Why do we SLIP to the basic level? A formal model

Frédéric Gosselin

Department of Psychology

University of Glasgow

Submitted for the Degree of Ph.D. to the Higher Degrees Committee of the Faculty of Social Sciences, University of Glasgow.

March, 2000

So out of the ground the LORD God formed […] every bird of the air, and brought them to the man to see what he would call them; and whatever the man called every living creature, that was its name.

(Genesis, 2:19)

Abstract

This dissertation introduces a new measure of basic-level performance (*Strategy Length & Internal Practicability, SLIP*). SLIP implements two computational constraints on the organisation of categories in a taxonomy: the minimum number of feature tests required to place the input in a category (*strategy length*) and the ease with which these tests are performed (*internal practicability*). The predictive power of SLIP is compared with that of four other basic-level measures: context model (Medin & Schaffer, 1978; modified by Estes, 1994), category feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), and compression measure (Pothos & Chater, 1998a), drawing data from the empirical work of Rosch et al. (1976), Murphy and Smith (1982), Mervis and Crisafi (1982), Hoffmann and Ziessler (1983), Corter, Gluck and Bower (1988), Murphy (1991), Lassaline (1990), Tanaka and Taylor (1991), and Johnson and Mervis (1997). Nine experiments further test the validity of the computational constraints of SLIP using computer-synthesised 3-D artificial objects, artificial scenes, and letter strings. The first five experiments examine the two constraints of SLIP in verification. Experiment 1 isolates the effect of strategy length on basic-levelness, and Experiments 2a and 2b that of internal practicability. Experiment 3 examines the interactions between the two factors. Experiment 4 tests, whether, as predicted by SLIP, there is a linear relationship between strategy length and response times. The last four experiments study the two computational constraints in naming. Experiment 5a isolates the effect of strategy length, and Experiment 5b that of internal practicability. Experiment 6 examines the time-course of the effect of strategy length. Finally, Experiment 7 looks at the effect of *robustness* (i.e., the idea an

approximate categorisation is better than none) on the order of feature tests in length 2 strategies.

Table of content

Preamble

The 20-question game

Most of you have played the 20-question game. A minimum of two players is required. A category name which applies to the real world is chosen by one of the players (we will call him the *answerer*). The other players (we will designate them as the *guessers* from now on) try to discover the selected name after having asked as few yes/no questions as possible. The answerer must answer truthfully to these questions. Any type of binary query is allowed in 20-questions. Thus, guessers could inquire whether the chosen word begins with a letter prior to "n" in the Latin alphabet, whether it has more than two vowels, whether it sounds like the mating call of a moose, and so on. The usual tactic, however, consists in asking relational questions, i.e. questions that reveal certain relations between the target category and another category, rather than orthographic, phonetic, or phonologic questions. There are four possible relations between two categories: independence (e.g., natural is independent of man-made), synonymy (e.g., physician is a synonym of medical doctor), partial overlap (e.g., blond partially overlaps woman), and inclusion (e.g., mammal is included in animal). By extension, there are four broad kinds of binary-relational questions that can be asked.

In the 20-question game, the preferred type of relational question is inclusion, starting with highly general categories and going on to more particular ones (Bendig, 1953). This type of relation between categories is also called an IS-A relation (Collins & Quillian, 1969). For example, the answerer could pick the name "Moby Dick", and the guessers could find it after the following dialogue: "Is it an animal?", "Yes."; "Is it a mammal?", "Yes."; "Is it a feline?", "No."; [...]; "Is it a cetacean?", "Yes.";

"Is it a whale?", "Yes."; [...]; "Is it a white whale?", "Yes."; "Is it Moby Dick?", "Yes. It is Moby Dick.". The fact that it is possible at all to use inclusions implies that at some cognitive level categories possess a tree organisation. Given the adaptability of humans, it is not too surprising that categories are organised this way, and that people use this property when playing the 20-question game: it is the fastest known search algorithm (e.g., Dewdney, 1989). In the best of worlds, it would enable someone to complete the game in $\log_2$(NUMBER_OF_CATEGORIES) questions.

Category feature-structure vs. category tree

Most category trees implemented by computer scientists are arbitrary insofar as they tell us nothing about the feature structures of their categories. From an optimal-search-algorithm standpoint, for example, it does not matter in which one of the three possible complete trees the four categories A, B, C, and D are inserted (i.e., {[(A) (B)] [(C) (D)]], {[(A) (C)] [(B) (D)]}, or {[(A) (D)] [(B) (C)]}). In humans, however, category trees are not arbitrary; they are also powerful feature inference machines: every category inherits the properties or attributes of related more general categories. For example, Moby Dick lives in the water like all other whales and breathes air just like any other mammal. We will say that humans organise their categories in feature-structures (category structures, hierarchies, and taxonomies). Category feature-structures are special category trees.

Are all the categories or nodes in humans' category feature-structures equal? More specifically, is there a level of organisation in this hierarchy which is psychologically superior? We will now briefly review the evidence that one level has a special psychological status.

Basic-level phenomenology, short version

In Rosch, Mervis, Gray, Johnson and Boyes-Braem's (1976, Experiment 7), participants were taught the name of 18 objects at three levels of categorisation–the *subordinate* (e.g., *Levis, Macintosh*), *basic* (e.g., *pants, apple*) and *superordinate* (e.g., *clothes, fruit*). These objects belonged to one of six possible non-biological taxonomies: *musical instruments*, *fruit*, *tools*, *clothing*, *vehicles*, and *furniture*. In a verification task, subjects were shown a category name followed by a stimulus picture, and had to determine whether they matched. Categories at the basic-level were fastest to verify, and categories at the subordinate level slowest (see also Hoffmann & Ziessler, 1983; Jolicoeur, Gluck & Kosslyn, 1984; Murphy, 1991; Murphy & Smith, 1982; Murphy & Brownell, 1985; Tanaka & Taylor, 1991).

The basic level is superior in many other respects: (1) objects are named quicker at this level than at any other level of abstraction (Hoffmann & Ziessler, 1983; Jolicoeur, Gluck & Kosslyn, 1984; Murphy, 1991; Murphy & Smith, 1982; Murphy & Brownell, 1985; Rosch et al., 1976; Tanaka & Taylor, 1991); (2) objects are designated preferentially with their basic-level names (Berlin, 1992; Brown, 1958; Rosch et al., 1976; Tanaka & Taylor, 1991; Wisniewski & Murphy, 1989); (3) many more features–especially shapes–are listed at the basic level rather than the superordinate level, with only a slight increase at the subordinate level (Rosch et al., 1976; Tversky and Hemenway, 1984); (4) throughout development, basic level names are learned before those of other categorisation levels (Anglin, 1977; Brown, 1958; Rosch et al., 1976; Horton & Markman, 1980; Markman, 1989; Markman and Hutchinson, 1984; Mervis and Crisafi, 1982); and (5) basic names tend to be shorter (Brown, 1956; Rosch et al., 1976). Convergence of these performance measures is crucial to establish a preferred categorisation level, even though verification speed is the most commonly used.

There is considerable evidence that a basic-level superiority holds across cultures for living kinds (for extensive reviews see Berlin, 1992, and Malt, 1995). Basic-level phenomenology also seems to hold across domains (for a review see Murphy & Lassaline, 1997), such as computer programs (Adelson, 1983), events (Morris & Murphy, 1990; Rifkin, 1985; Rosch, 1978), personality types (Cantor & Mischel, 1979), sign language (Newport and Bellugi, 1978), environmental scenes (Tversky & Hemenway, 1983), clinical diagnosis (Cantor, Smith, French, and Mezzich, 1980), and emotions (Shaver, Schwarz, Kirson, and O'Connor, 1987).

The aim of this dissertation

To summarise, people organise their categories in special category trees called category feature-structures. The different levels of these are not created equal. Many indexes of performance are maximised at the so-called basic-level. The main goal of this dissertation is to *explain* why this is so.

Basic terminology

It is worth pointing out at this point that the usage of "basic level" is ambiguous in the literature. It can refer to the middle-level of a three-level hierarchy (with the level above called "superordinate" and the one bellow "subordinate"–e.g., Markman, 1989), to an index of performance (the fastest level, or the one most often used to name things, and so forth–e.g., Corter and Gluck, 1992; Anderson, 1990, 1991), or to both the level of categorisation and the index of performance (e.g., Rosch et al., 1976; Mervis & Crisafi, 1982). This ambiguity is beautifully illustrated in Tanaka and Taylor's (1991) "basic to subordinate shift" which is nonsense unless one changes the meaning of "basic" from "index of performance" to "middle level of categorisation" in mid-air. Henceforth, the *basic-levelness* of a category will denote a measure of performance. Whenever possible, we will refer to the levels of abstraction as the subordinate, basic, and superordinate. Otherwise, we will use a set of unambiguous level descriptors–e.g., low, middle, and high. The subordinate-basic-superordinate trio has the advantage of having a phase known to most psychologists. Although most experiments have probed these three embedded categorisation levels, people can use many more levels in their interactions with objects. Berlioz, for instance, was a famous French composer, an artist, a human, a mammal, a living organism, a bunch of atoms, and so forth. Sometimes we will use the more general level descriptors: H (highest level), H - 1 (second highest), H - 2 (third highest), etc.

Chapter 1. General introduction: basic-level phenomenology, long version

In this chapter we shall discuss the basic-level phenomenology in more detail. The work of Eleonor Rosch and colleagues has been so influential that it provides us with a "natural breaking point", a *basic* event in the basic-level literature: we will first discuss the research conducted by Rosch et al.'s predecessors, then their own contribution, and finally the work of their successors.

We must point out that we understand the expression "basic-level literature" in a most inclusive way that is, all articles that examined levels of generality and have suggested that one of these is special in some respect. The name of this special level of categorisation varies throughout time and area of study; it is sometimes called *level of usual utility* (Brown, 1958), *entry level* (Biederman, 1987), *entry point* (Jolicoeur, Gluck and Kosslyn, 1984), *basic level* (Rosch et al., 1976; and most of the psychological literature), *folk generic level* (Berlin, 1972, 1992; as well as most of the anthropological literature), *genus level* (Anderson, 1989, 1990), *BOL* (from Rosch's "Basic Object Level"; Posey, 1979), B^0 (Taylor, 1990), and so on.

1.1 Before Rosch and colleagues

1.1.1 Brown's level of usual utility

In his article *How shall a thing be called?*, Brown (1958) asked us to consider a parent teaching a child the names of things in the world in his native language. The usual strategy consists of pointing at something and saying "this is an X". This is an *ostensive definition* (see?). Ostensive definitions are inherently ambiguous. Consider Brown's dime example. A dime could be named a "dime", but also "money", a "metal object", a “thing”, and moving to subordinates, it could be designated as a “1952 dime”, or as a “particular 1952 dime”. (Note that Brown is concerned only with the vertical ambiguity–an ambiguity associated with level of generality–in naming things, but a horizontal ambiguity–an ambiguity associated with synonyms–also exists. For example, “dime” is synonymous with “10 cents”.) How is a parent to select a name among all these possibilities? They could just select one randomly, and this would be the end of the story. This is not however what they do. Brown appeals to our intuition (like linguists often do) to convince us that a dime may be named “money” or “dime”, but probably not “metal object”, “thing”, “1952 dime”, and so on. He then goes on to make a more general claim: “Listening to many adults name things for many children, I find that their choices are quite uniform and that I can anticipate them from my own inclinations.” (Brown, 1958, p. 14). Why is it that adults consistently choose names at a certain level of categorisation? His answer: adults have a notion of the language appropriate for use with children. “It seems likely that things are first named so as to categorise them in a maximally useful way.” (Brown, 1958, p. 20). Brown calls the level of categorisation at which these uniform choices occur the *level of usual utility*.

Brown’s (1958) contribution is more theoretical than empirical. We will come back to this in Chapter 3. However, Brown is the first to have emphasised this empirical reality that not all levels of generality are equal. He also anticipated important research trends in the basic-level literature: He proposed that the level of usual utility could be revealed by several indexes of performance. Brown mentions Zipf’s (1935) seminal work which showed that word length is negatively correlated with word frequency which is correlated with usual utility. For example, the monosyllable “dog” has much higher frequency according to the Thorndike-Lorge list than do the polysyllables “boxer”, “quadruped”, and "animate being". Brown also suggested that children first learn names at the level of usual utility. This is founded on the fact that the *child's concrete vocabulary* comprises more words at the level of usual utility than at any other level of generality. For Brown, this follows from the parents' tendency to use usual utility names in designating things to their children. This was studied by developmental psychologists much later (e.g., Markman & Horton, 1980). Brown's utility principle was formulated with great care to include possible individual differences. The names used by parents to designate things to their children are the most useful to categorise them for non-linguistic purposes *in their experience*. For a numismatist (a coin collector), for example, a priceless 1910 dime is more a "priceless 1910 dime" than simply "money" or a "dime". For the kids from a particular neighbourhood a dog might be "Prince", but it is a "dog" for the rest of the world. Brown also ventured that maybe the level of usual utility of the parents differs from that of the children (c.f., the Mervis child's basic level). This led to some research in anthropology (e.g., Berlin, 1992; Berlin, Breedlove & Raven, 1973; Boster, 1980; Coley, Medin & Atran, 1997; Dougherty, 1978) as well as psychology (e.g., Rosch et al., 1976; Tanaka and Taylor, 1992; Johnson and Mervis, 1997).

1.1.2 Cognitive anthropology and the folk-generic level

Cognitive anthropologists are interested in how we segment the world into categories. They study the extent to which these categories are given by the input, that is, by the environment, and the extent to which they are created through constructive cognitive processes. They look at classification across cultures in a range of domains such as plants and animals (ethnobiology), colour, kinship, textiles, and household objects such as pots and bowls. Of these, the ethnobiological studies have the most commonalties with psychology; furthermore, they have been the most rigourous in their approach (Malt, 1995). We will thus focus on this ethnobiological literature here.

We will examine three intertwined research themes: (1) folk and scientific taxonomies comparisons, (2) evidence for a preferred level of categorisation, and (3) evidence for a universal preferred level of categorisation (i.e., which holds across all cultures). The third theme is especially interesting because it goes beyond what standard cognitive psychology has to offer. Themes (1) and (2) are prerequisites for (3). If there is a universal preferred level of generality, then taxonomies must be commensurable to a certain extent. At the very least, they must share this special level. This is what (1) establishes. We will spend some time on this because it fleshes out the methods used by cognitive anthropologists to address questions more relevant to us. And if there is a universally preferred level, then there must be a preferred level within each folk taxonomy. This is the question tackled in (2).

Many researchers have shown that folk and scientific ethnobiological classifications overlap greatly. (They have not compared two folk taxonomies–an endeavour which would tap into their goal more directly–because of the quantity of work this would involve; the scientific ethnobiological taxonomies are readily available (Malt, 1995).)

The scientific classification break-down for an African elephant, for example, is *Animalia* (animal) at the *kingdom* level, *Chordata* at the *phylum* level, *Vertebrata* (vertebrate) at the *sub-phylum* level, *Mammalia* (mammal) at the *class* level, *Proboscidea* at the *order* level, *Elephantidae* (elephant) at the *family* level, *Laxodonta Africana* (African elephant) at the *genus* level, and *Laxodonta Africana Vulgaris* at the *species* level (e.g., Perrott, 1971).

Berlin (1972, 1992; see also Berlin, Breedlove & Raven, 1973) has examined folk classification systems in the most thorough fashion: He studied plant classification in the Tzeltal Maya of southern Mexico and the

Aguaruna Jívaro of north central Peru, two traditional cultures. Psychologists know of this work especially for its description of folk classifications as taxonomies with a preferred level of generality–the *folk generic* level[1]–(see Rosch et al., 1976). However, another important contribution of Berlin's research is the comparison of folk and scientific taxonomies. For his analysis of the correspondence of the Tzeltal folk categories to the scientific categories, Berlin (1972; Berlin et al., 1973) concentrated on the so-called *folk-generic* categories. These correspond to the most specific categories labelled by single words and to the most common and salient categories (see Berlin, 1992). Latter Rosch et al. (1976) will argue that folk-generic and *basic* categories are roughly the same. Berlin discovered 61% one-to-one correspondence between the Tzeltal folk generic botanical categories and scientific botanical species. This is a rather large percentage of overlap.

Bulmer (1970) found a comparable level of overlap in the Kalam's of New Guinea taxonomy of vertebrate animals. Instead of using folk-generic categories like Berlin, he used *terminal* categories that is, the most specific categories known to a culture. Note that sometimes these are at the folk-generic level, and sometimes they are more specific. The Kalam terminal categories overlap with 60% of the scientific zoological species.

