber et al. (1979), Carr et al. (1982), Irwin and Lupker (1983), and Durso and Johnson (1979). Furthermore, Bajo (1988) sought priming data that contributed to the decision between the dual-code and the common abstract code hypotheses of semantic memory. As stimulus materials, a pool of 128 picturable concepts at basic level were chosen in such a way that always two of them were taxonomically related. Relatedness was widely defined as semantic association (e.g., cat–dog) or as membership in categories of a rather low superordinate level (e.g., part of the face: nose–eye; part of the body: arm–leg). Thus, 64 different category terms were available for the categorizing task, and each instance of a category was considered to be an effective prime if the other instance was the target. In the unrelated priming condition, instances of two different categories were combined.

In Experiment 1, a name-verification and a category-verification task were given. The modalities of primes and targets were factorially crossed so that picture–picture, picture–word, word–picture, and word–word pairs resulted. The subjects in the name-verification group were instructed to answer the question
“Does the stimulus name match the name of the concept . . . ?” for the target, disregarding the prime, by pressing a “yes” or “no” key. The subjects in the category-verification group heard the question “Does the stimulus belong to the category . . . ?” Pictures and words were equalled for visual angle at a width of 3 degrees. Thus, Theios and Amrhein’s (1989) objection does not apply. The prime was visible for 1000 ms and then followed by a 50-ms visual noise mask and the target. The prime-target relation (related, unrelated), the primc-target modalities (picture-picture, picture-word, word-picture, word-word), and the responses (yes, no) were factorially crossed and varied within subjects at random from trial to trial.

The results exhibited the rank order of mean latencies from verifying words (538 ms) over verifying pictures (559 ms) and verifying categories of pictures (625 ms) up to verifying categories of words (644 ms). Thus, they mirrored the reading-naming difference, the naming-categorizing difference for pictures, and the word-picture difference for categorizing as discussed in the last paragraph. Of course, the reduced amount of the word-picture difference in verifying, compared to reading and naming, is due to this task. Most interesting are the priming results. Firstly, there was no effect of prime modality, that is, word and picture primes had always the same effects. Secondly, the priming effect was about equal in verifying names of pictures (101 ms) and in verifying categories of pictures (112 ms) as well as of words (100 ms). On the other hand, verifying names of words showed only a small, residual effect of 11 ms from picture and 13 ms from word primes.

The common code hypothesis accords with these data. Verifying the name of a picture as well as verifying the category of a picture or a word requires semantic processing of the target that takes place in the abstract system. Thus, it is plausible that the subjects translate not only the target, but also the prime in its abstract semantic code, regardless if it is a word or a picture. In this system, processing of targets that are related to the primes takes advantage of codes that are pre-activated by the primes. Verifying the word, on the other hand, requires a match between spoken and printed modality of the word symbol that is carried out in the verbal system without semantic processing. Of course, this interpretation of the results provides new, independent evidence for the lexical hypothesis which postulates an extended system for word processing without semantic components.

However, there are two possible criticisms. Firstly, if the picture has privileged access to the semantic system, why, then, did picture primes have no greater effect than word primes? A plausible answer is that the SOA of 1050 ms was sufficient even to complete the slower semantic interpretation of the word primes. Secondly, the lack of facilitation of the word targets in the name verification task could result from functional differences between this task and the reading task.

In order to test this objection, Bajo (1988) carried out Experiment 2. Now, the
usual reading and picture-naming instructions were given and two additional independent variables were introduced:

(1) Type of instruction: half of the subjects were instructed to try using the prime information as far as possible to process the target more efficiently; half received a neutral instruction as in Experiment 1.

(2) For half the subjects, the modality conditions of picture–picture, picture–word, word–picture, and word–word were randomized between trials as in Experiment 1; for the other half of subjects they were blocked.

These two factors were crossed between subjects; all other factors were varied within subjects. The results exhibited the following pattern of facilitations. Again, naming picture targets was on the average 100 ms facilitated by related primes, disregarding modality of the prime, instruction, and mixed or blocked trial sequence. On the other hand, word targets now also showed a priming by 50 ms under “semantic” instruction and mixed presentation and by 48 ms under neutral instruction and blocked presentation. “Semantic” instruction together with blocked presentation yielded 65 ms facilitation. No priming effect (7 ms) occurred only if the neutral instruction was combined with mixed presentation.

The interpretation is clear: the task to read words aloud is usually carried out without activation of the word’s meaning and thus there is no effect of a word or picture prime. On the other hand, the subjects can voluntarily use semantic information also in word reading if they are instructed to do so or if there is no uncertainty about the modality of prime and target because modality is kept constant throughout blocks of trials. Now, semantic processing takes place as is indicated by the observed priming. Taken together, this result indicates an asymmetry of access to the common abstract code for pictures and words. Whereas pictures always contact the semantic code, words do only if the task requires semantic processing or if the subjects choose a strategy to use semantic information. This leads unambiguously from the common code hypothesis to the lexical hypothesis as presented in our second paragraph.