Hunn (1977) studied the agreement between Tzeltal animal folk categories and scientific animal categories at all levels of categorisation (i.e., *species*, *genus*, *family*, etc.). His measure of the degree of dissimilarity between a folk and a scientific category is perhaps best explained by an

---

[1] Although Berlin's (1972) "folk generic level" became the standard expression to designate this special level of categorisation in cognitive anthropology, it is worth mentioning that Conklin (1954) had already used the expression "basic name" relative to plant taxonomies to express the same idea.

example. Suppose that a folk category includes all the specimens of the "Rodentia" scientific category and only those. The dissimilarity measure between the folk and the scientific category would be zero. But if this folk category also included *moles* and *shrews* that is, members of the scientific order of "Insectivora", then one level would have to be climbed in the scientific taxonomy before all the folk category would be comprised in a scientific one; here the dissimilarity between the folk and the scientific category would be one (Malt, 1995). Hunn found 79% of zero dissimilarity for birds categories, and 78% for mammals. Most of the non-zero dissimilarities were low. Hunn concluded that the overlap between folk and scientific categories is important.

A few studies suggest that utility plays a role in the way cultures classify their environment. Diamond (1966) remarked that even though Fore had very specific categories for birds, they classified all butterflies in one category. Bird categories have a utility for the Fore because these categories help the Fore hunt the birds that they like; butterfly categories have little interest. Bulmer (1970) observed that the Kalam of New Guinea tend to group biological species together when they are of no use to them. A recent study conducted by Medin, Lynch, Coley and Atran (1997) examined the effects of goals and interests on classification and reasoning processes. They showed that classifications made by landscape workers were largely influenced by utilitarian factors. For example, landscapers' groupings of trees were frequently based on properties such as landscape utility, aesthetic value, size, and weediness.

To summarise, people from different cultures organise their environment in taxonomies, and these folk taxonomies overlap with scientific taxonomies, although incompletely. There is some evidence that utility might play a role in how complete or incomplete these folk taxonomies are. The next question that cognitive anthropologists addressed will ring a bell: are all these levels equivalent, or, alternatively, is one special in some respect?

The evidence for a special level of categorisation in ethnobiological studies comes from a number of independent studies of different cultures. The structure of the names of categories from different levels of categorisation is one type of evidence: Both folk generic categories, such as “oak” and “redbud”, and the next higher level categories, such as “tree” and “herb”, are reliably named with *primary lexemes*. Folk-generic categories are the most specific categories designated by primary lexemes. Usually primary lexemes are single words (e.g., “maple”, “bass”), but they can also be compound nouns (e.g., “poison oak” or “baby breath”), if they do not contain the name of an immediate superordinate category (e.g., *poison oak* is not included in neither the *poison*, nor the *oak* category) and contrast with categories that are primary lexemes (e.g., *tulip tree* contrasts with primary lexeme categories such as *oak* and *maple*, and therefore is a primary lexeme). Lower-level category names are almost always composed of two words: the name of a superordinate category, preceded by a modifier (e.g., “pine warbler”). Moreover, the categories they contrast with comprise the same superordinate category name in their compound name (e.g., “palm <u>warbler</u>” contrasts with “Canada <u>warbler</u>”) (see Malt, 1995).

Hunn’s analysis of the degree of correspondence between folk and scientific taxonomies gives additional support to the claim that the folk-generic level is superior to the others (see our earlier discussion of Hunn’s research). Ninety-one percent of the Tzeltal animal folk generic categories have levels of dissimilarity of zero or one, whereas 85% of their subordinate categories and 51% of their superordinate categories have the same levels of dissimilarity.

Another evidence is that folk-generic categories outnumber all the other kinds of categories. This happens because only 20% of the folk generic categories are subdivided (Berlin, 1992). This is in itself evidence for the salience of those categories. Investigators have also reported that generic names are those most easily and commonly elicited from informants (Berlin et al., 1973; Taylor, 1990) although these findings are not as robust. In other words, the most typical answer to the question "what is this?" is a folk generic name. Finally, there is some data that suggest that children learn folk-generic names first. Stross (1973) asked 25 Tzeltal children (4 to 13 years old) to name 209 different plants. Children most often produced the folk-generic names.

To summarise, there is ample evidence from ethnobiology that the folk-generic level of abstraction has a special psychological status. Is this preference universal? In other words, does it hold across all cultures?

The folk-generic level found in most anthropological studies corresponds roughly to the *scientific* generic level (which, in most cases, is coextensive with single species since frequently only one species of a genus is present in a given local environment; Berlin, 1992).

However, there is also evidence that the folk-generic level may vary to a certain extent as a function of the individual's expertise about the domain. Boster (1980) discovered that the members of the Aguaruna community responsible for cultivating manioc refered to manioc plants by sub-folk-generic terms, whereas other members of the community used the folk-generic label. Berlin (1992) further noted that Aguarana women fail to differentiate among members of some scientific genera of forest birds which the men, who spend more time in the forest, do name distinctively. Similarly, Dougherty (1978) remarked that urban American children, in contrast to the Tzeltal children studied by Stross, appear to learn supra-generic distinctions among plants first and may never learn more than about a dozen folk generic distinctions. This is consistent with Rosch et al.'s (1976) finding that for biological categories such as trees, their college student subjects seemed to have a basic-level above the folk generic level (the basic-level usually corresponds to the family level in scientific taxonomies) reported by Berlin et al. (1973). Remember that these individual differences were already suggested by Brown (1958).

More recently, Coley, Medin and Atran (1997) examined the relationship between privileged levels in folk biological taxonomies and inductive inference. They predicted that the principles that lead to basic-level phenomenon (e.g., high within-category similarity relative to between-category similarity–see *1.3.1 Tests of the differentiation model*) would lead to inductive privilege. Differences in the location of the folk-generic level across cultures should thus be reflected in differences in which level appeared privileged for induction. For Itzaj Maya adults, results were as predicted by anthropological accounts of folkbiological taxonomy: Inferences to scientific-generic (i.e., the folk-generic level) categories were consistently stronger than ones to more general categories. However, for Americans college students, results showed that the the preferred level for naming, etc. was the family level (i.e., the Rosch's basic level), whereas the privileged level for induction was the scientific generic level.

1.2 Rosch and colleagues' basic-level

In a series of papers, Rosch and her colleagues introduced the basic-level problema to cognitive psychology: Rosch, Mervis, Gray, Johnson, and Boyes-Braem's (1976) paper is the single most influential article published on the basic-level and its ramificatons[2]. Although the Rosch and

[2] Just to give you an idea: Rosch et al. (1976) was cited 965 times since its publication (BIDS Citation Index). Compare this citation count with those of a few classics:

Mervis (1975) paper is better known for the family resemblance idea it is the first published article to use the expression "basic-level". Rosch (1977 and 1978) as well as Mervis and Rosch (1981) are review articles, and do not add much to Rosch et al.'s (1976) contribution.

Theoretically, Rosch et al. (1976) proposed a utility model of basic-levelness (*cue validity*) and a similarity-based one (*differentiation model*), and by the same token created the two attractors around which most other basic-level models would be organised (see Chapter 3). Empirically, they showed that of three levels (*subordinate*, *basic*, and *superordinate*), the basic was psychologically superior in many respects. We will describe Rosch et al.'s experiments in detail because they set the agenda for all future research.

Rosch et al. were very much influenced by Berlin's (1972; Berlin et al., 1973) early work. In fact, Experiments 1 to 4 are aimed at correcting three shortcomings of the ethnobiological studies: (1) they refers only to biological classes, (2) the claims for natural groupings are generally supported by few correlated attributes (this has been corrected in Berlin's latter work), and (3) the location of natural groupings at a particular level of abstraction is defined by linguistic-taxonomic, rather than psychological criteria.

Miller's (1956) magical number paper is cited to date on 1489 occasions, Shepard and Metzler's (1971) mental rotation paper, on 713 occasions, Medin and Schaffer's (1978) context model article, on 579 occasions, and Tversky's (1978) contrast model paper, on 981 occasions (BIDS Citation Index). And with those of early influential basic-level articles: Brown (1958) is cited on 99 occasions, Murphy and Smith (1982), on 85 occasions, Horton and Markman (1980), on 57 occasions, Jolicoeur, Gluck and Kosslyn (1984), on 118 occasions, Mervis and Crisafi (1982), on 101 occasions, and Tversky and Hemenway (1984), on 192 occasions (BIDS Citation Index). This is evidence that Rosch et al. (1976) constitutes a basic event in the short basic-level literature history.

In Experiment 1, Rosch et al. systematically studied the co-occurrence of attributes in the most common taxonomies of man-made and biological objects in our culture. They used 90 object names belonging to three levels of categorisation in nine taxonomies. The names were chosen so that they would be representative of categories in Western culture. These nine taxonomies were used in all of Rosch et al. experiments. They can be seen in detail in Table 1.

Table 1: Taxonomies used by Rosch et al.. Adapted from Rosch et al. (1976, Table 1).

Nonbiological taxonomies			
Superordinate	Basic	Subordinate 1	Subordinate 2
Musical instrument	Guitar	Folk guitar	Classical guitar
	Piano	Grand piano	Upright piano
	Drum	Kettle drum	Bass drum
Fruit	Apple	Delicious apple	Mackintosh apple
	Peach	Freestone peach	Cling peach
	Grapes	Concord grapes	Green seedless grapes
Tool	Hammer	Ball-peen hammer	Claw hammer
	Saw	Back hand saw	Cross-cutting hand saw
	Screwdriver	Phillips	Regular screwdriver
Clothing	Pants	Levis	Double knit pants
	Socks	Knee socks	Ankle socks
	Shirt	Dress shirt	Knit shirt
Furniture	Table	Kitchen table	Dining room table
	Lamp	Floor lamp	Desk lamp
	Chair	Kitchen chair	Living room chair
Vehicule	Car	Sports car	Four door sedan car
	Bus	City bus	Cross country bus
	Truck	Pick up truck	Tractor-trailer truck
Biological taxonomies			
Superordinate	Basic	Subordinate 1	Subordinate 2
Tree	Maple	Silver maple	Sugar maple
	Birch	River birch	White birch
	Oak	White oak	Red oak
Fish	Bass	Sea bass	Striped bass
	Trout	Rainbow trout	Steelhead trout
	Salmon	Blue back salmon	Chinook salmon
Bird	Cardinal	Easter cardinal	Grey tailed cardinal
	Eagle	Bald eagle	Golden eagle
	Sparrow	Song sparrow	Field sparrow

Two hundred students listed attributes for each of the categories. For example, a participant could have listed the three features *roots*, *Canadian emblem,* and *leaves* for the category of "sugar maple". Only the attributes that were listed at least six times were included in the analysis. For nonbiological categories, they found that people tended to list many more features at the basic-level than at the superordinate-level, with only a slight increase at the subordinate-level. This pattern of results was shifted up one level for the nonbiological taxonomies. In other words, the biological superordinates and basics showed the same proprieties as the nonbiological basics and subordinates, respectively[3]. (This experiment is further discussed in Chapter 4.)

In Experiment 2, Rosch et al. studied the similarities between the motor programs associated with the use of these objects for the three levels of abstraction. They asked participants to describe their interactions with the category objects at different levels of categorisation. For instance, a participant could *play music with musical instruments*, but *hold a guitar on her lap, pinch its strings with one hand, modulate the strings length with the other, and so on*. Nonbiological superordinate categories have few, if any, motor movements that can be made to the category as a whole and few movements in common. Nonbiological basic-level categories receive descriptions of many specific movements made to all members of the

---

[3] Rosch et al. (1976) chose their categories armed with the early lessons of cognitive anthropology. Sparrow, trout, and oak are folk-generic categories (or, in Table 1, basic categories) and should therefore be preferred to their respective superordinates bird, fish, and tree. Rosch et al. found the opposite pattern of preferences. Dougherty (1978) observed the same phenomenon in urban American children. This suggests that the folk generic level is not universal. In any case, cognitive psychology has forgotten the biological nomenclature illustrated in Table 1. Now bird, fish, and tree are basic categories, and sparrow, trout, and oak some of their respective subordinates.

category and many of these movements are described by a sufficient number of different participants to form a picture of movement sequences made in common to all members of the basic class of objects. Subordinate nonbiological categories did not differ significantly from the basic ones either in the specificity of the descriptions or in the number of common movements made in interacting with the object.

Again, the characteristics of the nonbiological taxonomies in Table 1 were shifted down one level relative to the biological ones. From Experiment 3 on, Rosch et al. shifted their original biological taxonomy (see Table 1) so that the categorisation level exhibiting the best performance coincided with the level called "basic". Bird, fish, and tree thus became basic categories; and sparrow, trout, and oak some of their respective subordinates. This is the biological basic-level we know and love.

It seems plausible prima facie that a high degree of similarity for motor programs in Experiment 2 results from things sharing many parts and these parts being organised in similar ways. In other words, things might be handled the same way because they "look" the same. This is also suggested by the shared features of Experiment 1: an important proportion of these shared features are shapes. In Experiments 3 and 4, Rosch e al. examined this hypothesis. In Experiment 3, the ratio of overlap to nonoverlap between the outline of sets of objects (normalised for size and orientation) was computed. Particular care was taken to select objects without bias; they were randomly selected from a large database. The superordinate categories were: *clothing*, *vehicle*, *animals*, and *furniture*; there were four basic categories per superordinate category and four subordinate categories per basic category. A large and consistent increase in similarity of the overall look of objects was obtained for basic level over superordinate categories. A significant but significantly smaller increase was observed from basic to subordinate.

In Experiment 4, averages of the shape of objects were computed for categories at all levels of abstraction. These are Platonic prototypes so to speak. This procedure might appear less arbitrary if you consider that for Rosch the basic-level is to levels of categorisation what a prototype is to its contrasting categories (see also Halverston, 1992, on Palaeolithic art). Averages of superordinate objects could not be identified as such any better than at chance level; basic-level objects were the most inclusive categories at which averaged objects were readily identifiable.

Experiments 5 and 6 used the nonbiological objects from Experiments 1 and 2. The former two experiments were designed to assess the nature of representations at different categorisation levels. In Experiment 5, the name of an object was given at one of the three levels of categorisation to subjects as a cue before they were shown that object for 200 ms either on the right hand side of a screen or on the left hand side (a mask was presented on the other side). The experiment is based on the premise that if the representation is shape-based and isomorphic to the object, then cueing with its name will help its detection. Superordinate names did not help; and basic names helped just as much as subordinate ones (this has now been replicated by a number of researchers; e.g., Biederman, 1990). In Experiment 6, the same procedure was used with a same-different task. Under physical identity, only the basic-level and subordinate names primed the speed at which the subject could say that it was the same. Which suggests, again, that an isomorphic representation could be activated by subjects in these cases.

Rosch et al. conducted the first verification experiment at different levels of abstraction. The rational was the following: maybe some of the basic-level advantages come from objects being first apprehended at this level (cf. prototypical categories are verified faster than the others; Rosch and Mervis, 1975). If this is true, people should be able to verify membership at the basic level more rapidly than at any other level. This has become the standard procedure to assess the performance at various categorisation levels. In Experiment 7, they used the objects and taxonomies from Experiments 5 and 6. A name was first presented to participants, which was followed by the picture of an object. The participants had to say whether or not the two matched as quickly as possible. The verification RTs for correct positive items were fastest at the basic level and slowest at the subordinate. (This experiment is further discussed in Chapter 4; the average verification times are reported in Table 6a.)

If objects are first apprehended at the basic-level by parents, maybe they are first learned at that level by children. This had already been suggested by Brown (1956) and by Stross (1973). In Experiment 8, triads of objects were shown to 3-yr-olds, 4-yr-olds, kindergarteners, first graders, third graders, and fifth graders. Their task was to pick out the odd object. In half the triads, a pair of objects matched at the superordinate level (e.g., *Mackintosh apple*, *Freestone peach*, and <u>*claw hammer*</u>); in the other half a pair matched at the basic level (e.g, *Mackintosh apple*, *Delicious apple*, and <u>*claw hammer*</u>). Note that no subordinate triads were used. This was decided because children do not know many subordinate categories. At all ages, the children picked out the odd element in the basic-level triads. For the superordinate triads, 3 yr-olds succeeded at 55% and 4 yr-olds at 96%.

Rosch et al.'s Experiment 9 was similar to their Experiment 8 but it used a different sorting task. A set of pictures were freely grouped into categories. If they were not taxonomic, the experimenter asked whether a different grouping was possible, and if so which one. The results are similar to those of Experiment 8. All children could sort objects at the basic level; only third graders and above could do so for the superordinate level.

Experiment 11 examined a well-documented case history: the protocol for spontaneous speech of infant Sarah (Brown, 1974). Basic-level names were by far the most frequently present. This suggests again that basic-level names are the first ones acquired by children.

Next Rosch et al. asked whether the cause of this learning age effect was the parents giving objects basic names in free naming. You probably recognised Brown's (1958) hypothesis. This was hinted at in the verification experiment. Experiment 10 tested the free naming preference hypothesis more directly. All the objects in Experiment 1 were used. Adults overwhelmingly named the objects at the basic level. Remember that this is supported by informal observations made by cognitive anthropologists as well as by Brown's (1958) linguistic intuitions.

Rosch et al. also tested whether these findings generalise to other languages. They reasoned that if a language has a poorer vocabulary than say English, its basic-level names will be the least affected. Experiment 12 is an informal consideration of American sign language. For nonbiological things, more basic-level names have consistent signs or sign combinations than names at any other level of categorisation.

One more issue discussed in Rosch et al. became an important research trend in the basic-level literature: the effect of expertise on basic-levelness. To most people the basic category "airplane" is basic. However, for one participant of Experiment 1, an airplane mechanic, the superior categories were at the subordinate-level of abstraction (e.g., "Boeing 747", "Concorde"). Along the same line, you will remember that Rosch et al. (1976) found that for biological categories such as trees, their college student subjects seemed to have a basic level above the folk generic level reported by Berlin (1972, 1992; Berlin et al., 1973) with aborigines (i.e., at the family level rather than at the genus level). They argued that this was due to a lack of expertise triggered by a lack of usefulness. This is quite reminiscent of Brown's numismatist example.

1.3 After Rosch and colleagues

The empirical research to date on the basic-level has been completely shaped by the work of Rosch and colleagues. We have divided the more recent basic-level literature into five themes that you will immediately recognise: (1) tests of Rosch's differentiation model (cf. Rosch et al., 1976, Experiments 1 and 7), (2) attempts to generalise the basic-level phenomenology to other domains (cf. Rosch et al., 1976, Experiment 12), (3) assessments of the importance of shapes as a basic-levelness determinant (cf. Rosch et al., 1976, Experiments 3 and 4), (4) studies of the effect of expertise on basic-levelness (cf., Rosch et al.'s, 1976, general discussion), and (5) evaluations of the hypothesis that basic-level names are the first learned by children (cf., Rosch et al., 1976, Experiments 8, 9, and 11; Brown, 1956; Stross, 1973).

1.3.1 Tests of the differentiation model

Rosch et al. (1976) proposed a model–*category differentiation*–that somehow optimises *distinctiveness* (i.e., "[...] have the least attributes shared with members of other [contrasting] categories.", Rosch et al., 1976, p. 435) and *informativeness* (i.e., "[...] have the most attributes common to members of the category [...]", Rosch et al., 1976, p. 435). Just how distinctiveness and informativeness are integrated is unclear (see Chapter 3 for a more detailed discussion). However, most experimenters–including Rosch et al.–have understood the differentiation model as the *sum* of the within-category similarity and the between-contrasting-category dissimilarity (cf. Tversky's, 1978, special contrast model). Rosch et al. showed that basic categories are indeed more differentiated in this restricted sense of the term than natural taxonomies.

Murphy and Smith (1982) were the first to test this special differentiation model using artificial objects (see Figure 8 for sample objects). Their experiments were aimed at dissociating Rosch et al.'s differentiating account from three alternative explanations: basic-level categories are superior because (1) they are first learned, (2) of the frequency of the category, or its name, or both, and (3) the conjoint frequency between a category, or its name, and an object (remember: this is one of Brown's, 1956, suggestions). As measures of informativeness and distinctiveness, they used, respectively, the number of uniquely defining *construction features* that were shared within category and between contrasting categories. To illustrate, consider their Experiment 1 taxonomy shown at the bottom of Figure 1 (see Chapter 2). Underneath the category names (e.g., "hob", "bot", and "com"), we give in an abstract form all the uniquely defining construction features. For example, a hob can be identified by feature *a*, a bot by any one of features *c*, *d*, or *e*, and a com by feature *o*. The informativeness of the hob category is 1 (it has a single within-category feature shared by all its members), and its distinctiveness is 1 (it has a single between-contrasting-category feature shared by all its members). The differentiation of hob is thus 2 (an informativeness of 1 + a distinctiveness of 1). Similarly, we find that the bot and com categories have informativeness (as well as distinctiveness) scores of 3 and 1, respectively. The differentitiation of bot is thus 6, and that of com, 2. Therefore, differentiation is maximised for the bot category, and is equally low for the hob and com categories. Generalising to categorisation levels: the middle-level is more differentiated than the high- and low- levels which are equally differentiated. In a series of three experiments they showed that basic-levelness varied with differentiation, not with the other factors (see Chapter 4 for more details).

The most systematic tests of the differentiation model were carried out by Murphy (1991a). To minimise confounds and maximise control, he used artificial taxonomies (see Figures 8 and 15 for sample objects). In a series of five experiments, he showed that basic-levelness was function of differentiation (see Chapter 4 for more details).

Experiments with artificial taxonomies are suggestive but still leave us with doubts as to whether their results can be generalised to much richer natural taxonomies (e.g., Tversky and Hemenway, 1991). We mentioned already that Rosch et al. demonstrated that basic categories were the most differentiated in natural taxonomies. This has been replicated over and over again (e.g., Mervis and Crisafi, 1982; Tversky and Hemenway, 1984; Tanaka & Taylor, 1991; Johnson and Mervis, 1997). The trouble with natural taxonomies is control. How can one show that differentiation and basic-levelness *vary* together, that they do not just co-occur by accident, in natural taxonomies? Two tactics have been used: Brown (1956), Berlin (1992), and Rosch et al. (1976) suggested that experts have more differentiated subordinates. Palmer, Jones, Hennesy, Unze and Pick (1989) demonstrated it experimentally with musicians and nonmusicians using a musical instrument taxonomy. Then Tanaka and Taylor (1991)–using dog and bird experts–showed that basic-levelness does co-vary with differentiation. We will soon devote a whole section on expertise and basic-levelness. For now we will leave it as it is.

Two groups of researchers arrived at another tactic independently (Jolicoeur, Gluck and Kosslyn, 1984; Murphy and Brownell, 1985): Subordinate categories are by definition very informative but they lack distinctiveness, and this is the basis of the basic-level advantage. However, there are some subordinates that are very distinctive, namely atypical subordinates (Rosch and Mervis, 1975). For example, penguins are "birds", but they are very distinct from other birds; electric knives are "knives" but distinct from other knives. So *basic-atypical* subordinate categories–as Murphy and Brownell named them–should behave more like basic-level categories than subordinate categories. This is exactly what they found. Note that this also refutes other accounts: atypical terms are typically longer, learned later, and less frequently used.

Markman and Wisniewski (1997) pointed out that the notion of differentiation (i.e., high within-category similarity and low between category similarity) does not acknowledge that a pair of categories can be dissimilar either (a) because it has few *commonalties* (e.g., no common dimension) or (b) because it has many *alignable differences* (e.g., many different values on the same dimensions). A car is dissimilar to a motorcycle because it has four wheels instead of two; this is an example of an alignable difference. A car is also dissimilar to a motorcycle because it has a jack. This is an example of lack of commonalty. Markman and Wisniewski showed that natural superordinate categories are dissimilar because of few commonalties, and that natural basic categories are dissimilar because they have many alignable differences. Alignable differences have a number of advantages over nonalignable ones: they are more focal in similarity comparisons (Markman and Gentner, 1996); they are more likely to be used in other cognitive processes involving comparisons such as decision making (Markman and Medin, 1995), conceptual combination (Wisniewski, 1996; Wisniewski and Markman, 1993), and concept formation (Wisniewski and Markman, 1996). It seems appropriate that the level of categorisation that has an advantage in a variety of cognitive tasks is marked by the presence of differences that are also privileged in a variety of tasks.

1.3.2 Other domains with a basic-level

If the basic-level phenomenology is a normal consequence of categorisation in taxonomies, then it must be found everywhere taxonomic categorisation occurs. In fact, the basic-level phenomenology seems to hold across a variety of domains (for a review see Murphy & Lassaline, 1997), such as (1) American sign language (Newport and Bellugi, 1978), (2) environmental scenes (Tversky & Hemenway, 1983), (3) events (Morris & Murphy, 1990; Rifkin, 1985; Rosch, 1978), (4) personality types (Cantor & Mischel, 1979), (5) clinical diagnosis (Cantor, Smith, French, and Mezzich, 1980), (6) emotions (Shaver, Schwarz, Kirson, and O'Connor, 1987), and (8) computer programs (Adelson, 1983). It is difficult to avoid a dull enumeration here.

(1) Remember that Rosch et al. (1976) found that, in American Sign Language (ASL), more basic-level names have more consistent signs or sign combinations than names at any other level of categorisation. Newport and Bellugi (1978) further explored the basic-level and ASL question. They learned that basic names are usually depicted by primary ASL signs (e.g., *chair*) whereas the superordinate terms are represented by a series of prototypical basic terms included in it (e.g., *furniture = chair-table-lamp*, etc.) and the subordinate names are either made of (a) compound signs composed of regular basic ASL signs (e.g., *kitchen chair = cook-chair*), (b) compound signs composed of regular basic signs in conjunction with size-and-shape standard specifiers (e.g., *park bench = chair-"oblong"*), or (c) conjuncts of regular basic signs and mimetic non-standard depiction of the shape of objects (e.g., *hacksaw = saw-"hacksaw-shaped"*). The special character of the basic-level terms is shown both by the fact that they are represented by primary terms and by the fact that they are often the components of superordinate and subordinate terms.

(2) Tversky and Hemenway (1983) examined whether or not environmental scenes possess anything like a basic-level. Scenes are processed differently than objects. For example, a city scene can be recognised as such before any of its individual components (e.g., Schyns and Oliva, 1994; Oliva and Schyns, 1997). Tversky and Hemenway used "indoor" and "outdoor" as their superordinate categories. They then selected the four more frequently named scenes in these two superordinate categories to create their basic-level categories: "home", "restaurant", "store", "school", "mountain", "park", "beach" and "city". Their subordinate categories were two subsets for each basic category. The increase in listed attributes, actions, and parts was greater from superordinate to basic (e.g., "indoor" to "school") than from basic to subordinate (e.g., "school" to "elementary school").

(3) Rifkin (1985) investigated Rosch's (1978) idea that event taxonomies also have a basic-level. He asked a group participants to produce what they thought were "basic events" (e.g., meal, entertainment, sports, crime), then he asked another group of participants to produce superordinates (i.e., "This is a type of what?") and subordinates (i.e., "What are examples of this activity?"). A third group of participants were asked to list attributes for all these categories. Rifkin found that the increase of listed attributes augmented significantly more from the superordinate to basic than from basic to subordinate, and that few attributes were listed at the superordinate.

Morris and Murphy (1990) replicated this with several basic-level correlates, including verification latencies and frequency of use in free naming. They found that verifications were fastest and that people preferred to name event at the basic level.

An effort by Landau (1996) to study event taxonomies in a more natural setting deserves mention. He studied a large corpus of cinematic shots from the movies of Alfred Hitchcock, and found that most of them were basic events.

(4) Cantor and Mischel (1979) searched for a basic-level in personality taxonomies. They asked participants to list attributes for hierarchically organised categories such as “emotionally unstable person”, “committed person”, and “cultured person” at their highest level of abstraction; “criminal madman”, “religious devotee”, and “patron of arts” at their intermediate level; and “strangler”, “Buddhist monk”, and “supporter of community orchestra” at their lowest level of abstraction. More attributes were listed at the middle-level than at the two others.

(5) Cantor, Smith, French and Mezzich (1980) studied the same question with categories used in diagnostic manuals at that time. They asked experienced clinicians to list features for categories at different levels of abstraction (e.g., “functional psychosis”, “schizophrenia”, and “paranoid schizophrenia”). The basic-level signature was discovered: more attributes were listed at the middle-level than at the two others.

(6) Shaver, Schwarz, Kirson and O’Connor (1987) showed that emotions are hierarchically organised and that they have a preferred level of abstraction. Their participants sorted 135 emotion terms. A hierarchical cluster analysis was conducted on the data from the sorting. The emotions clustered into a three-level hierarchy: They were two clusters, “positive” and “negative” emotions, at the most general level; there were five groups, which they interpreted as “love”, “joy”, “anger”, “sadness”, and “fear”, at the intermediate level; and there were 25 clusters (e.g., “cheerfulness”, “contentment”, and “pride” were subdivisions of “joy”), at the most specific level. The intermediate-level emotion categories corresponds roughly to what people say when asked to name emotions (plus “hate”, perhaps); to emotions children learn to name first; and, finally, to emotions that theorists have classified as primary.

(7) Finally, Adelson (1985) examined computer programming concepts hoping to find evidence for a basic level. Expert programmers were asked to enumerate the attributes of names of concepts at a general level (e.g., "algorithm", "data structure"), at an intermediate level (e.g., "sort", "tree"), and at a specific level (e.g., "insert", "binary"). The increase from most general to intermediate was larger than from intermediate to specific, and few attributes were listed at the most general level. Participants also chose the intermediate categories more often in a partitioning task.

1.3.3 Importance of shapes for basic-levelness

Remember that Rosch et al.'s (1976, Experiment 3) found a large and reliable increase in similarity of the overall look of objects from basic to superordinate categories, and a significant–but significantly smaller–increase from basic to subordinate. Furthermore, Rosch et al. (1976, Experiment 4) found that averages of basic-level objects are the most inclusive categories at which average objects are readily identifiable[4]. These findings suggest that shape plays an important role at the basic-level.

Shapes are composed of *parts* in a certain *configuration*. For example, a prototypical house is a *wedge* on top of a *cube*. What is more determinant for basic-levelness? Tversky and Hemenway (1984) found a sharp increase of listed part features–not spatial relationships between them–from the superordinate to the basic level (e.g., *handle* and *blade* for

[4] Halverson (1992) argued that Upper-Palaeolithic drawings are just such basic-level averages (e.g., bison or horse). Interestingly, the drawings from this period have canonical views (Palmer, Rosch & Chase, 1981), they are "abbreviated" (missing feet, missing head, etc.), of variable sizes, and, typically, only made of shapes (neither colour, nor texture is represented).

"knife"; *peel* and *pulp* for "banana"), but little rise from the basic to the subordinate level for a broad range of natural categories including both objects and living things.

Additional evidence comes from a study by Klatzky and Lederman (1995) on identifying objects from haptic glances that is, only touching the object for a short duration. This procedure gives participants access only to local parts (and texture) information. At 200 ms exposure time, basic naming accuracy was above chance, and providing the superordinate name did not increase the accuracy significantly.

It seems that parts–not their spatial configuration–are the critical determinant of the correlation between shape and basic-levelness. Is it possible to restrict the range even more?

McMullen and Jolicoeur (1992) suggested that the basic-level categories are defined by *geons* and that subordinate categories require additional shape processing such as determining the spatial relationship between geons. These include parts such as *cube*, *sphere*, *cylinder*, and so on. Biederman (1987) showed that *geons* are invariant through most views and most orientations. If McMullen and Jolicoeur are correct, basic-level categorisation should be mostly orientation and size invariant, like geon recognition. However, subordinate-level categorisation should be orientation dependant (e.g., ON THE RIGHT OF could become ON THE LEFT OF), but size invariant.

Hamm and McMullen (1998) tested the orientation predictions. They asked participants to name as quickly as they could rotated objects (between 0° and 120°) at the superordinate (e.g., animal), basic (e.g., dog), and subordinate (e.g., collie) levels. As predicted, they found a large effect of rotation on subordinate naming, but little on basic and superordinate naming.

The size predictions were examined by Archambault, Gosselin and Schyns (in press). They used eight animal species, such as "dog", "cow", and "frog", each divided in two subordinate categories, such as "Doberman dog", "German shepherd dog", "Holstein cow", "Friesian cow", "leopard frog", and "rhino frog". The exemplars were presented at six different sizes between .38 and 12 deg. Participants were submitted to one of two tasks: In the discrimination task, participants were shown two animal exemplars simultaneously, and then were asked either the subordinate question "Was it the same animal?", or the basic one "Was it the same animal category?". In the categorisation task, participants were presented one animal exemplar and then asked either a subordinate question such as "Was it a Holstein cow?" or a basic question such as "Was it a cow?". For both tasks accuracy was size invariant at the basic but not at the subordinate level. This goes against McMullen and Jolicoeur's (1992) predictions.