Two important further variables were introduced in the priming task with picture–picture and word–word pairs by Huttenlocher and Kubicek (1983). Firstly, these authors asked whether the priming effects arise automatically from the repeated activation of semantic features that were common to prime and target, or from the controlled expectation of a target that is semantically similar to the prime. Since Posner and Snyder (1975), this question has been investigated by varying the predictive validity of the prime. Thus, a low-expectancy group of subjects was given 12.5% semantically related prime–target pairs in the trial sequence, whereas the high-expectancy group received 87.5% related pairs. Secondly, the influence of word frequency on the naming and reading latencies as well as on the priming effect was evaluated. The modality of the stimulus pairs, picture–picture or word–word, was blocked, and there were no picture–word or word–picture pairs.
In picture naming, related targets exhibited a mean latency of 765 ms in the low-expectancy and of 781 ms in the high-expectancy group. The responses to the respective unrelated targets were 59 ms longer under low and 175 ms longer under high expectancy. The authors conclude that the first of these numbers represents an automatic facilitation by similarity between prime and target that is also contained in the second number as a component. The far greater amount of the second number is explained as an effect of “surprise” in those rare trials in which the target was not similar to the prime in the high-expectancy group. Although the authors do not explicitly mention this source, that result fits perfectly the inhibition that Posner and Snyder (1975) found if controlled expectation was directed to a stimulus other than the presented one. The responses to high-frequency pictures were 116 ms faster than to low-frequency ones, but the interaction of frequency and relatedness was not significant. In word reading, high-frequency words were read by 24 ms faster than low-frequency words, but there were only marginal priming effects in all conditions.

Huttenlocher and Kubicek (1983) divide the processing chain of picture naming into four parts: visual processing, activation of the concept, retrieving of its name, and articulation. Given this, the results were interpreted as follows:

(1) In picture naming, there is an automatic priming by similarity between prime and target of about 59 ms. If a controlled expectation is induced, then processing of unexpected targets is inhibited by about 116 ms, a result which is in line with theory and data by Posner and Snyder (1975). Because these effects are not obtained in reading, they can be attributed only to the first three stages.

(2) Because object decision tasks concerning pictures (cf. Kroll & Potter, 1984) showed priming effects in the same magnitude without naming responses, the name-retrieving stage is also eliminated as a source of this effect, and there remains only visual processing or activation of the concept.

(3) The large effect of name frequency should arise in other stages than visual processing or concept activation, because there was no interaction between this factor and relatedness. Because frequency effects are far less in reading, the articulation stage is also eliminated and only the name-retrieval stage remains. In short, priming effects originate in concept identification, frequency effects in name retrieval. However, there remains the question why some authors found priming effects in reading. Huttenlocher and Kubicek (1983) hypothesize that in the reading task access to the articulatory plan and to the meaning are independent from one another. If easily readable words are to be spoken aloud, then their meaning plays no role. On the other hand, different experimental conditions can activate the access to meaning and thus render reading primeable. Visual degradation of word stimuli (Meyer et al., 1975; Sperber et al., 1979) seems to be such a condition.

Taken together, the priming study by Huttenlocher and Kubicek (1983) also provides evidence for the hypothesis of a common semantic code together with a
word system that can operate uninfluenced by word meanings. This is the lexical hypothesis of our section 2.

6. Stroop-like interference

The reading–naming difference as introduced in our first section was often investigated with colour words and colour patches as stimuli. Stroop (1935) looked for an explanation in terms of associative interference: every object, including colours, can be named in several ways, and these different naming tendencies could inhibit one another, whereas a reading response is uniquely determined by the print. This gave him the idea to add intentionally an irrelevant response tendency to the colour-naming task by writing a different colour word using the colour that is to be named, for example the word red in GREEN ink. The idea led to a discovery: the latency to name the colour of such a stimulus is seriously increased compared to naming a colour patch, and the subjects show strong signs of stress and effort. The rise of naming latency due to the irrelevant word is called Stroop interference. Often, it amounts to more than 80–100 ms and sometimes it reaches the reading–naming difference. Thus, it is one of the most marked effects in cognitive psychology. It is very reliable and, because of its magnitude, easy to replicate.

To read the word component of a Stroop stimulus, on the other hand, is practically not longer than to read a word that is printed in black. This difference, the serious impact of the irrelevant word on colour naming and the immunity of the reading response against the irrelevant colour, is called the asymmetry of the Stroop effect.

From the beginning, Stroop research had two aims which were tightly interwoven: to look for an explanation of this strong effect, and to use this effect as a tool to investigate reading, naming, and selective attention. (For comprehensive reviews see Dyer, 1973; Glaser & Glaser, 1989; MacLeod, 1991.) In this article, we confine ourselves to applications of the Stroop technique to picture naming and picture categorizing. Our main interest concerns the question how far Stroop interferences can provide evidence in favour of the particular hypotheses given in the second section.

Glaser and Glaser (1989) present a general characterization of the Stroop experiment. They started from two premises:

(1) The colour of the Stroop stimulus is the limiting case of a picture. Therefore, colour–word/colour and word–picture interference are both instances of a general reading–naming interference.

In the following text, concepts as represented by colours, pictures, or internal concept codes are given in capitals; words that are meant as physical symbols, stimuli, or internal word codes are given in italics.
(2) The word and the colour or picture, respectively, of the Stroop-like stimulus can be considered as two stimuli. That is, the Stroop stimulus is the limiting case of double stimulation (Kantowitz, 1974) in which both stimuli are spatially integrated and temporally synchronized.

A further characteristic of the Stroop experiment is the instruction. The subject has to respond to one stimulus component, the target, according to a certain rule. The usual instructions demand reading, naming, categorizing, or some kind of matching or semantic decision. The other stimulus component, the distractor, must be ignored. Therefore, the Stroop experiment is similar to the priming experiment, although the instruction to ignore the prime is given only rarely.