In a series of five experiments, Murphy (1991a; see also 1991b) tested whether the relationship between geon and basic-levelness is *necessary* (i.e., the degree to which a taxonomy does not have parts collected at one level, it will not display basic-level phenomena) or *sufficient* (i.e., the degree to which a taxonomy has parts collected at one level, it will tend to display basic-level phenomena). He found that parts are neither necessary, nor sufficient. His experiments are described in greater detail in Chapter 4.

Corter, Gluck and Bower (1988) also refuted Murphy's necessary clause. They used artificial disease categories defined in terms of *verbal* or *conceptual* features rather than perceptual ones. The artificial diseases were characterised by symptoms such as *swollen*, *discoloured*, *bleeding*, or *sore gums*; *puffy*, *sunken*, *red*, or *burning eyes*; and *blotchy* or *scaly rash*. The middle level categories were verified faster (or, if you prefer: they had greater basic-levelness) than the two others which were verified equally as fast. With hindsight we could say that Murphy's necessary condition has been refuted a number of times by basic-level experiments in domains without shapes (Adelson, 1985; Cantor and Mischel, 1979; Cantor, Smith, French and Mezzich, 1980; Morris and Murphy, 1990; Rifkin, 1985; Shaver, Schwarz, Kirson and O'Connor, 1987; see section *1.3.2 Other domains with a basic-level*).

To summarise, the correlation observed between basic-levelness and geons seems to be accidental, and therefore could reveal more about the structure of the world than about cognition.

1.3.4 Expertise and basic-levelness

Brown's (1956) principle of utility made it possible for different people to have different "levels of usual utility". Remember his coin collector example: To this connoisseur a priceless 1910 dime is a "priceless 1910 dime"; to most of us, it is a "dime" or "little money". In the cognitive anthropological literature, the effect of expertise on basic-levelness is also mentioned: Boster (1980) found that manioc cultivators among Aguaruna tend to refer to manioc plants by specific rather than generic names, whereas other members of the community used the generic label as expected; Berlin (1992) noted that Aguaruna women, who spend less time in the forest than men, may fail to differentiate among members of some scientific genera of forest birds which men do name distinctively; and Dougherty (1978) observed that urban American children, in contrast to the Tzeltal children studied by Stross, appear to learn supra-generic distinctions among plants first and may never learn more than about a dozen folk generic distinctions. Finally, you will remember Rosch et al. (1976) found that for biological categories, such as birds, their college student subjects seemed to have a basic level above the folk generic level reported by Berlin et al. (1973). Rosch et al. proposed that expertise increases differentiation (or the number of unique listed features) of subordinate categories.

Palmer, Jones, Hennesy, Unze and Pick (1989) tested this hypothesis. They asked musicians and nonmusicians to list features for categories, such as “string” and “woodwind”, at the musical instrument families level, and for categories, such as “clarinets” and “violin”, at the individual instrument level. They found an orthogonal pattern of responses: for musicians, categories were more differentiated at the level of individual instruments than at the musical instrument families level; and vice-versa for the nonmusicians.

Tanaka and Taylor (1991) conducted a more elaborate study involving dog and bird experts. These people performed three categorisation tasks involving exemplars from both the expert domain and the less familiar (novice) domain. The tasks were modelled after those of Rosch et al. (1976) and involved attribute listing, free naming, and category verification. In the feature-listing task, participants were asked to list attributes of superordinate-, basic-, and subordinate-level categories. Participants listed almost as many features for the subordinate-level categories within the domain of expertise (e.g., “robin” for the bird experts; “collie” for the dog experts) as they did for the basic-level categories (e.g., “bird” for the bird experts; “dog” for the dog experts). Outside their area of expertise they behaved like Rosch et al.’s (1976) subjects: the increase in the number of attributes was greater from superordinate to basic than from basic to subordinate. Bird experts were more likely to name pictures of birds with subordinate-level names than with basic names, whereas dog experts did not show any preference for either basic or subordinate-level names for pictures of dogs. For object within the domain of expertise, subordinate-level verifications were as fast as basic-level ones, and faster than superordinate-level ones. In other words, they found an overall increase in the accessibility of the subordinate level. However, the basic level retained its privileged status. (This experiment is further described in Chapter 4.)

Johnson and Mervis (1997) undertook the most ambitious of all research projects on the effect of expertise on basic-levelness. They conducted six experiments that extended Tanaka and Taylor's work in several ways: They used more degrees of expertise: advanced birdsong experts, intermediate songbird experts, tropical freshwater fish experts, and novices both in songbirds and tropical freshwater fishes. They studied four levels of generality that they called the superordinate, the basic, the subordinate, and the sub-subordinate. Finally, they employed many more basic-levelness measures: attribute generation, object naming, silhouette identification, silhouette discrimination, verification task, and an auditory priming task. Together, these experiments support the so-called *constrained basic-level malleability* view. "According to the *constrained basic-level malleability* view, the level that functions as the original universal basic level is determined through [fixed] perceptual structure. However, more specific sub-basic levels may also come to function as basic as result of intracultural variations in knowledge or intercultural variations in domain salience." (Johnson and Mervis, 1997, p. 249) (These experiments are described in detail in Chapter 4.)

All the expertise experiments reviewed so far lack control: Are all their expert participants experts to the same extent? Do they use the same strategies? It is impossible to say. A promising solution is the controlled creation of experts. Gauthier and Tarr (1997) have demonstrated that this is a possibility by having people become experts of "Greeble" categories (i.e., categories containing complex artificial creature-like objects). This inspired two basic-level experiments.

Lin, Murphy and Shoben (1997, Experiment 3) asked participants to perform a knowledge assessment task for half the subordinate-level categories (e.g., sedan, collie, coffee table, and jeans) of a taxonomy composed of vehicle, animal, furniture, and clothing at the superordinate level, and of car, truck, dog, bird, table, chair, pants, and shirt at the subordinate level. As a result these participant became "new experts" of half the subordinate-level categories. In a verification task, a significant change in the advantage of the basic over the subordinate level was observed between "unexposed" (122 ms) and "exposed" (33 ms) items.

Archambault, O'Donnell and Schyns (1999) investigated the hypothesis that expertise could influence the basic percept of an identical distal object. In Experiment 2, subjects learned to categorise four objects (two mugs and two computers) at the specific level and 26 objects at the general level (13 mugs and 13 computers). When objects were learned at the general level, a sentence printed at the bottom of each picture would either say "This is a mug" or "This is a computer," depending on its category. When objects were learned at the specific level, the sentence would individuate each object–e.g. "This is Mary's mug" or "This is Peter's computer," depending again on the category. Participants thus became experts of half the mugs and computers. These mugs and computers were inserted in a complex natural office scene. To tap into the visual encoding of objects (or percept), the authors used a change detection task (see Simons & Levin, 1997). In a trial, two office photographs were sequentially presented. Between the two frames, a mug could change (be replaced by a different mug) or disappear, a computer could change or disappear, or other office objects could disappear. The subjects task was to identify the difference between the two photographs. The frame sequence was repeated until subjects could identify the change correctly. The number of repetitions was used as a basic-levelness measure. Results indicated that physically identical changes were better perceived when subjects knew the objects at the specific level than when they knew them at the general level: subjects perceived a specific-level object change faster than a general object change.

1.3.5 The developmental literature

Brown (1958) proposed that children first learn "usual utility level" names because parents have a tendency to first use these names when asked "what is this thing?". Rosch et al. (1976) gave experimental support to this hypothesis. To summarise their findings: On the one hand, they demonstrated in a number of ways that basic-level names are first accessed by parents. On the other hand, they showed that 3 yr-olds are much more accurate at sorting objects at the basic- than at the superordinate-level. Moreover, they analysed the spontaneous speech of infant Sarah (Brown, 1974) and found that she acquired basic-level terms before superordinate names which in turn were learned before subordinate ones (Rosch et al., 1976).

There is an a priori reason to believe that Rosch et al's and Brown's explanation for the linguistic development of children is not the whole story. There *is* compelling empirical evidence that parents do in fact use basic names for things more often than any other kind (e.g., Rosch et al., 1976; Jolicoeur, Gluck & Kosslyn, 1984; Tversky & Hemenway, 1984; Murphy & Brownell, 1985; Tanaka & Taylor, 1991; Jonhson & Mervis, 1997). Let us assume that this is the case for the sake of argument. There is, however, a catch: how are children to match these basic names with their corresponding concepts (read: a category without a public tag or, alternatively, meaning in the extensional sense)? Children face the inverse problem parents face (remember our discussion of Brown's problem): among all the concepts that apply to a situation, how do they know which one to associate with the name uttered by their parents? This is known as Quine's "gavagai" problem in philosophy of mind[5].

Quine (1960) asks us to imagine that a linguist visits an unknown country and attempts to learn the native language. A rabbit passes by and a native of the country says "gavagai", pointing his index finger toward the little furry animal. How is the linguist to figure out what "gavagai" means? It could refer to "white", "furry", "medium-sized", "animal", "passing by", "this individual rabbit at the particular moment", and so on. The linguist has to formulate a hypothesis about the meaning, and that hypothesis requires testing. To test for the potential meaning of "gavagai", the linguist will point to certain things and ask whether it is a "gavagai". If the native denies that the thing is a "gavagai" then the hypothesis is rejected. This method can never settle on a single meaning because there will always be an infinite number of other hypotheses that are also consistent with the data.

To sum up: Learning a category consists in making the appropriate connection between a name and a concept. We assumed that parents designate things to their children with basic names. Thus if children are to learn new categories at all they must have a bias for basic-level concepts (Markman & Horton, 1980; Markman, 1987, 1989). Of course, this does not solve Quine's "gavagai" problem entirely. In the rabbit example, it does not explain why the linguist would understand "gavagai" as an

---

5 Hofstadter and the Fluid Analogies Research Group (1995) studied several such underconstrained situations. The "do-this" toy-problem is perhaps the most telling. Suppose that someone touches his nose with his index finger and asks you to do the same. What are you going to do? Are you going to touch your nose with your index finger? Or touch your nose with his index finger? Or touch his nose with your index finger? Or touch is nose with your nose? Or wiggle your index finger? And so on. Most people choose the first possibility listed above. Why is that so?

equivalent to the English "rabbit" rather than say "finger". But this basic-level bias solves the vertical component of Quine's problem.

Markman and Horton (1980) were the first to reveal this bias empirically. They assessed the ease with which children can learn basic and superordinate artificial animal categories, independently of parents' production biases. Their artificial animals were real animals with novel features. For example, a basic category was composed of salamanders with added *wings*, *feet*, and *"fourchu" tail*. Some distracting features were also added: *texture*, *locations of the wings and feet*. And a superordinate category was composed of animals with *horns* and a *feather tail*. The overall shape and some added features (e.g., texture, number of feet, etc.) were distracters here. They taught children either by showing them exemplars or by verbally describing the relevant features. Basic-level categories were easier to learn for pre-schoolers, kindergarteners, and first graders. Giving them verbal descriptions helped only for superordinate sets. And the pre-schoolers did not benefit from the verbal descriptions.

Mervis and Crisafi (1982) replicated Markman and Horton (1980) with a different basic-level measure and a different artificial taxonomy (see the middle taxonomy of Figure 7 and the sample objects of Figure 9 in Chapter 4). They used an odd task similar to the one devised by Rosch et al. (1976). They showed three sets of triads: subordinate, basic, and superordinate triads. All participants were equally good for the basic-level triads. For superordinate triads, the 5-6-yr-olds were just as good as 4-yr-olds and better than 2-3-yr-olds; and, for subordinate triads, the 5-6-yr-olds were better than the 4-yr-olds who were just as good as the 2-3-yr-olds. (This experiment is further described in Chapter 4.)

What the "gavagai" problem demonstrates without a doubt is that learning the correspondence between a name and a concept is underconstrained. So additional (and contingent) constraints must be used by children (see Marr, 1982, for similar situations in vision). If parents have a production bias for basic-level names, children must somehow constrain their comprehension towards basic-level concepts. However, there is some evidence that parents use a different language when talking to children (e.g., Brown, 1956; Callanan, Repp, McCarthey and Latze, 1994; Markman and Hutchinson, 1984; Mervis, 1987; Mervis and Mervis, 1982). Maybe children need a different kind of constraint to understand their parents. This is precisely what Markman and Hutchinson (1984) investigated. They used Rosch et al.'s odd task that is, out of triads objects they had to select they odd ones. Here is a sample triad from their experiment: police car (*standard object*), sports car (*taxonomic choice*), and policeman (*thematic choice*). For example, children had to respond to a puppet which said either "See this 'sud'? Find another." (new word condition) or just asked "See this. Find another." (no word condition). Eighty-three percent of children as young as 2 to 3-yr-old chose the basic-level option over the thematic one in the new word condition. In the no word condition, both options were chosen equally often (see also Michnick, Golinkoff, Shuff-Bailey, Olguin and Ruan, 1995; but see Callanan, Repp, McCarthey and Latze, 1994). This corroborates the basic-level comprehension bias hypothesis.

Where does this bias come from? Little works has been done on this question. One possibility is that the only concepts children can learn are basic-level ones, but some experiments rule this out. It seems that three- to 4-month-old infants can form categorical representations of basic- and superordinate-level for natural kinds and for artefacts (Eimas and Quinn, 1994; Behl-Chadha, 1996). This was found using a familiarisation-categorisation paradigm. In the learning phase, two stimuli from the same category are presented at the same time (e.g., two exemplars of the “cat” or the “chair” basic-level category; or two exemplars of the “mammal” or the “furniture” superordinate category). In the testing phase, two new stimuli–a new exemplar from the familiarised category (e.g., a new cat, chair, mammal, or furniture) and a distracter (e.g., a horse, bed, bird, or car)–are presented and staring preferences are recorded. If infants prefer staring at the distracter this is taken as an indication that a category was extracted during the learning phase. The premiss being that infants prefer novelty (e.g., Karmiloff-Smith, 1995). We will propose an explanation for the origin of the basic-level comprehension bias in Chapter 2.

Chapter 2. Strategy length & internal practicability[6]

In this chapter, we will introduce SLIP (Strategy Length & Internal Practicability). We will begin with the presentation of its two computational factors. We will then give an informal account of the model before formalising it. Following this we will adapt SLIP to disjunctions, to naming, and–to a lesser extent–to the other basic-levelness correlates. Finally, we will discuss where SLIP fits into the *diagnostic recognition* framework.

2.1 Two principles of organisation of information in hierarchies

Consider the top taxonomy in Figure 1 (adapted from Hoffmann & Ziessler, 1983, Hierarchy I). Each letter represents a feature. Underneath the category names, we give the abstract optimal strategies fed to SLIP. We will come back to this shortly. At the bottom of the taxonomy, the abstract feature constitution of all exemplars is given. Participants accessed the high- and mid- levels categories equally fast, and were slower for low-level categories.

The input can be classified as a *ril* if it possesses feature a. A categorisation strategy is thus $Strat(X, ril)$ = [{"does X possess a?"}]. Going down the hierarchy, a strategy for classifying the input as *kas* would be $Strat(X, kas)$ = [{"does X possess c?"}]. Note that this and the former strategy are equally long, because the feature they test (a vs. c) is unique to the category (*ril* vs. *kas*). In contrast, a strategy of length 2 is needed to classify the input in a low-level category. The input is a *lun* whenever one

[6] This chapter elaborates on Gosselin and Schyns (1997, 1998b, 1999) as well as Gosselin, Archambault and Schyns (in press).

of these two strategies succeeds: *Strat*(*X*, *lun*) = [{"does *X* possess *a?*"} & {"does *X* possess *g?*"}] or, alternatively, *Strat*(*X*, *lun*) = [{"does *X* possess *c*?} & {"does *X* possess *g?*"}]. These strategies are of length 2 because the feature *g* is present in the two low-level categories *lun* and *nub*[7]. One further feature test (on *a* or *c*) is necessary to determine the category membership of the object. Overlap between features is common in object taxonomies (think, e.g., of the number of objects having the same *colour*, or *having wheels*, or *having legs* and so forth). The length of a categorisation strategy measures the overlap between the features defining a target category and its contrast categories. At this stage, it is worth pointing out that *all* published verification studies–except Hoffmann and Ziessler's (1983, Hierarchy I)–have so far neglected the length of categorisation strategies. Because features do overlap across many real world categories, we will here acknowledge this fact and make strategy length the first computational determinant of SLIP.

[7]In the figures, we use a shorthand notation: [{"does X possess *a*?"}] = *a*; [{"does X possess *c*?", "does X possess *d*?", "does X possess *e*?"}] = *cde*; [{"does X possess c?} & {"does X possess *g*?"}] = *c&g*; and [{"does X possess *a*?} or {"does X possess *b*?"}] = *a* | *b*.

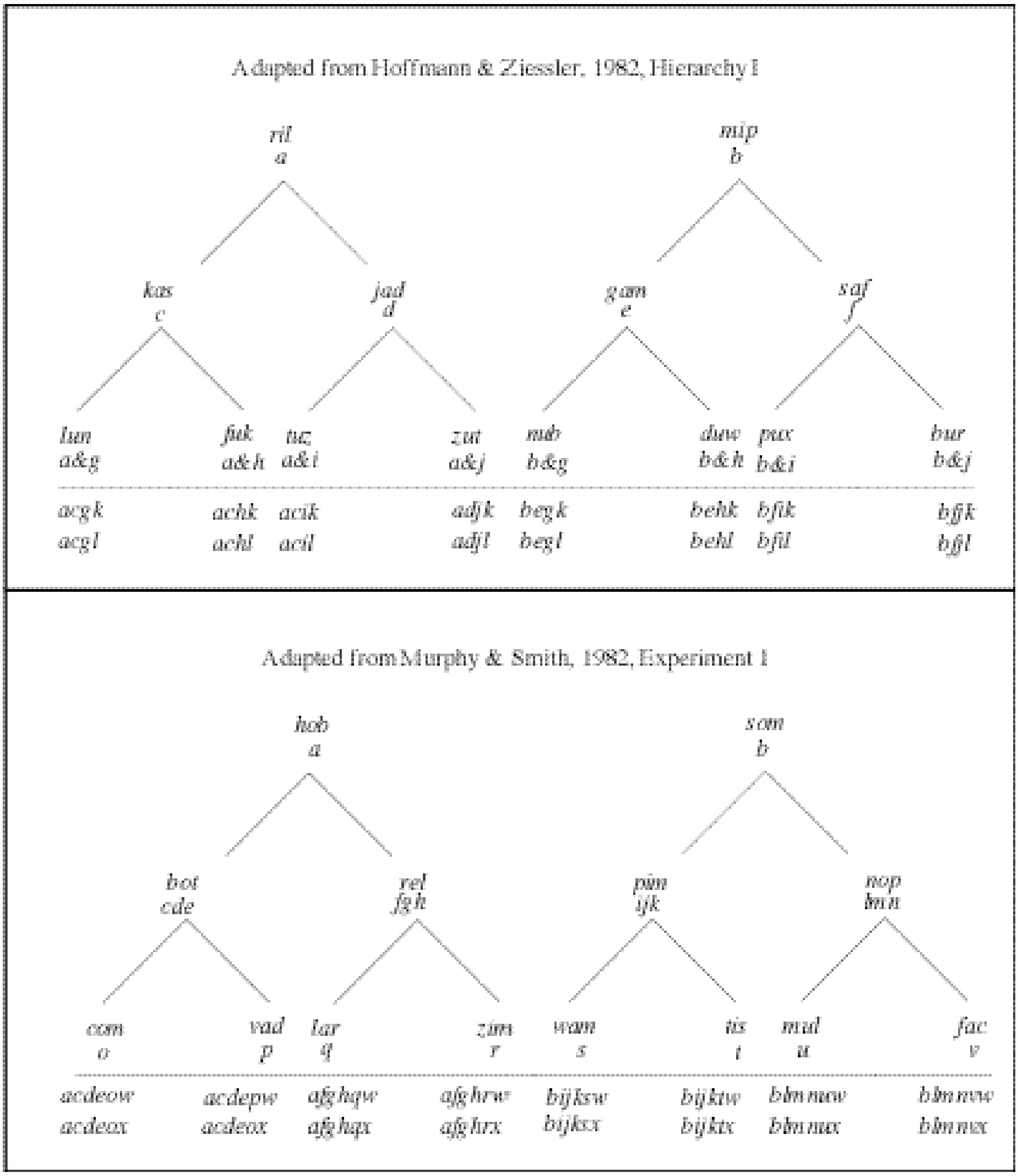


Figure 1. The top taxonomy is that of Hoffmann & Ziessler (1983, Hierarchy I), and the bottom taxonomy is that of Murphy & Smith (1982, Experiment 1). Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy.

The bottom taxonomy of Figure 1 (adapted from Murphy & Smith, 1982, Experiment 1) illustrates the second determinant of SLIP. Participants were faster at the middle-level, and slower at the higher-level. First note that features do not overlap between the categories and so strategies at all levels have the same length of 1. However, categories at the middle level all have many different features, each one of which is sufficient to access the category. To determine that the input is a *bot*, one can apply any one of the following feature tests: "does *X* possess *c?*", "does *X* possess *d?*", or "does *X* possess *e?*". For the purpose of categorisation, these tests on different features are redundant, and taken together, they form the exhaustive set of redundant feature tests to access the category. Feature redundancy is known to be an important component of speed of access to the levels of a taxonomy (e.g., Rosch et al., 1976; Murphy & Smith, 1982). It is the second computational determinant of SLIP. So, in the bottom taxonomy of Figure 1, a length 1 strategy to place an object in *bot* is *Strat*(*X*, *bot*) = [{"does *X* possess *c?*", "does *X* possess *d?*", "does *X* possess *e?*"}].

2.2 SLIP: an intuitive account

We will now develop SLIP (Strategy Length & Internal Practicability). It is an ideal categoriser insofar as it applies optimal feature testing strategies to determine the category membership of objects. We will explain what we mean exactly by "optimal" in the next section. A strategy comprises sets of features and SLIP tests their presence, one set at a time, in a specific order. Because features in a set are redundant, only one of them needs to be successfully tested to test the entire set. We assume that response time is a linear function of the total number of features tested when SLIP executes a strategy. With the varying strategy lengths of the top taxonomy of Figure 1, SLIP predicts faster verification speeds for the high- and mid- levels categories (both have length 1 strategies) than for low-level categories (which have strategies of length 2).

So far the model outlined never slips from an ideal feature testing strategy. However we wish to implement the idea that human categorisers approximate ideal strategies. To this end, we will assume that the processes of SLIP are noisy and sometimes slip off the ideal track to test random object features.

In general, slippage will increase the number of feature tests and the time taken to reach a category decision. However, slippage to a diagnostic feature is more likely for categories with many redundant features than for those with fewer features. Redundant features make categories more resistant to noise. In the bottom taxonomy of Figure 1, SLIP predicts a faster access to the more redundant middle level, even though the strategies have all an identical length of 1.

In summary, SLIP predicts that an object should be categorised faster in category X than in category Y (1) if the length of the optimal strategy that identifies the object as X is smaller than the length of the optimal strategy that identifies the same object as Y and (2) if the optimal strategy associated with category X comprises more redundant attributes than that of category Y.

2.3 SLIP: a formal model

We will make the simplifying assumption that SLIP always uses the strategy leading to a category decision in the shortest possible time. No doubt, this will turn out to be an oversimplification. For example, if an optimal strategy is very complicated, a simpler–but less efficient–strategy might be preferred.

A strategy succeeds whenever *all* of its sets of redundant attributes have been verified in order (e.g., Pashler, 1998; Treisman & Gelade, 1980; Wolfe, 1999; Woodman & Luck, 1999). A set of redundant attributes is successfully tested whenever one attribute[8] of the set is successfully tested. The probability of the success for the set of redundant attributes *j* is ${}_j$. Leaving aside the computation of ${}_j$ for the moment, we can cast the testing of a basic unit of a strategy (one set of redundant attributes) as a *Bernouilli trial.* A *geometric density function* specifies the probability that the set *j* has succeeded after exactly *t* attempts:

$$\left(1 \quad {}_j\right)^{t \quad 1} {}_j . \quad (1)$$

$\left(1 \quad {}_j\right)^{t \quad 1}$ is the probability that set *j* has not been successfully tested in the $t - 1$ first attempts, multiplied by ${}_j$, the probability that it succeeds on the *t*th attempt. When strategies comprise only one set of redundant attributes (length 1 strategies), Equation 1 fully describes the behaviour of SLIP. When ${}_j$ is large, $\left(1 \quad {}_j\right)$ is small, and thus the probability of completing this length 1 strategy decreases rapidly with *t*. It follows that a length 1 strategy is completed quickly on average.

${}_j$ implements internal practicability. It corresponds to the probability that one test in a set of redundant attributes succeeds at time *t*. A number of factors probably contribute to internal practicability. We have already mentioned redundancy or the cardinality of the set of redundant values. Saliency (e.g., Garner, 1978, 1983; Rensink, O'Regan & Clark, 1997) and expertise (e.g., Archambault and Schyns, 1999; Biederman & Schiffrar, 1987; Christensen, Murry, Holland, Reynolds, Landay & Moore, 1981; Goldstone, 1994; Norman, Brooks, Coblentz & Babcock, 1992; Quinn, Palmer & Slater, 1999) are other likely contributors. Here we will only consider redundancy and spatial configuration (as a secondary attribute). Greater redundancy makes the verification of a set

---

[8] In our view, there is no fundamental difference between feature and dimension. The latter is an ordered set of the former, and both express the idea that objects vary on N attributes. We will here use attribute, feature, and dimension interchangeably.

of redundant attributes easier, and a greater number of spatial configurations makes it tougher.

We know that configurations of features are important for recognition (e.g., Garner, 1974; Schyns & Rodet, 1997) and SLIP is sensitive to them. SLIP assumes that a configuration provides the context for the identification of single attributes. Suppose three primary redundant attributes *a*, *b*, and *c* arranged in this order to form a secondary attribute, the configuration *abc*. Suppose further that this configuration is one of 6 possible combinations of 3 attributes presented in an experiment. C_j is the probability that the categoriser identifies a configuration of redundant attributes (it is 1 over the number of possible configurations of redundant attributes; 1/6 in the example).

When SLIP verifies the presence of an attribute, its processes can randomly slip with a fixed probability S. Two independent events can lead to a successful test: (1) SLIP guesses the right configuration j and tests an attribute of set j. This happens with a probability of $C_j(1\text{-}S)$. (2) SLIP guesses the right configuration and slips, by chance alone, on an attribute of the tested set j. The probability of this event is C_jSR_j, where R_j is the number of redundant attributes divided by the total number of attributes in the input object (the index of redundancy). Thus ${}_j$, the probability of successfully testing one set j of redundant attributes, is $C_j(1\text{-}S+SR_j)$. Note that this constant C_j implies that a SLIP observer has no memory whatsoever of the checked configurations. There is some empirical evidence for this in humans. Horowitz and Wolfe (1998) asked people to search for a letter “T” among “L” distractors. In some trials, the letters were jumbled during the search, making it impossible for participants to keep track of their progress. This made no difference in the efficiency of the search.

So far, we have described SLIP for strategies composed of only one set *j* of redundant features (length 1 strategies). In this case, the speed of access to a category varies with only the redundancy of its attributes. We now turn to the modeling of SLIP for the general case of length *n* strategies. This requires the implementation of *strategy length*. The probability that a length *n* strategy is completed after *t* trials in a particular configuration of successes and misses is

$$\prod_{j=1}^{n} \left(1 \quad {}_{j}\right)^{\;{}_{j}} \;{}_{j}\,, \quad (2)$$

where ${}_{j}$ is a function of the tested set *j* that counts the number of failed attempts to verify this set after *t* trials (it will be further specified below). When $n = 1$, Equation 2 reduces to Equation 1. When $n > 1$, the solution is more involving. We know that the last set of an ordered strategy is always verified at time *t*. This implies that the previous sets have been verified sequentially in the *t*-1 preceding steps. There are different patterns of successes and misses of verifications of sets of attributes. For example, with $n = 3$, if set 3 is verified at time *t*, set 2 might have been verified at *t*-1 (or *t*-2, or *t*-3, or *t*-4, and so on), and set 1 anywhere between time *1* and *t*-2 (or *t*-3, or *t*-4, or *t*-5, and so on). For any ordered collection of sets of attributes in a strategy, given that the last set *n* succeeds at time *t*, the number of possible combinations of successes of verification of *n*-1 set in *t*–1 discrete steps is

$$= \begin{pmatrix} t \quad 1 \\ n \quad 1 \end{pmatrix} = \frac{(t \quad 1)!}{(t \quad n)!(n \quad 1)!}.$$

With the combinatorics of successful tests of multiple sets of attributes, the probability that a length *n* strategy is completed after *t* trials becomes

$$P(t) = \sum_{i=1} \prod_{j=1}^{n} \left(1 \quad {}_{j}\right)^{\;{}_{ij}} \;{}_{j}\,, \quad (3)$$

where ${}_{ij}$ specifies the number of failed verifications of the *j* th set of category attributes for the *i* th configuration of successes and misses.

The multinomial expansion $(a_1 + a_2 + \ldots + a_n)^{t \quad n}$ implements $_{ij}$: it expands into different terms, and the sum of the n exponents of each term is equal to t - n. Thus, the ith exponent of the jth term of the ordered expansion provides the number of failed verifications of the ith set of redundant attributes for the jth configuration of successes and misses. For example, with a strategy of length $n = 3$ verified in $t = 5$ discrete steps, we obtain the multinome $(a_1 + a_2 + a_3)^2$ which expands into the following = 6 terms: $a_1{}^0 a_2{}^0 a_3{}^2 + a_1{}^0 a_2{}^1 a_3{}^1 + a_1{}^0 a_2{}^2 a_3{}^0 + a_1{}^1 a_2{}^0 a_3{}^1 + a_1{}^1 a_2{}^1 a_3{}^0 + a_1{}^2 a_2{}^0 a_3{}^0$.

Random numbers variate for this general case can be obtained by adding the n appropriate geometric random number variate $\mathbf{G}_j$: $_j; j = 1,\ldots,n$, the geometric random numbers being computed from unit rectangular random number variate $\mathbf{R}$ by the relationship

$$\mathbf{G}: \quad \sim \log(\mathbf{R}) / \log(1 \qquad).$$

2.3.1 Summary

So far, we have identified two computational constraints on the organisation of information in a taxonomy of categories: the overlap of features between categories and the redundancy of features within categories. These two constraints determine different feature testing strategies to access different categorisations of the input. In general, greater feature overlap augments the length of a strategy, and higher feature redundancy augments its accessibility. SLIP implements the two constraints to predict the average number of feature tests required to resolve one categorisation strategy. Figure 2 schematises the functioning of SLIP.

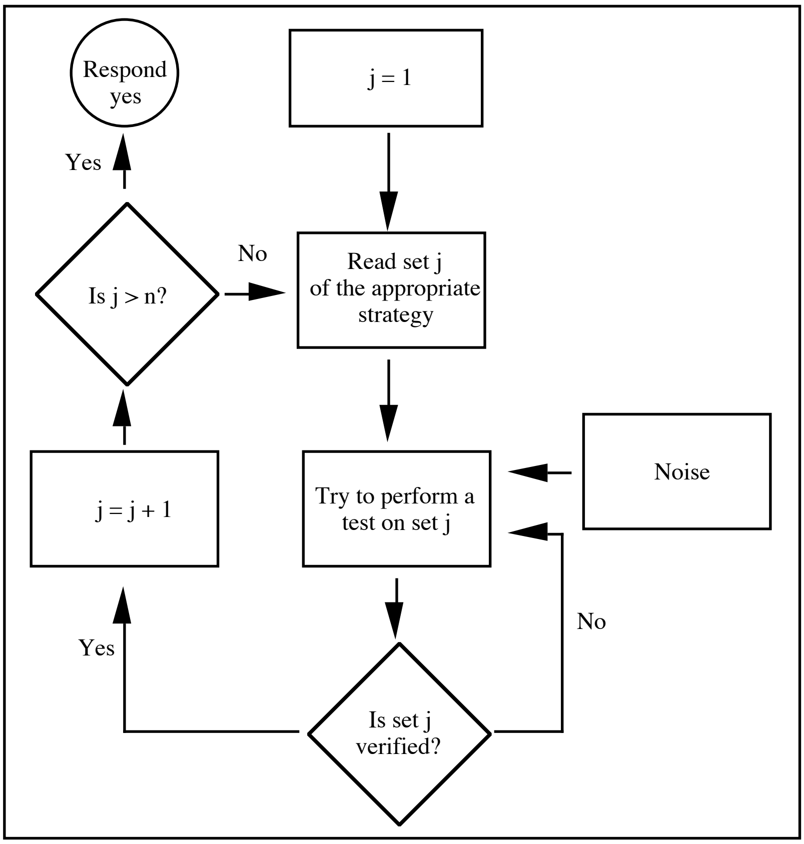


Figure 2. SLIP box diagram for positive verification items. The variables *n* and *j* are, respectively, the length of the considered strategy and a pointer to a set of redundant attributes.

Equation 1 implements the redundancy of features with the probability that one set of redundant features is successfully tested in *t* discrete tests. Equation 3 generalises Equation 1 introducing the idea that the *n* sets of attributes in a length *n* strategy have been successfully tested after *t* attempts.