In the Stroop experiment, the following independent variables play a role:

(1) Stimulus modality: the conventional Stroop stimulus is modally mixed, that is, it contains a pictorial and a verbal component. Modally pure word-word stimuli were used by Dallas and Merikle (1976a, 1976b), Warren (1977), Shaffer and LaBerge (1979), and La Heij et al. (1985). The systematic investigation of Stroop effects with word-word and colour-colour stimuli, compared to conventional colour-word stimuli, began with the studies by Van der Heijden (1981) and Glaser and Glaser (1982). Stroop effects of picture-picture stimuli were investigated by Glaser and Glaser (1989). The main results of these studies are that the interference fails to appear only if the word of a modally mixed stimulus is to be read. Modally pure, that is colour-colour, picture-picture, and word-word stimuli, always exhibit Stroop interference as well as modally mixed stimuli in the naming task. The categorizing task, on the other hand, reverses this result in a dramatic way. Word-word and picture-picture stimuli show again the interference (Glaser & Glaser, 1989), with modally mixed stimuli now word processing is disturbed by adequate picture distractors, whereas picture categorizing becomes immune against distracting words (Smith & Magee, 1980; Glaser & Dünkelhoff, 1984).

(2) A central independent variable in Stroop experiments is distractor-target pairing. It determines essentially the amount of interference. In the congruent condition of this variable, both stimulus components match at basic level (e.g., RED-red, HOUSE-house). Usually, the congruent distractor facilitates the processing of the target, whereby picture targets are more facilitated than word targets (e.g., Glaser & Glaser, 1989, Experiment 6). That is in accordance with the results from priming studies as discussed in the previous section. The control condition is created by using a neutral distractor, that is, a non-word, non-colour or non-picture. In Stroop research, this condition provides the usual baseline for the evaluation of the distractor effects, whereas only a part of the priming studies contain such a non-word control. Incongruent distractor-target pairings cause the usual Stroop inhibition (e.g., RED-green). In picture naming, they show a semantic gradient: naming a picture is more disturbed by a distractor from the
same superordinate category (e.g., HOUSE–castle, same category BUILDING) than by a distractor from another category (e.g., HOUSE–fish, different category ANIMAL; Glaser & DÜngelhoff, 1984; Glaser & Glaser, 1989; Guttentag & Haith, 1978; Lupker, 1979). It is noteworthy that a mere associative relatedness between distractor and target (e.g., MOUSE–cheese) does not cause the rise of interference that results from common membership in a superordinate category (Guttentag & Haith, 1978; Lupker, 1979). In the categorizing task, interference is observed if the instruction concerning the target would lead to a different response if it would be applied to the distractor. That is, if the category name "building" is to be said as response to the target word house, then the distracting picture of a FISH (different category ANIMAL) strongly interferes, whereas that of a CASTLE (same category BUILDING) does not (Glaser & DÜngelhoff, 1984).

(3) Despite its semantic components, the Stroop interference is very sensitive to variations of the spatial configuration of distractor and target. Maximum interference is obtained if both components are spatially integrated, for example if the colour is used to print a different colour word, or if the picture surrounds the word, as is the case in the usual picture–word stimulus pairs. On the other hand, the facilitation in the congruent condition is less dependent on stimulus geometry. We will not pursue this complicated matter further here (see Glaser & Glaser, 1989, p. 14).

(4) In the priming studies, the prime usually precedes the target by about 400–1000 ms. In the Stroop studies, the distractor is most often presented synchronously with the target. However, there are some investigations that systematically varied the SOA between distractor and target within a range from about −500 ms (distractor pre-exposed) to about +400 ms (distractor post-exposed) in order to trace the time course of the interference (Dyer, 1971; Flowers, 1975; Flowers, Nelson, Carson, & Larsen, 1984; Glaser & DÜngelhoff, 1984; Glaser & Glaser, 1982, 1989; Goolkasian, 1981; Neumann, 1980; Posner & Snyder, 1975; Warren, 1977).

In the resulting SOA functions, three characteristic components are observed:

(1) The facilitation due to congruent distractors is often weak at SOA = 0 ms and shows a flat maximum at distractor pre-exposure by 200–400 ms. Thus, facilitation is a slow effect. It can contain automatic and controlled components (Taylor, 1977). This facilitation in the Stroop experiment seems to be the same effect that is observed in priming studies.

(2) There is an inhibition that has a similar time course as the facilitation, that is, it also shows a flat maximum at distractor pre-exposure of 200–400 ms and more. It seems to be the inhibitory counterpart of the facilitation, and it occurs when an unexpected target is presented after the voluntary build-up of a certain expectation by the subject (cf. Glaser & Glaser, 1982; Posner & Snyder, 1975; Taylor, 1977). Thus, it is essentially a controlled effect.
(3) The Stroop inhibition shows a strong maximum within a small SOA window around synchrony of distractor and target; usually, it is observed from distractor pre-exposure of 100 ms over simultaneous exposure of distractor and target up to distractor post-exposure of 100 ms. Thus, it is a fast effect. It is essentially automatic because it occurs without conscious strategies, but it is intensified by voluntarily directed expectations.