2.3.2 The Pascal density function

Equation 3 is rather cumbersome. Fortunately, when all $_j$ within a category are equal (which is the case for all taxonomies reported in this dissertation), it reduces to

$$P(t) = \quad (1 \quad)^{t \ n} \quad ^{n}, (4)$$

a *Pascal density function.* Figure 3 shows four particular Pascal density functions. The *central limit theorem* implies that the limit of the Pascal density function as n approaches infinity is a Gaussian.

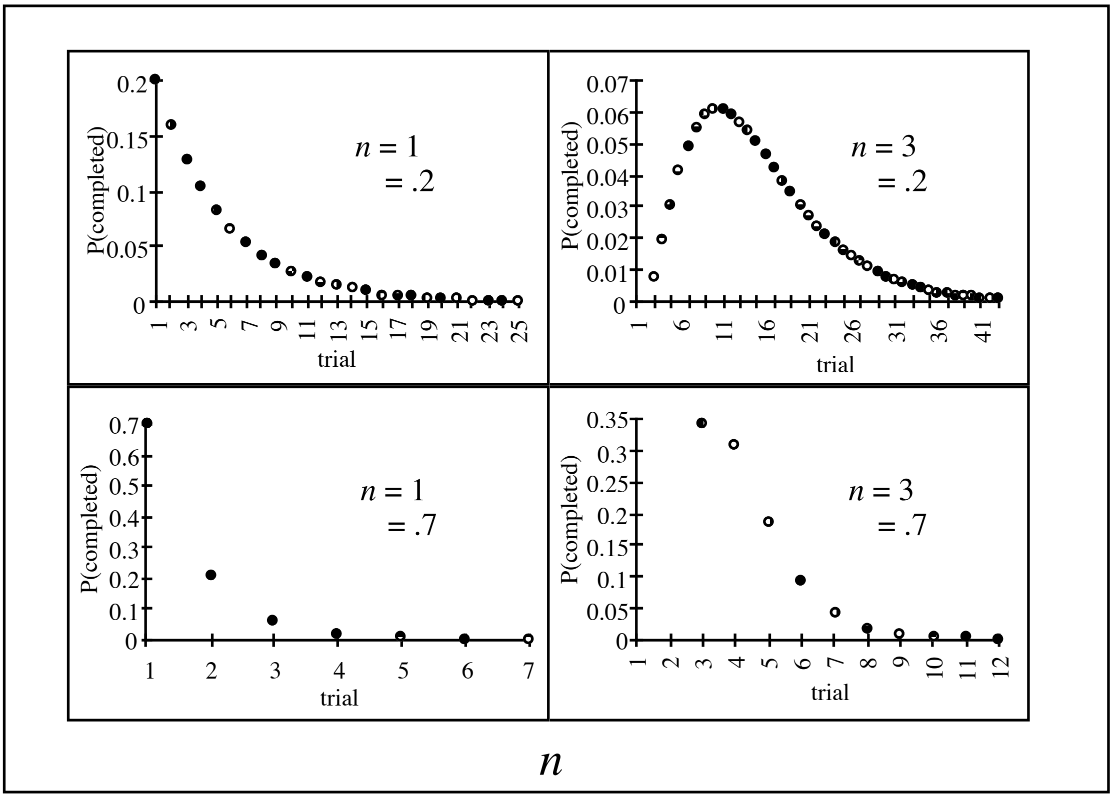


Figure 3. Instances of Pascal density functions.

The Pascal density function has a number of well-known characteristics. Of particular interest to us here: Its mean is equal to $n/$ and its variance to $n(1 \quad)/ \ ^2$; the gamma density function provides us with a continuous approximation of the Pascal density function:

$$\frac{{}^{n}t^{n \quad 1}e^{\quad t}}{(n)}$$

where (n) is the gamma function, a continuous function that approximates $(n \quad 1)!$, $n>0$ is the so-called *shape* parameter, and $0\leq \quad \leq 1$ is the *scale* parameter (e.g., Johnson and Kotz, 1969; Hastings and Peacock, 1975).

LaBerge (1962) used the *negative binomial* density function–a close parent of the Pascal density function–to model reaction times (for an excellent review see Luce, 1986). The negative binomial function gives the probability that n hits will have been encountered after x misses, hence $x = t - n$.

Throughout this dissertation, we will use the mean of the appropriate Pascal density function as a global measure of basic-levelness.

To illustrate, we will now apply this restricted version of SLIP to the taxonomies in Figure 1. Remember that in the top taxonomy strategy length varies and internal practicability is constant whereas the opposite applies to the bottom taxonomy. In SLIP terms, $_j$ will be constant in the top taxonomy, but vary in the bottom taxonomy, whereas n (the length of a strategy) will vary in the top taxonomy and be constant in the bottom taxonomy.

In the top taxonomy, the index of redundancy R (remember: it is the number of redundant attributes divided by the total number of attributes in the input object) is equal to .25 (i.e., 1 / 4 = .25), and the probability that the categoriser has properly identified a configuration, C, is 1 (i.e., 1 / 1 = 1). With the default value of $S = .5$, all $_j$ are equal to .63 (i.e., $C_j(1-S+SR_j) = 1*(1-.5+.5*.25) = .63$).

Thus, for the top taxonomy of Figure 1, SLIP predicts a mean number of feature tests of 1.6 (i.e., $n/$ = 1/.63 = 1.6) for the high- and mid-level categories, and 3.2 tests (i.e., $n/$ = 2 / .63 = 3.2) for the low-level categories. To compute the basic-levelness of a level of generality, we average the mean basic-levelness of all its categories. In this case, the high- and middle levels have a basic-levelness of 1.6 tests, and are accessed faster (with fewer feature tests) than the low-level one that has a basic-levelness of 3.2 feature tests.

Let us now consider the bottom taxonomy of Figure 1. For the high- and middle- levels, R = .17 (i.e., 1 / 6 = .17), C = 1 (i.e., 1 / 1 = 1); with S = .5, $_j$ is .58 (i.e., $C_j(1\text{-}S+SR_j)$ = 1*(1-.5+.5*.17) = .58). For the middle-level, R = .5 (i.e., 3 /6 = .5), C = 1 (i.e., 1 / 1 = 1), and $_j$ is .75 (i.e., $C_j(1\text{-}S+SR_j)$ = 1*(1-.5+.5*.5) = .75). Using the Pascal density function for each category and averaging within level of categorisation, we obtain a SLIP prediction of faster access to the middle level (SLIP = 1.33–i.e., *n*/ = 1 / .75 = 1.33), and slower access to the top and bottom levels (SLIP = 1.71–i.e., *n*/ = 1 / .58 = 1.71).

2.3.3 Disjunctions

So far, we have assumed that a conjunction of several attributes defines the categories of a taxonomy. "If we rely on intuitions (our own and those published by semanticists) and restrict ourselves to concepts about naturally occurring objects (flora and fauna), [...] we can think of no obvious disjunctive concepts." (Smith & Medin, 1981, p. 28) Even though we basically share this view, some artefact concepts are nevertheless obviously disjunctive. For example, a *strike* in baseball is either a *called*, or a *swinging strike* (Bruner, Goodnow, and Austin, 1956). Besides, several basic-level experiments have examined disjunctive categories.

SLIP can be modified to handle disjunctive categories. Consider, for example, the optimal strategy to access the high-level *ril* category in Hoffmann and Ziessler's (1983, Hierarchy II) taxonomy (see the top taxonomy of Figure 14 in Chapter 4): *Strat*(*X*, *ril*) = [{"does *X* possess *c*?"}] or [{"does *X* possess *f*?}]. The average number of feature tests required to determine that *X* belongs to the disjunctive *ril* is computed as follows: We start with the average number of attempts to complete the first strategy term (i.e., [{"does *X* possess *c*?"}]), weighted by the probability that it applies to the object. We then compute the average number of attempts to determine that the first strategy term does not apply and that the second one does apply; this figure is weighted by the probability that this situation will occur. This procedure is repeated until all the terms of the disjunction have been evaluated. The sum of all these quantities is the average number of feature tests needed to determine that *X* belongs to *ril*.

The first term strategy of the conjunction defining *ril* is completed after an average of 1.6 tests (i.e., = 5/8) and applies with a probability of .5. On average, 3.05 attempts are required to find out that the first term does not apply (we explain how to compute this in the following section), and an additional 1.6 tests to complete the second term of the disjunction–which applies with a .5 probability. Summing the tests across the two terms of the disjunction is .5 * 1.6 + .5 * (3.05 + 1.6) = 3.13 tests.

2.3.3.1 Negative verification items[9]

In a verification experiment, category name and object can either match (positive items) or mismatch (negative items). To respond to negative items, three qualitatively different methods can be used within SLIP: (1) a feature-counting method whereby object X is not a member of category Y if the features of X are exhausted and it has not yet been classified as a Y, (2) a contrast method whereby object X is not a member of category Y if it has already been classified as a member of a contrast category Z, and (3) a Fisherian probabilistic representation. We only describe the third method here.

If a classifier has failed to complete a verification of category Y after t attempts ($t \geq$ strategy length of Y), either the item is negative, or the classifier has so far slipped on irrelevant attributes. In SLIP, we can compute the likelihood of the latter because we know the density function of the number of trials necessary to complete the strategy. Based on this

---

[9] This section expands on Gosselin and Schyns (1998b).

distribution, a classifier could conclude that an item is false after *t_stop* trials, where *t_stop* is the number of trials beyond which the probability that the item is true is smaller than an acceptable level of error . This is the logic of Fisherian statistical testing. We used =.05 in all the simulations of this dissertation.

For length 1 strategies, *t_stop* is $\log(\ )/\log(1\ \ )$ (Gosselin and Schyns, 1998b). This is the *inverse geometric survival function* of probability . For example, before concluding that an object is not a *ril* (or, alternatively, that the *ril* strategy does not apply) in Hoffmann & Ziessler (1983, Hierarchy II, see top taxonomy of Figure 14), *t_stop* = 3.05 attempts are required, using the above estimate of and the default = .05 (i.e., log(.05) / log(1-.63) = 3.05).

For strategies of length n, the item can be discovered to be negative after 1, 2, 3, ..., or n - 1 successful set of redundant attribute tests. Thus the average number of attempts needed to conclude that a length n strategy does not apply (and hence that the item is negative) is a pondered sum of n terms. The first term is the average number of attempts required to conclude that the first set of redundant attributes does not apply (i.e., $\log(\ )/\log(1\ \ )$) pondered by the probability that this situation will occur (this depends on the design of the experiment). The second term is the sum of the average number of attempts required to check the first set of redundant attributes (i.e., $1/\ $) (In theory, one should modify the mean computation so that it includes points up to *t_stop*, not up to infinity. If is small, however, it makes very little difference. We will not modify it here.) and the mean number of attempts needed to decide that the second one does not apply (i.e., $\log(\ )/\log(1\ \ )$), weighted by the probability that this situation will happen. And so on until the nth term. The general formula is thus

$$\sum_{x=1}^{n} P_x \left(\frac{x \quad 1}{\ } + \frac{\log(\ )}{\log(1 \quad)} \right),$$

where P_x is the probability that the realisation that the item is negative happens after x - 1 successful tests of redundant attributes.

Figure 4 shows the SLIP complete box diagram for verification items, positive and negative items combined. It is essentially the same as that for positive verification items (see Figure 2) except for the more elaborate attribute-checking loop and the addition of a disjunctive-term loop.

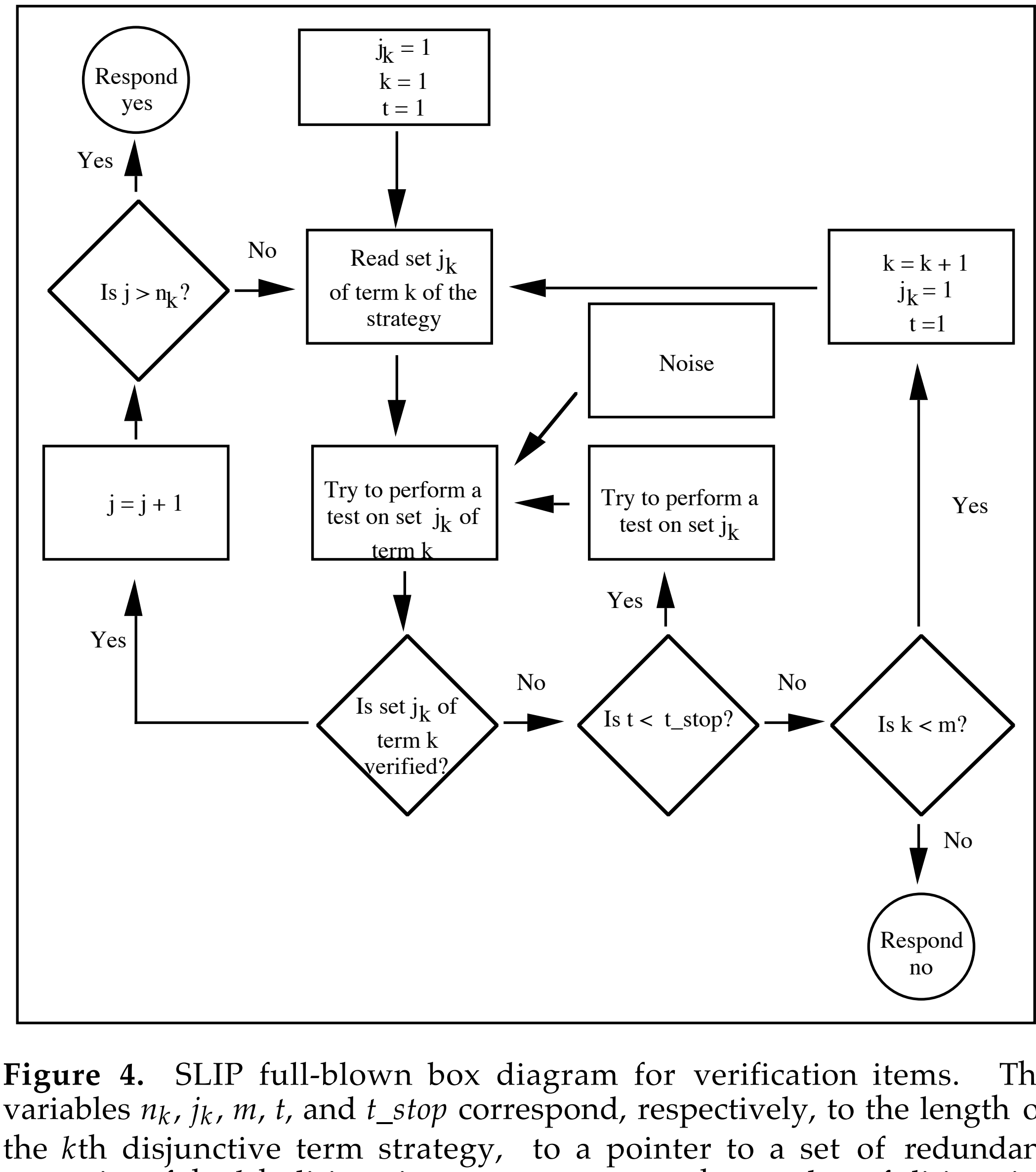


Figure 4. SLIP full-blown box diagram for verification items. The variables n_k, j_k, m, t, and *t_stop* correspond, respectively, to the length of the *k*th disjunctive term strategy, to a pointer to a set of redundant strategies of the *k*th disjunctive term strategy, to the number of disjunctive terms in the considered strategy, to a pointer to the number of attempts at testing the *j*th set of redundant attributes of the *k*th disjunctive term, and to the number of attempts at testing it before the probability that the item is positive reaches an acceptable level of error.

2.3.4 Naming

Although SLIP is primarily designed to model (positive) verification tasks–the most widely used in basic-level experiments–it is quite straighhtforward to extend it to predicting naming performance. When asked "what is this thing?", SLIP can apply most of its strategies in parallel, and output the name associated with the first completed strategy. Within each strategy, SLIP follows the order of sets of redundant features, but it performs its feature tests at random (i.e., $S = 1$). The rational for the latter is that there is no a priori reason why any particular attribute value tests should be performed. Again, this is an oversimplification. Consider, for example, the following situation: someone is asked to name an object from the top taxonomy of Figure 1 at the high-level. Two attributes values are diagnostic (i.e., *a* and *b*), and so this person would be better off testing for these rather than for all attribute values indiscriminately.

The functioning of SLIP for a naming task is summarised in the box diagram of Figure 5. Note that the model for verification (see Figure 2) and naming are very similar.