Almost all theoretical explanations share the assumption that the distractor in the Stroop task is involuntarily processed up to a certain stage. The response to the target is delayed because the internal distractor signal absorbs mental capacity or hampers the accumulation of evidence in favour of the target or its response. The main question for any Stroop theory arose from the asymmetry, the immunity of the reading task against an interfering picture, and the immunity of the picture-categorizing task against an interfering word. For a long time, the horse-race or relative-speed hypothesis was most widely accepted (Dyer, 1973; Morton & Chambers, 1973; Palef & Olson, 1975; Posner & Snyder, 1975; Warren, 1972, 1974): the internal signals from distractor and target are processed in temporal proximity and in parallel; interference occurs only if the distracting signal "wins the race", that is, reaches a certain stage before the target's signal. Thereby, the critical stage is captured by the irrelevant signal and released for target processing only after a refractory period.

However, the relative-speed hypothesis of the Stroop interference was unambiguously rejected by the studies with SOA variation (Glaser & Dünkelhoff, 1984; Glaser & Glaser, 1982, 1989; Neumann, 1980): they showed that it is impossible to disturb the reading response by giving the colour or picture distractor a head start. According to a similar logic, Dunbar and MacLeod (1984) demonstrated that the Stroop asymmetry is preserved even if the words of the Stroop stimuli are transformed so that reading becomes longer than naming. Thus, Glaser and Glaser (1989) concluded that occurrence or absence of the Stroop inhibition does not depend on the speed with which target and distractor are processed. Rather, the Stroop interference should exhibit the degree to which one pathway from perception to action (e.g., from printed word to reading it aloud) is functionally privileged compared to another one (e.g., from picture to pronouncing its name). Thus, the pattern of Stroop interference in adequate series of experiments should provide information on cognitive structures that go beyond that from basic reaction times and from priming data.

Now, let us have a look at Stroop-like results with picture naming. The first investigations were carried out by Rosinski, Golinkoff, and Kukish (1975), Ehri (1976), and Rosinski (1977) using stimulus matrices. These studies aimed at the development of reading skills in children. The main idea was that the degree to which the word stimuli interfere with picture naming should indicate the level of reading ability. The results showed that the interference was very high after the children read fairly well as second graders. Later on, the interference decreased
steadily up to adulthood. Underwood (1976) used the picture-word interference to explore attentional effects in reading of adults.

Lupker (1979) carefully investigated the influence of several independent variables:

1. A non-word distractor (e.g., MOUSE–wydem) prolonged the naming latency by 65 ms, compared to the lack of any distractor.
2. A neutral word without any relation to the target (e.g., MOUSE–hand) inhibited 21 ms more.
3. The same effect was produced by a distractor with an associative relation to the target (e.g., MOUSE–cheese). This means that associative relatedness has no particular influence on the Stroop interference.
4. If the distracting word belonged to the target picture’s superordinate category (e.g., MOUSE–dog), then the interference increased by an additional 31 ms.
5. Whether the distractor had a high or low typicality as an instance of the target’s superordinate category (cf. Rosch, 1975) did not influence the amount of interference.
6. A distracting word without semantic or associative relatedness to the target inhibited target naming by 24 ms more if it had a high imageability (e.g., BUTTERFLY–wspaper) than if its imageability was low (e.g., BUTTERFLY–law).

In a very accurate study, La Heij et al. (1990) conducted a time-course analysis of the effects on picture naming of a common category membership of distractor and target on the one hand and of associative relatedness between them on the other hand. The results demonstrate that common category membership produces the fast Stroop-like inhibition at SCAs from 0 ms (synchrony) to +150 ms (slight post-exposure of the distractor; cf. Glaser & Düngelhoff, 1984), whereas associative strength facilitates markedly with pre-exposed distractors from SOA = −800 ms up to −400 ms. Because the two variables are confounded in most priming and Stroop studies, these results contribute importantly to a solution of the fundamental question why “semantic relatedness” facilitates in priming and inhibits in Stroop experiments.

Taken together, these results show that the Stroop interference is a sensitive indicator of cognitive processes which responds differentially to different independent variables: associative relatedness between distractor and target as well as typicality of the distractor show no inhibition, whereas membership in the target’s superordinate category and imageability are very effective. Lupker (1979) supposes that this interference originates in a semantic common code system. Its partially hierarchical structure (see Collins & Loftus, 1975) should cause the effect of common category membership of distractor and target. The stronger effect of highly imageable distractor words should result from their easier access to the semantic system.
Lupker and Katz (1981) investigated Stroop-like effects in a semantic decision task. The targets were a set of (1) different pictures of a dog, (2) pictures of four-legged animals that were not dogs, and (3) pictures of inanimate objects. The subjects had to decide whether the target represented a dog or not. In the control condition, there was no distractor word. The “yes” responses were facilitated 7 ms by the distractor word dog. A non-word distractor inhibited the “yes” response by 11 ms, the name of an inanimate object by 15 ms, and an animal name other than dog by 39 ms. The “no” responses were only inhibited 25 ms by the distractor word dog. It is noteworthy that these effects are very low compared with the Stroop inhibition as obtained in naming tasks that often exceed 100 ms. The authors conclude “that the automatically available semantic information from the word only causes problems in making decisions (a) when the word can supply information similar to but obviously not identical with that available from the picture and (b) when the two stimulus components are not compatible with the same decision” (Lupker & Katz, 1981, p. 277).

In a second experiment, the subjects had to name the superordinate category label of the target picture. Distracting words inhibited under several conditions by 22–33 ms without any significant difference. Only a category label that was an element of the set of the responses, but was different from the actual response, inhibited by 53 ms. Now, the authors conclude: “words do not cause problems for the response selection process by suggesting responses other than their names” (Lupker & Katz, 1981, p. 279). Later on, this conclusion was corroborated by the lack of any semantic component of the word–word interference in the reading task as given in the experiments by Glaser and Glaser (1989). Generally, these results by Lupker and Katz (1981) are to be interpreted in the context of the results by Smith and Magee (1980) and by Glaser and Düngelhoff (1984): the semantic processing of pictures is relatively immune against distracting words, except if the distracting words are possible responses as is the case in the usual picture-naming task.

Of course, the experiments of Lupker and Katz (1981) require complementary studies in which the targets are words and the distractors are pictures. Such a study was reported by Lupker and Katz (1982). Generally, greater inhibitions were now obtained. That means that in semantic decision tasks word processing is more inhibited by picture distractors than picture processing by word distractors. This is in accordance with the priming results discussed above. However, the experiments by Lupker and Katz of 1981 deviated in detail from those of 1982, so that subtle comparisons are not possible.

Smith and Magee (1980) presented a very important study with four experiments. The reading–naming and categorizing tasks were factorially crossed with word or picture as modality of the target. In three of the four experiments, the stimuli were presented as tables; in the categorizing task, the subject had to give a “yes–no” decision whether or not the target was an instance of a pre-specified superordinate category. The time to work through the table was divided by the
number of its elements. In Experiment 1, picture naming was inhibited 191 ms by an incongruent word, and word categorizing 194 ms by an incongruent picture. On the other hand, reading was prolonged 19 ms if a distracting picture was given, and categorizing a picture was prolonged 50 ms by an incongruent word.

The authors interpreted this result in terms of the relative-speed hypothesis of the Stroop inhibition: reading is faster than naming, and thus only the naming response should be disturbed by an irrelevant word. On the other hand, picture categorizing is faster than word categorizing, and thus now word categorizing should be disturbed by an irrelevant picture but not vice versa. These data agree with the relative-speed hypothesis of the Stroop interference as well as with the hypothesis of a faster access of the picture to the internal semantic code:

Experiment 1 examined the hypothesis that a reversal in the pattern of interference would occur when the task was changed from naming to categorization, a task that requires semantic analysis. In accordance with the hypothesis of more rapid semantic access by pictures, the pattern of interference did indeed reverse: Word categorization was delayed in the presence of distracting pictures, whereas picture categorization was left relatively immune to interference by the simultaneous presentation of incongruent words (Smith & Magee, 1980, p. 389).

It is noteworthy that the authors base this argument on the usual word–picture difference in categorizing, although they failed to replicate this difference in three experiments: words were categorized 16 ms or 11 ms faster than pictures (Experiments 1 and 3), and pictures were categorized only 2 ms faster than words in Experiment 2. Thus, the argument must be reversed: word categorizing is disturbed by an incongruent picture distractor even in experiments that do not show a faster picture than word categorizing. That is the same result for categorizing as was found for naming by Dunbar and MacLeod (1984): the word retains its power to inhibit the naming response even if word processing is experimentally delayed.

This line of argument was further pursued by Glaser and Düngelehoff (1984). They used a set of 36 basic objects that belonged to nine superordinate categories. All stimuli consisted of a word and a picture; four Stroop conditions were used. In the congruent condition, word and picture matched at basic level (e.g., chair–CHAIR). In the control condition, a non-picture or non-word provided neutral stimulation at the time and place of the distractor. In the category–congruent condition, words were different from pictures, but had their superordinate category in common (e.g., chair–TABLE). Finally, in the incongruent condition, words and pictures were different and were instances of different superordinate categories (e.g., chair–RABBIT). Beside that, SOA was the essential independent variable. Thus, the time courses of the Stroop-like effects were obtained.

Three essential results were found:

(1) Congruent and category-congruent stimuli showed the same amounts and time courses of facilitation and inhibition, respectively, as the conventional Stroop
stimuli that consist of colour words and colours. Thus, the identification of Stroop and picture-word interference with one another is further justified. Again, also a head start of the picture could not inhibit reading, whereas the word disturbed picture naming maximally within a SOA range from $-100 \text{ ms}$ to $+100 \text{ ms}$.

(2) The incongruent distractor words exhibited less interference than the category-congruent ones. This replicated the results of Guttentag and Haith (1978) and of Lupker (1979) that picture naming is maximally disturbed by distractor words of the same superordinate category. In particular, the fast, steep maximum of Stroop interference at $-100 \text{ ms} \leq \text{SOA} \leq +100 \text{ ms}$ was slurried with incongruent distractors.

(3) Word categorizing showed essentially the same time course of a strong inhibition as picture naming, whereas picture categorizing was only scarcely influenced by distracting words. Again, a head start of the word could not improve the word's impact on picture categorizing. That fits well with the data, but not with the theory of Smith and Magee (1980) insofar as the latter rests on the relative-speed hypothesis.

Glaser and Düngelhoff (1984) rejected the relative-speed hypothesis and concluded that the Stroop asymmetry, which is so dramatically reversed if the reading-naming task is changed to categorizing, indicates different priorities of internal pathways. Glaser and Glaser (1989) went a further step in this direction: if the Stroop asymmetry as found with modally mixed stimuli results from different internal priorities of the stimulus modalities in different tasks, then both components of modally pure stimuli should be processed along pathways with equal priorities. Thus, modally pure word-word or picture-picture stimuli should exhibit interferences that again should provide information about these pathways. This was corroborated empirically: the authors found strong Stroop-like inhibitions with the usual time courses for modally pure stimuli. Strikingly, the word-word interference lacked semantic components in the reading task, whereas the picture-picture interference showed them. Thus, new evidence was obtained that the pathway of reading aloud does not contain a mandatory stage for semantic evaluation. That is also in accordance with the results of some priming studies (e.g., Bajo, 1988; Irwin & Lupker, 1983).