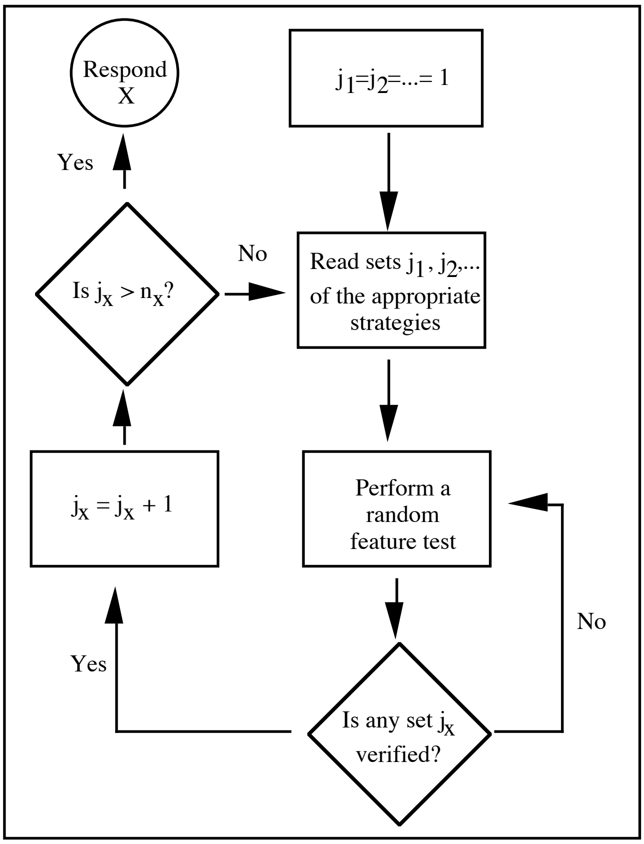


Figure 5. Box diagram of SLIP for naming. Variable j_x is a pointer to a set of redundant attribute of strategy X, and n_x the length of this strategy.

The main difference is the value of one parameter: S, the probability of slipping to a random feature (in the box diagram of Figure 5 this led to the removal of the noise box). When this probability increases, the number of feature tests required to complete a strategy increases proportionally (because more tests are made on irrelevant features). Thus, SLIP predicts the same qualitative order of speed of access in naming and verification, but it also predicts that the naming of an object will on average take longer than its verification (Rosch et al., 1976).

We will illustrate this with the two taxonomies in Figure 1. Remember that in the top taxonomy strategy length varies and internal practicability is constant whereas the opposite applies to the bottom taxonomy. In SLIP terms, $_j$ will be constant in the top taxonomy, but vary in the bottom taxonomy, whereas n (the length of a strategy) will vary in the top taxonomy and will be constant in the bottom taxonomy.

In the top taxonomy, the index of redundancy R is equal to .25 (i.e., 1 / 4 = .25), and the probability that the categoriser has properly identified a configuration, C, is 1 (i.e., 1 / 1 = 1). With the S naming value of 1, all $_j$ are .25 (i.e., $C_j(1\text{-}S\text{+}SR_j)$ = 1*(1-1+1*.25) = .25). SLIP predicts a mean number of feature tests of 4 (i.e., $n/$ = 1/.25 = 4) for the top and mid levels categories, and 8 tests (i.e., $n/$ = 2 / .25 = 8) for the bottom level categories (compare Hoffmann & Ziessler's data for the top taxonomy of Figure 1 in naming: high = ~750 ms, middle = ~1250 ms, and low = ~3000 ms; and in verification: high = ~500 ms, middle = ~500 ms, and low = ~700 ms).

Let us now consider the bottom taxonomy of Figure 1. For the high- and middle- levels, R = .17 (i.e., 1 / 6 = .17), C = 1 (i.e., 1 / 1 = 1); with S = 1, $_j$ is .17 (i.e., $C_j(1\text{-}S\text{+}SR_j)$ = 1*(1-1+1*.17) = .17). For the middle-level, R = .5 (i.e., 3 /6 = .5), C = 1 (i.e., 1 / 1 = 1), and $_j$ is .5 (i.e., $C_j(1\text{-}S\text{+}SR_j)$ = 1*(1-1+1*.5) = .5). We thus obtain a SLIP prediction of faster access to the middle level (SLIP = 2–i.e., $n/$ = 1 / .5 = 2), and slower access to the top and bottom levels (SLIP = 6–i.e., $n/$ = 1 / .17 = 6).

SLIP can also predict the probability that a particular name will be used in a naming task. This is also a common measure of basic-levelness. For the sake of simplicity, suppose that we are only processing length 1 strategies for a particular naming task. Let X be the completion of the target set of redundant features; N, the completion of one set of redundant features relevant for the considered naming task; and T, the completion of any set of redundant features. Probability theory gives us the following relationship: P(X | T) = P(X | S) * P(S | T). We are interested in finding P(X | S) that is, the probability that the target set of redundant features has been completed given that one set of redundant features relevant for the naming task has been completed. $P(X|T)$, the probability that the target set of redundant features has been completed given that one has been completed in the total set of redundant features, is equal to $P(X \quad T)/P(T) = P(X) = \ _X$. Similarly, $P(S|T) = P(S)$; all the sets of redundant features being independent, $P(S) = \ _i{}_i$, where i spans all the strategies that apply. Therefore $P(X|S)$ is equal to

$$\frac{_X}{_{i}\ _i}. \qquad (5)$$ [10]

The numerator is an inverted completion average, and the denominator is the sum of all of them. Hence the faster a strategy is completed in denomination, the more frequently it will be used. For the bottom taxonomy of Figure 1, for example, SLIP predicts probabilities of use of .6 for mid-level names (i.e., .5 / (.17 + .17 + .5) = .6), and .2 for high- and low-level names (i.e., .17 / (.17 + .17 + .5) = .2).

For a set of strategies of various lengths, the calculation is more arduous because of what happens over n_{max} (i.e., the length of the longest strategy considered) geometric variates have to be considered for all strategies. By "geometric variate", we mean a series of attribute tests that lead to the completion of a set of redundant attributes. For instance, in the top taxonomy of Figure 1, each object has three names, and n_{max} is equal to 2. The length 2 strategy (i.e., the low-level strategy) has won the race against the other two strategies when it has won both geometric variates. Both have a probability of .33 (i.e., .25 / (.25 + .25 + .25) = .33), and thus their combined probability is .33 * .33 = .11. Any length 1 strategy (i.e., the high- and the mid- level strategies) has won the race when it either won against the other two strategies on the first geometric variate, or on the

[10] Ties between competing strategies complicate the picture a little bit. If n is large, however, Equation 5 is a good approximation.

second. For example, the probability that the high-level strategy wins on the first variate (i.e., .25 / (.25 + .25 + .25) = .33) plus the probability that none of the two length 1 strategies wins on the first variate but that the high-level strategy wins on the second (i.e., [.25 / (.25 + .25 + .25)] * [.25 / (.25 + .25 + .25)] = .11) is equal to .44.

2.4 Generalisation to other correlates of basic-levelness[11]

We designed SLIP to model category verification, and we have extended its reach to naming. However, we pointed out earlier that a critical aspect of basic-levelness is that it optimises a *number* of indexes of performance (see Preamble and Chapter 1). It is important to show that SLIP is not limited to model category verification and naming.

You will remember that the basic-level is superior in many other respects: more features–especially shapes–are listed at the basic level than at the superordinate level, with only a slight increase at the subordinate level (Rosch et al., 1976; Tversky & Hemenway, 1984); and throughout development, basic level names are learned before those of other categorisation levels (Brown, 1958; Rosch et al., 1976; Horton & Markman, 1980; Markman, 1989; Markman and Hutchinson, 1984; Mervis and Crisafi, 1982). Furthermore, basic-levelness seems quite universal across domains (e.g., Murphy & Lassaline, 1997) as well as cultures (e.g., Berlin, 1992; Malt, 1995).

We will differentiate two types of basic-levelness correlates: the ones that we believe are connected to SLIP's inputs (i.e., strategy length or internal practicability), and the ones that we think are related to its outputs (i.e., verification or naming latencies). We have to stress that this section is highly speculative.

---

[11] This section expands on Gosselin and Schyns (1999).

2.4.1 Input correlates: proliferation of features, expertise effect, atypical subordinates, and cultural universality

Another correlate of the basic-levelness is that people tend to list many more features (especially shape) at this level than at others (Rosch et al., 1976; Tversky and Hemenway, 1984). Remember that most features of one basic-level category do not overlap with those of contrasting categories (e.g., Tversky & Hemenway, 1984, Tanaka & Taylor, 1991). Following the principles of SLIP, the addition of such diagnostic features in a category increases its internal practicability, and its basic-levelness. SLIP therefore predicts a proliferation of listed features at the basic level. The fact that these features are mostly shape rather than colour and texture could reflect the organisation of our perceived world.

A similar reasoning applies to the discovery of Tanaka and Taylor (1991), and Johnson and Mervis (1997) that expertise induces faster verification times (and number of listed features) for subordinate categories. It also applies to the observation of Jolicoeur, Gluck and Kosslyn (1984) and Murphy & Brownell (1985) that atypical subordinates (e.g., *penguins*, *electric knifes*) behave more like basic level categories than other subordinate categories (e.g., *robin*, *Swiss knife*). Murphy & Brownell (1985) have shown that these atypical subordinates are more informative (i.e., they have more listed features) and distinctive (i.e., they share fewer of these listed features with contrasting categories) than other subordinates. In other words, atypical subordinate categories have more internal practicability (are more redundant) than other subordinate categories. In sum, the computational principles of SLIP can account for the most important correlates of basic-levelness: faster verification, naming, and number of listed features.

Cognitive anthropologists have shown that folk taxonomies across cultures have roughly the same preferred level of abstraction. Given that the human experience is quite homogenous throughout cultures (we all breath air, eat, sleep, and procreate)–and thus so is the input to humans' SLIP modules–, we would expect the preferred level of categorisation to be roughly culturally universal.

2.4.2 Output correlates: learning rate and domain universality

Turning to development, it has been suggested that children have a *comprehension* bias (innate or learned) for the basic-level. Because adults show the bias in *production*, this would enable children and adults to resolve the level of categorisation ambiguity and understand each other (Markman & Horton, 1980; Markman, 1989). We have argued here that the production bias for basic names in adults arises from the organisation of their taxonomic knowledge and the resulting strategies that access the categories. The developmental literature is unclear about the origin of the bias for children to comprehend at the basic level. SLIP suggests that infants acquire concept taxonomies (e.g. Eimas & Quinn, 1994), and access them following the general principles of SLIP. Adults would produce basic names because they are first accessed in their "mental race", and children would connect these names with basic concepts because the latter are also accessed first in their "mental race". This does not imply that the taxonomic organisations of adults and children are identical, only that the same categories are first accessed. In other words, adults and children can differ markedly in the number of categories and levels of categorisation they have in memory, but still access the same basic level categories.

Note that this applies to all taxonomies, including taxonomies in domains as varied as computer languages, emotions, events, and so on. Hence the domain universality of basic-levelness.

2.5 A special diagnostic recognition model[12]

It is worth remembering that people who categorise a visual input seek to obtain a close match between a category representation and a representation of the object in the input. This match between memory and input information is what we call *a task* for the observer. Generally speaking tasks are not rigid. Instead, different categorisations of an identical object tend to change the information requirements of the task at hand. For example, to assign a visual event to the *Porsche, collie*, *sparrow*, *Mary*, or *New York* category comparatively more specific information may be necessary than when categorising the event as a *car*, *dog*, *bird*, *human face* or *city*. Task constraints have traditionally been the main focus of categorisation research, but they are an irreducible factor of *any* recognition task, and the first factor of the *diagnostic recognition* framework outlined here (Schyns, 1998; see also Hill, Schyns & Akamatsu, 1997; Schyns & Oliva, 1999). Recognition is successful resolution of task constraints on a given input.

The second factor of diagnostic recognition is the *a priori* structure of perceptual information available to construct hierarchically organised categories. We group objects into perceptual categories because they "look alike"–i.e., they share cues such as a similar silhouette or global shape, distinctive sets of parts similarly organised (e.g., *nose, mouth, eyes, ears, hair* and their *relationships*), typical surface properties (e.g., *smooth* vs. *discontinuous*, *symmetric* vs. *asymmetric*, and *textural*, *colour* and *illumination cues*), or biological motion. Generally speaking, not all image cues are equally available; there are perceptual limitations to their extraction from the image. The structure and perceptual availability of object information

[12] This section has been further developed in Gosselin, Archambault and Schyns (in press).

has traditionally concerned perceptually-oriented object recognition researchers. However, perceptual cues are an irreducible factor of any object categorisation, and the second factor of diagnostic recognition.

In the diagnostic recognition framework the two factors just discussed interact: When the information required to assign an object to a category matches with input information, a subset of object cues become particularly useful (i.e., diagnostic) for the task at hand. Diagnosticity is the first component of recognition performance. However, perceptual limitations on the extraction of diagnostic cues should also affect performance. Thus, diagnostic recognition frames explanations of performance as interactions between cue diagnosticity and cue availability. It is our view that the nature and the implications of these interactions have been largely neglected both in object recognition and in object categorisation research.

SLIP incorporates these two critical aspects of diagnostic recognition: task constraints and information availability. Task constraints correspond here to the different strategies associated with a vertical organisation of categories–i.e., the idea that different categorisation strategies can be applied to the same object. This enables us to examine whether the two determinants of SLIP (i.e., strategy length and internal practicability) determine the speed of access to the different levels of a taxonomy. We will provide several examples of this in Chapters 4 and 5. Furthermore, SLIP has two perceptual constraints: First, the features prescribed in a categorisation strategy are the only ones sampled in the input. This implies that changing the features of a strategy (e.g., via the acquisition of conceptual expertise) could, to some extent, control the features that are (or are not) seen in a given object. Archambault, O'Donnell and Schyns (1999) gave support for this: they showed that two groups of people with different strategies applied to the same object were relatively blind to changes that fell outside the feature tests prescribed by their strategies. Second, SLIP postulates that feature sets in a strategy are tested serially, in a specific order. Some of the implications of this last perceptual constraint will be explored in Experiment 7 (see Chapter 5).

Chapter 3. Other models of basic-levelness[13]

In this chapter we will review all formal models of basic-levelness before turning to a detailed comparison of their performance with that of SLIP. All of these formal models originate from two fundamental ideas: *utility* and *similarity*. The former led to the suggestion that the basic level is the most useful level of abstraction in a taxonomy. This was first exploited by Brown (1956) with his *level of usual utility* and by Rosch (Rosch & Mervis, 1975; Rosch et al., 1976; Rosch, 1978). Models that embrace utility are Rosch et al.'s (1976) *cue validity*, Jones's (1983) *category feature-possession*, Corter and Gluck's (1992) *category utility*, Fisher's (1986) *COBWEB*, Anderson's (1990) *rational analysis*, Pothos and Chater's (1998a) *compression*. The second class of models assumes that the basic level maximises a measure of exemplar similarity at this level of abstraction. These are Rosch's (1976) *differentiation model*, Tversky's (1977) *contrast model*, and Medin and Schaffer's (1978; modified by Estes, 1994) *context model*.

For each model, we will discuss whether or not it can predict the standard basic-level phenomenon, i.e. a preference for an intermediate level of abstraction. A successful example suffices to prove that a model possesses this potential; to show the contrary, a formal proof is required. The models that do not fulfil this requirement will not be included in our numerical simulations later on. To illustrate the functioning of the successful models, we will provide detailed numerical simulations of Murphy and Smith's (1982, Experiment 1) and Hoffmann and Ziessler's (1983, Hierarchy I) category structures (see Figure 1).

---

[13] This chapter expands on Gosselin and Schyns (1997, 1999).

We will mention two more explanations of basic-level performance: Murphy and Smith's (1982) *preparation model* which emerged neither from the utility, nor from the similarity tradition. It is not a fully worked-out model, but it is the only explicit attempt–before SLIP–to formalise participants' behaviour in verification tasks. Finally, we will review the most influential part-based accounts.

3.1 Utility and category cue validity

Brown (1956) suggested that "... things are first named so as to categorise them in a maximally useful way." (p. 20) For example, a dime is a "dime", instead of "metal object" because, for most purposes, this is what is relevant about the object. Rosch et al.'s (1976) *cue validity* builds on this idea. The cue validity of category c_i corresponds to the sum of the conditional probabilities that an object belongs to c_i given that it possesses each one of n features. Formally,

$$\sum_{j=1}^{n} P\left(c_i | f_j\right).$$

Cue validity can be framed as a measure of category utility because the more informative a cue is, the higher its cue validity is, and the more useful it is. Murphy (1982) proved that the cue validity of a more general category (e.g. *white whale*) is necessarily greater or equal to the cue validity of a more specific category (e.g. *Moby Dick*). Hence, cue validity cannot predict a superiority for an intermediate categorisation level, a minimal requirement of models of basic-levelness.