Further research was devoted to the variables that determine the amount of Stroop inhibition. La Heij et al. (1985) accentuated the similarity between priming and Stroop-like experiments and asked why semantic similarity between prime or distractor, respectively, and target facilitates in the priming and inhibits in the interference experiments. They identified two important variables that contribute to these different results: (1) number of “semantic domains” and (2) the relation between the sets of distractors and targets.

(1) In the usual Stroop experiment, only one semantic domain, that of colour concepts, is used. Picture-word interference is usually investigated using a small
number of two to ten superordinate categories, whereas in priming studies often more categories are used (e.g., 64 by Bajo, 1988). Furthermore, in Stroop-like experiments, a small number of stimuli are repeated several times, whereas in most priming studies a large number of stimuli is presented only one or two times. In a word-word variant of the Stroop experiment, La Heij et al. (1985) could demonstrate a facilitation by associative relatedness between distractor and target that increased with increasing number of associated stimulus pairs that were active within a block of trials.

(2) In the usual Stroop experiments, the set of the distractor words is identical to the set of the spoken responses. However, these two sets can also be intersecting or disjoint. Generally, other conditions being constant, target processing is far more inhibited by distractors that are elements of the response set than by those that are not (Glaser & Glaser, 1989; Klein, 1964; Proctor, 1978). We called that variable set relation (Glaser & Glaser, 1989, p. 25). La Heij et al. (1985) found that the degree of inhibition due to set relation decreases with increasing number of categories.

La Heij and Vermeij (1987) varied the stimulus set size in the steps of 2, 4, or 8 stimulus alternatives in a picture-naming experiment and found that the interference due to incongruent words decreased and the facilitation due to congruent words increased with ascending number of target alternatives.

A very careful investigation of the variables that influence the Stroop-like interference of word distractors on picture naming was carried out by La Heij (1988). In Experiment 1, six pictures from the two categories musical instruments and tools were used as targets. The visual similarity of the pictures was kept constant within and between categories. Thus, the increased interference due to semantically related distractors compared to unrelated ones (Glaser & Düngelhoff, 1984; Guttentag & Haith, 1978; Lupker, 1979) should disappear if it resulted only from different visual similarity within and between categories as supposed by Neumann and Kautz (1982) and, for a face-name interference task, by Young, Ellis, Flude, McWeeny, and Hay (1986). It should also be remembered that Snodgrass and McCullough (1986) suggested this as the cause for the word-picture difference in categorizing.

Semantic relatedness was varied by combining a picture with a distracting word from the same (e.g., PIANO—guitar) or from a different (e.g., PIANO—chisel) category. Furthermore, one half of these distractors were labels of the pictures used in the experiment, the other half were not. Thus, set relation was combined factorially with relatedness. As control distractors, series of Xs were used, and names of objects that were neither musical instruments nor tools provided an additional semantically irrelevant condition. The results exhibited four independent effects: (1) the mere presence of a distracting word prolonged the latency by 45 ms compared with the non-word distractor; (2) the task relevance of the
distractor, that is, its usefulness as a potential response in the task to name musical instruments and tools, prolonged the latency by a further 23 ms; (3) semantic relatedness contributed an additional inhibition of 13 ms; and (4) set relation added 25 ms. There was no interaction between semantic relatedness and set relation so that the effects of these two factors are orthogonal to one another. In his Experiment 2, La Heij (1988) essentially replicated these results with other pictures that were controlled for visual similarity by an additional reaction-time experiment.

In line with Lupker and Katz (1981) and Huttenlocher and Kubicek (1983), La Heij (1988) supposes four stages in the picture-naming pathway: visual processing, semantic activation of the concept, retrieving of its name, and articulation. Because some studies have shown a very reduced Stroop interference if the response to the non-verbal stimulus component is given by pressing keys or with other non-naming responses like sorting cards or humming tones (McClain, 1983; Palef, 1978; Virzi & Egeth, 1985), La Heij (1988) concludes that the reading-naming interference does not arise whilst the picture is semantically identified, but during name retrieval. Glaser and Glaser (1989, p. 32) also tried to incorporate these empirical results in a model and arrived at the same theoretical conclusions. The Stroop interference is generated in the lexicon, because the code of the distracting word is activated from three sources: semantic spread of excitation from the target in the semantic system that is transferred to the lexicon, an increased basic activation as a potential response in the block of trials, and an activation through automatic reading. The Stroop inhibition is explained by the extra effort to select the target’s label against this strong evidence in favour of the distracting word.

An additional property of the picture-word interference was found by Lupker (1982) and by Rayner and Springer (1986). In Lupker’s (1982) Experiment 1, an unrelated word inhibited picture naming by 78 ms compared to a picture-alone condition. This interference was reduced by 72% to 22 ms if the word was orthographically similar to the picture’s name. In his Experiment 2, an inhibition of 69 ms was reduced by 23 ms (33%) if the distractor word rhymed with the response word, and by a further 32 ms (46%) if in addition it was orthographically similar.

In Rayner and Springer’s (1986) picture-naming task, there were congruent (e.g., BALL-ball, category TOY), category-congruent (e.g., BALL-drum), and incongruent (e.g., BALL-pear) distracting words as in the study by Glaser and Düngelhoff (1984). Additionally, there were distractors that shared initial letter and word contour with the label of the picture (e.g., category-congruent: BALL-bell; incongruent: BALL-bill). With the usual distractors, the usual results also were replicated: a congruent distractor facilitated the response, and a category-congruent distractor inhibited far more than an incongruent one. However, if the distracting word was not identical to the picture’s label, but had the initial letter and the shape in common with it, then the interference was reduced on average by
61% in the category-congruent and by 37% in the incongruent condition. This means that the distractor produces two involuntary effects: its property of being a readable word and its semantic relatedness to the target inhibit, whereas its graphemic similarity to the response word facilitates target processing.

At first glance, it does not seem to be impossible to reconcile these data with some kind of logogen model (Rayner & Springer, 1986). However, there remains a fundamental theoretical problem. The interference due to semantic relatedness is explained by assuming that the decision between target and distractor signal becomes more difficult with increasing similarity. This is a known fact about discrimination in psychology. On the other hand, increasing graphemic similarity of the distracting word facilitates target processing, although it should activate a different word that is also very similar to the correct response word and that should also compete with it for the control of the spoken response. In other words, the data of Rayner and Springer (1986) claim two different discrimination processes: one that becomes more difficult with increasing similarity among the alternatives and that operates on the semantic components of the naming task, and another that becomes easier with increasing similarity and works on the graphemic components.

However, perhaps a solution can be found within the framework of Levelt's (1989, p. 231) hypothesis that the word entries of the lexicon have two separate components: (1) the lemma which contains syntax and meaning (the latter refers to the semantic information that is necessary for the selection of the word by activated concepts in the semantic system); and (2) the morpho-phonological form which governs the articulatory programming. That means that lemma retrieval and articulatory programming are distinct processes which occur in strong temporal seriality. Thus, the semantic features of the distractor could become effective by activating their lemma that would compete with the correct response at the level of the lemmas, but the resulting inhibition terminates at this stage and does not reach articulatory programming (Levelt et al., 1991). On the other hand, linguistic features of the distractor work by activating phonemes on the basis of grapheme-phoneme correspondence at the morpho-phonological level of the lexicon. Thus, rhyme, orthography, and shape of the distractor could prime the particular phonemes that are used in an incremental articulatory programming of the correct response.

Levelt's hypothesis was essentially derived from analyses of speech errors (e.g., Dell & Reich, 1981; Fay & Cutler, 1977; Garrett, 1988). Later on, Levelt and his co-authors looked for additional evidence in Stroop-like experiments. Thus, Schriefers, Meyer, and Levelt (1990) gave their subjects a picture-naming task in which verbal distractors were presented acoustically. Beside neutral and unrelated words, semantically and phonetically related distractors were presented at the three SOA levels of −150 ms (distractor first), 0 ms, and +150 ms (distractor second). The main results were an increased inhibition due to semantically related
pre-exposed distractors and a facilitation due to phonologically related post-
exposed distractors.

These data clearly demonstrate two temporal windows for semantic inhibition
and phonological facilitation of the naming response. Semantic interference is
maximal, if the distractor is exposed early, whereas phonological facilitation
profits most from a post-exposed distractor. This is well in accordance with the
assumption of two separate steps of lexical access that follow one another, even if
it does not uniquely reject the alternative hypothesis of an overlapping or
continuously flowing transition between these two stages.

By the way, there is an undiscussed discrepancy between these authors' results
and those of the other Stroop-like experiments that varied SOA (Glaser &
Düngelhoff, 1984; Glaser & Glascr, 1982, 1989; La Heij et al., 1990). Whereas
the semantic interference usually reaches its maximum at SOAs from 0 ms
(synchrony) up to a post-exposure by +50 ms to +150 ms of the distractor,
Schriefers et al. (1990) found it only at a pre-exposure by −150 ms. Perhaps this
is caused by the auditory modality of the distractors. An experimental clarification
of this issue is still lacking.

Of course, there remains the problem that semantic similarity inhibits, whereas
phonological similarity of the distractor facilitates target processing. The authors
hypothesize that the articulation of the target's label does not use the distractor
word in its entirety - that should lead to the activation of a different, competing
response - but have accelerated access to the single phonemes that are primed by
the distractor. As discussed above, this hypothesis can also explain the data from
other investigations (e.g., Lupker, 1982; Rayner & Springer, 1986). Nevertheless,
进一步 converging evidence would be useful.

In a series of six experiments, Levelt et al. (1991) looked for further evidence for
the two-step hypothesis of lexical access. Their subjects received a speeded picture-
naming task. In the relevant trials, an additional word or non-word stimulus,
respectively, was presented acoustically. Contrary to the usual picture–word inter-
ference task, now the subjects had to make a lexical decision concerning the acous-
tical stimulus as a secondary task that was also timed. The main dependent variable
was now the latency of the lexical decision. Besides SOA, the relatedness of picture
and word was the essential independent variable. It was varied in seven steps:

1. In the identical condition, the name of the picture matched the word (e.g.,
   SHEEP–sheep).
2. In the semantically related condition, the word was a close associate of the
   picture (e.g., SHEEP–wool).
3. In the condition of semantic alternatives, the word denoted a concept of the
   same superordinate category as the picture (e.g., SHEEP–goat).
4. The phonologically related condition was made up by words that shared
   the initial phonemes of the target's name (e.g., RADIO–radar).
(5) There were also unrelated words that were associatively as well as semantically and phonemically distant to the picture's label.

(6) A further condition contained words that were phonemically similar to a close associate of the target (e.g., SHEEP—wood because wool is an associate of sheep).

(7) A final condition contained words that were phonemically similar to a semantic alternative of the target (e.g., SHEEP—goal because goat belongs to the same superordinate category as sheep).

SOA was varied in the three steps short (word on average 73 ms after picture), medium (word on average 373 ms after picture, and long (word on average 673 ms after picture).

The main results were as follows. The unrelated condition was used as a baseline with respect to which inhibitions and facilitations were evaluated.

(1) The lexical decision of words that were identical to the response was inhibited at short and facilitated at long SOAs. This is reasonable if one assumes that at short SOAs phonological access of picture and word overlap and hamper one another. At long SOAs, on the contrary, word processing is facilitated because the identical word is already phonologically activated as a response to the picture.

(2) Semantically related words were inhibited at short SOAs but showed no effects at medium and long SOAs. Apparently, this semantic interference occurs only if the word stimulus taps into the semantic processing of the picture. If this stage of picture processing is terminated, there are no more semantic effects on word processing.

(3) At short SOAs, semantic alternatives were inhibited far more than semantically related words (Experiment 6). That is in accordance with the data from the usual picture-word experiments (cf. Lupker, 1979).

(4) The phonologically related words were strongly inhibited at all three SOA levels. This result contrasts to the data from picture-word interference studies in which phonological similarity between the distracting word and the target's label always facilitated the response (Lupker, 1982; Rayner & Springer, 1986; Schriefers et al., 1990). This discrepancy, of course, needs further discussion which, however, is beyond the scope of the present article.

(5) Finally, words that were phonologically related to an associate or a semantic alternative of the target's label did not show any effect different from unrelated words. This is strong evidence for the two-step hypothesis of lexical access: if a spoken word is required as a response in picture naming, then other semantically related concept codes are coactivated through a spreading semantic excitation. This activation terminates at their lemmas. That means that only the one word that has to control the overt response is also programmed phonologically, whereas the coactivated codes do not reach the phonological and articulatory stages.
7. Conclusions

The three domains of experimental investigations on picture naming arrived at an impressive corpus of data:

(1) The comparison of naming with reading as well as with picture and word categorizing showed three reliable effects: reading is markedly faster than picture naming, naming pictures is faster than categorizing pictures, and picture categorizing is faster than word categorizing. Furthermore, picture categorizing is accelerated for superordinate categories that share characteristic visual features with objects at basic level (Hoffmann & Klimesch, 1984), and it is inhibited if the pictures are especially similar to one another between the categories (Snodgrass & McCullough, 1986). The word-frequency effect is far greater in naming than in reading responses.

(2) The priming results are partially different from one another, so that it is difficult at present to integrate them all into a completely consistent theoretical framework (cf. Snodgrass, 1984, p. 14). Nevertheless, one characteristic result is that in the reading and naming tasks pictures are more effective as primes and are more primeable as targets. Within-modality priming is more effective than between-modality priming. Often, the reading response is only marginally or not at all primed. In some cases, even picture naming was not primed (e.g., Irwin & Lupker, 1983), whereas in other cases the reading response was markedly primed depending on voluntary strategies of the subjects (cf. Bajo, 1988). Thus, the contribution of the priming studies to the evaluation of the hypotheses given in the second section seems so far to be limited. Nevertheless, further productive contributions of priming studies could be possible (e.g., De Groot, 1990).

(3) Although the Stroop-like interferences have been used only recently as a tool to investigate picture processing, they have provided some strong, reliable, and consistent effects. Thus, the Stroop asymmetry is reversed if the reading and naming task is replaced by the categorizing task: word processing is inhibited by an incongruent picture distractor. Modally pure stimuli, that is, word-word and picture-picture pairs, always exhibit interference, but lack semantic components in the reading task. In a way that is not yet fully understood, the Stroop-like interference is very sensitive to a common category membership of target and distractor, but unaffected by typicality (e.g., Lupker, 1979). Associative relatedness among distractor and target label facilitates to a small degree at synchronous exposure and to a marked degree if the distractor precedes the target by about -400 ms and more (La Heij et al., 1990). Task relevance of the distractor, set relation between target and distractor and number of semantic domains are also theoretically informative independent variables (cf. La Heij, 1988). Finally, there is a nearly complete paradox insofar as a distractor that is semantically similar to the target inhibits its processing, whereas a phonemically similar distractor facilitates. Perhaps this effect is related to the facilitation by associative strength.
As was demonstrated in detail throughout this article, these data can be integrated within a theoretical framework that contains two large components as shown in Figure 1(c): (1) an abstract semantic memory that nevertheless is functionally connected with picture processing and, in a broader sense, with perception and action; and (2) a lexicon that provides storage and processing facilities for all linguistic knowledge and abilities beyond semantics.

References


Neumann, O., & Kautz, L. (1982). *Semantische Förderung und semantische Interferenz im Be*
nennungsexperiment [Semantic facilitation and semantic interference in a naming experiment] (Bericht Nr. 23/1982). Bochum, Germany: University of Bochum.