3.1.1 Category feature-possession

Jones (1983) proposed that the basic level maximises the average category feature-possession, another measure of usefulness. Like cue validity, it starts from the probability that an object belongs to category c_i given that it possesses feature f_j, $P\left(c_i | f_j\right)$. However, it also considers the probability that the object possesses feature f_j given that it belongs to category c_i, $P(f_j|c_i)$. Together, the conjunction of these two probabilistic events, $K_{ij} = P(c_i|f_j)P(f_j|c_i)$, is called the *collocation* of category c_i and feature f_j. Collocations are computed for all categories and features, and the largest collocations of each feature are extracted, $K_{ij} = \max(K_{1j}, K_{2j}, \ldots, K_{mj})$. For each category, the number of largest feature collocations are counted. This number weighted by a number between 0 and 1 (following Jones, 1983, we set this weight to 1) is the feature possession of a category. It reflects the number of strong bi-directional links between a category and its features, or their mutual predictability. The category feature-possession of a level of categorisation is the average of the category feature-possession of its categories.

Consider the taxonomies of Figure 1. First, we must compute $P(c_i|f_j)$'s and $P(f_j|c_i)$'s for $i, j \;\; \{f_1, f_2, \ldots, f_n\}$. For example, in the top taxonomy, $P(a|ril)$ and $P(ril|a)$ are equal to 1; and, in the bottom taxonomy, $P(d|hob)$ is equal to .5 and $P(hob|d)$ to 1. Second, we calculate all the collocations. In the top taxonomy, the collocation of category *ril* and feature *a* is $P(a|ril)P(ril|a)$, that is 1 (see Table 2 for all collocations of category feature-structures of the top taxonomy in Figure 1); and the collocation of category *hob* and feature *d* is $P(d|hob)P(hob|d)$, or .5 (see Table 3 for all collocations of category structure of the bottom taxonomy in Figure 1).

Table 2: Collocations for the numerical simulation of the category feature-possession (Jones, 1983) with the top taxonomy in Figure 1 (Hoffmann & Ziessler, 1983, Hierarchy I).

	Feature												
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	
ril	1	0	.5	.5	0	0	.125	.125	.125	.125	.25	.25	3
mip	0	1	0	0	.5	.5	.125	.125	.125	.125	.25	.25	3
kas	.5	0	1	0	0	0	.25	.25	0	0	.125	.125	1
jad	.5	0	0	1	0	0	.25	.25	0	0	.125	.125	1
gam	0	.5	0	0	1	0	0	0	.25	.25	.125	.125	1
saf	0	.5	0	0	0	1	0	0	.25	.25	.125	.125	1
lun	.25	0	.5	0	0	0	.5	0	0	0	.063	.063	1
fuk	.25	0	.5	0	0	0	0	.5	0	0	.063	.063	1
tuz	.25	0	0	.5	0	0	0	0	.5	0	.063	.063	1
zut	.25	0	0	.5	0	0	0	0	0	.5	.063	.063	1
nub	0	.25	0	0	.5	0	.5	0	0	0	.063	.063	1
duw	0	.25	0	0	.5	0	0	.5	0	0	.063	.063	1
pux	0	.25	0	0	0	.5	0	0	.5	0	.063	.063	1
bur	0	.25	0	0	0	.5	0	0	0	.5	.063	.063	1

Table 3: Collocations for the numerical simulation of the category feature-possession (Jones, 1983) with the bottom taxonomy in Figure 1 (Murphy & Smith, 1982, Experiment 1).

	Feature																
	a	*b*	*cde*	*fgh*	*ijk*	*lmn*	*o*	*p*	*q*	*r*	*s*	*t*	*u*	*v*	*w*	*x*	
hob	1	0	.5(*3)	.5(*3)	0	0	.25	.25	.25	.25	0	0	0	0	.25	.25	3
som	0	1	0	0	.5(*3)	.5(*3)	0	0	0	0	.25	.25	.25	.25	.25	.25	3
bot	.5	0	1(*3)	0	0	0	.5	.5	0	0	0	0	0	0	.125	.125	3
rel	.5	0	0	1(*3)	0	0	0	0	.5	.5	0	0	0	0	.125	.125	3
pim	0	.5	0	0	1(*3)	0	0	0	0	0	.5	.5	0	0	.125	.125	3
nop	0	.5	0	0	0	1(*3)	0	0	0	0	0	0	.5	.5	.125	.125	3
com	.25	0	.5(*3)	0	0	0	1	0	0	0	0	0	0	0	.063	.063	1
vad	.25	0	.5(*3)	0	0	0	0	1	0	0	0	0	0	0	.063	.063	1
lar	.25	0	0	.5(*3)	0	0	0	0	1	0	0	0	0	0	.063	.063	1
zim	.25	0	0	.5(*3)	0	0	0	0	0	1	0	0	0	0	.063	.063	1
wam	0	.25	0	0	.5(*3)	0	0	0	0	0	1	0	0	0	.063	.063	1
tis	0	.25	0	0	.5(*3)	0	0	0	0	0	0	1	0	0	.063	.063	1
mul	0	.25	0	0	0	.5(*3)	0	0	0	0	0	0	1	0	.063	.063	1

Third, we locate the largest collocation for every feature in the columns of Tables 2 and 3 (see the shaded figures of Tables 2 and 3). For example, the largest collocation for feature *a* in Table 2 is equal to 1 and the largest one for feature *d* in Table 3 is also equal to 1. Fourth, a count of the number of shaded figures provides the category feature-possession measures (see the rightmost column of Tables 2 and 3). Both the *ril* and the *hob* feature-possession scores are 3. Finally, these category feature-possessions are averaged within level of categorisation. For the top taxonomy of Figure 1, feature-possession predicts that reaction times should be fastest at the high level of categorisation (high-level feature-possession = 3), and that they should be equally slow at the middle- and low- levels (middle- and low- levels feature-possession = 1). And for the bottom category taxonomy of Figure 1, feature-possession predicts that reaction times (RT) should be fastest at the middle- and high- levels of categorisation (high- and mid- level feature-possession = 3), and that they should be slowest at the low-level (low-level feature-possession = 1). Feature-possession accounts for the observed mid-level preference in Murphy & Smith (1982, Experiment 1).

When category organisations are entirely composed of non-overlapping features and no nondiagnostic features are present, the category feature-possession measure is equal to the number of added features for this category (we will see that, in this case, SLIP's predictions are proportional). Proof: collocation is equal to 1–the maximum–for such categories and feature pairs; otherwise it is smaller than 1 (either the feature is added above the category and $P(c_i|f_j)$ is smaller than 1, or the feature is added below the category and $P(f_j|c_i)$ is smaller than 1). Each feature is added at one category (by definition of the considered hierarchies). Thus each feature will be associated with one and only one collocation of 1. It follows that, here, category feature-possession is equal to the number of unique features added to this category. The number of unique added features is a measure of redundancy. For non diagnostic features (the ones with equal probability of occurring in all low-level categories), the highest level always wins. This is simply because $P(c_i|f_j)$ (which is equal to $P(c_i)$ here) is always maximum at the highest level and because $P(f_j|c_i)$ is constant for non diagnostic features. It is more difficult to grasp what happens when strategy length varies. Judging by our above simulation of Hoffmann and Ziessler (1983, Hierarchy I) as well as by the other simulations with varying strategy length presented in this dissertation (see Chapter 5), category feature-possession seems to be biased for high-level categories.

3.1.2 Corter and Gluck's category utility measure

Corter and Gluck's (1992) category utility is grounded on strong logical principles. For the authors, a useful category is more capable of predicting the features of its members. Starting from $P(f_j|c_i)$, the probability of feature f_j given category c_i, the probability of guessing correctly this feature is $P(f_j|c_i)^2$. If the category is useful, the informed feature guess should be better than a guess made without knowledge of the category. If $P(f_j)$ is the probability of such a raw feature guess, the probability of being correct is $P(f_j)^2$.

The category utility of c_i for feature f_j is $P(c_i)\left[P(f_j|c_i)^2 - P(f_j)^2\right]$, the subtraction between the informed and the raw feature guesses, given $P(c_i)$, the probability that the object belongs to c_i. Summed across all features of the input, category utility becomes

$$P(c_i)\sum_{j=1}^{m}\left[P(f_j|c_i)^2 - P(f_j)^2\right]. \quad (6)$$

Equation 6 computes the basic-levelness of category c_i. The basic-levelness of a level of abstraction is the average basic-levelness of its categories.

Let us illustrate this with the taxonomies from Figure 1. We start with the computation of all $P(c)$'s, $P(f_k)$'s, and $P(f_k|c)$'s. In both taxonomies in Figure 1, the $P(c)$'s are equal to .5, .25, and .125, respectively, for the higher, middle, and lower levels of categorisation. In the top taxonomy in Figure 1, the $P(f_k)$'s are equal to .5 for *a,b,k, and l*, and to .25 for the other features. For the bottom taxonomy of Figure 1, the $P(f_k)$'s are equal to .5, .25, and .125, respectively, for features *a* and *b*, *c* to *n*, and *o* to *x*. Here are a few $P(f_k|c)$'s samples: *P(a | ril)* is 1, *P(b | ril)* is 0, and *P(g | ril)* is .25; and *P(a | hob)* is 1, *P(b | hob)* is 0, and *P(d | hob)* is .5. Next, we subtract the squares of the $P(f_k|c)$ from the squares of the $P(f_k)$'s. For instance, [$P(g\,|\,ril)^2$-$P(g)^2$] is equal to 0, and [$P(d\,|\,hob)^2$-$P(d)^2$] to .188 (Table 4 summarises all these differences for the high-level category *ril*, the middle-level category *kas*, and the low-level category *lun* from the top taxonomy of Figure 1; and Table 5 summarises them for the high-level category *hob*, the middle-level category *bot*, and the low-level category *com* from the bottom taxonomy of Figure 1).

Table 4: Key computations for the numerical simulation of the category utility measure (Corter & Gluck, 1992) with the top taxonomy of Figure 1 (Hoffmann & Ziessler, 1983, Hierarchy I).

	Feature												
	a	*b*	*c*	*D*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	
ril	.75	-.25	.188	.188	-.063	-.063	0	0	0	0	0	0	.75
kas	.75	-.25	.938	-.063	-.063	-.063	.188	.188	-.063	-.063	0	0	1.5
lun	.75	-.25	.938	-.063	-.063	-.063	.938	-.063	-.063	-.063	0	0	2

Table 5: Key computations for the numerical simulation of the category utility measure (Corter & Gluck, 1992) with the bottom taxonomy of Figure 1 (Murphy & Smith, 1982, Experiment 1).

	Feature																
	a	*b*	*cde*	*fgh*	*ijk*	*lmn*	*o*	*p*	*q*	*r*	*s*	*t*	*u*	*v*	*w*	*x*	
hob	.75	-.25	.188	.188	-.063	-.063	.047	.047	.047	.047	-.016	-.016	-.016	-.016	0	0	1.374
bot	.75	-.25	.938	-.063	-.063	-.063	.234	.234	-.016	-.016	-.016	-.016	-.016	-.016	0	0	3.119

Then, we sum all the differences within categories (i.e., the rows in Tables 4 and 5). The totals appear in the rightmost column of Tables 4 and 5. Fourth, we obtain the category utility scores by weighting each sum by the appropriate $P(c)$. For the top taxonomy of Figure 1, the *ril* category utility is 0.375 (.5 * .75), that of *kas*, 0.375 (.25 * 1.5), and that of *lun*, 0.25 (.125 * 2). For the bottom taxonomy of Figure 1, the *hob* category utility is 0.687 (1.374 * .5), that of *bot* is 0.780 (3.119 * .25), and that of *som* is 0.452 (3.619 *.125). Finally, the average of all category utilities is computed within a level of abstraction. For the top taxonomy of Figure 1, the high-and mid- levels category utility is 0.375, and the low-level category utility is 0.25. This is exactly what Hoffmann and Ziessler (1983, Hierarchy I) observed. For the bottom taxonomy of Figure 1, the middle-level utility is the greatest which is 0.780, followed by the high-level utility which is 0.687, and trailed by the low-level utility which is 0.452. Murphy and Smith (1982, Experiment 1) observed a slightly different pattern of response times: the high-level categories RTs trailed the pack and the low-level categories were verified the second fastest. Nevertheless, this last example demonstrates that the utility of a level of abstraction can be maximum at an intermediate level of categorisation.

Category utility also has a bias for higher levels of categorisation. The following equation

$$P(c) \sum_{k=1}^{m} P(f_k|c)^2 \quad \sum_{k=1}^{m} P(f_k)^2 \; ($$

was obtained by distributing the summation of Equation 6 over the two terms of the subtraction. $\sum_{j=1}^{m} P(f_j)^2$ is a constant and the two remaining variable terms are biased for higher levels. The probability $P(c_i)$ that an object belongs to category c_i decreases exponentially with increasing levels of category specificity, quickly reducing utility at each level (e.g. in the top taxonomy of Figure 1, starting from the top level, utility is halved at each level down the hierarchy). At the same time, $\sum_{j=1}^{m} P(f_j|c_i)^2$ usually increases almost linearly with increasing specificity, and can only compensate the exponential reduction of $P(c_i)$ with an exponential addition of redundant features at lower levels. Hence the bias of category utility for the higher levels of a taxonomy.

The addition of features at any level of the hierarchy increases utility, but as just discussed, exponential additions at lower levels are necessary to compensate for the exponential decrease in the likelihood of these categories. When the added features are unique to the category their number is proportional to redundancy, and thus category utility is sensitive to redundancy. However, when added features overlap between categories category utility and SLIP tend to diverge in their predictions.

3.1.2.1 Fisher's COBWEB

Fisher's (1987, 1988) COBWEB is an incremental clustering algorithm. Clustering consists in placing items into contrasting categories based on some entirely unsupervised rules; it is thus different from other categorisation tasks in which feedback informs people on whether or not an object belongs to a category. COBWEB's clustering criteria is Corter and Gluck's (1992) category utility. When an object is encountered, COBWEB places it either into an already existing category, or into a new singleton category. To be exact, COBWEB considers all possible categorisations for this object, including the one putting it in a new singleton category; for each categorisation, COBWEB computes the average category utility:

$$\frac{\sum_k P(C_k)\left[\sum_i \sum_j P\left(f_{ij}|k\right)^2 - \sum_i \sum_j P\left(f_{ij}\right)^2\right]}{n},$$

where $P\left(f_{ij}|k\right)$ is the probability of attribute i possessing value j in category k, $P\left(f_{ij}\right)$ is the base rate of attribute i possessing value j, and n is the number of categories in the partition. The categorisation with the greatest average category utility score is the selected one. Fisher proposed that COBWEB partitions objects into their basic categories.

COBWEB's clustering criteria is Corter and Gluck's (1992) category utility summed across categories and divided by the number of categories, and it thus predicts more or less the same thing.

3.1.3 Rational analysis

Anderson (1990, 1991) has devised an iterative algorithm of categorisation. It inputs the dimensional descriptions of objects one at a time, and outputs the "optimal" probability that a new object will display a certain value on a dimension given a set of dimensional values for that object. For example, it could inform an animal whether or not a novel object has a positive value on the "dangerous" dimension. As a by-product, this algorithm partitions objects into "horizontal" categories (i.e., this partitioning does not possess a "vertical" or hierarchical dimension). Anderson proposed that the most robust partitioning occurs at the abstraction level with highest basic-levelness. The algorithm is optimal inasmuch as it realises a version of Rosch's (1978) *cognitive economy* principle which states that "what one wishes to gain from one's categories is a great deal of information about the environment while conserving finite resources as much as possible" (p. 28). We will not explain its derivation because it is quite complicated. The interested reader is referred to Anderson (1990, 1991).

We will follow Anderson's 1991 formulation of his iterative categorisation algorithm with the exception of Equation 7 taken from his 1990 book.

The first object is categorised into a new singleton category. Then every time a new object is encountered a three-step process is repeated: (a) given a partitioning for the first m objects, calculate for each category k the probability $P(k|F)$ that m+1st object comes from category k (this includes the existing categories and a new category) given that the object has features F; (b) create a partitioning of the m+1 objects with the m+1th object assigned to the category with maximum probability; and (c) to predict value j on an unobserved dimension i for the n+1st object with observed features F, calculate

$$P_i(j|F) = \sum_k P(k|F) P_i(j|k),$$

where $P(k|F)$ is the probability that the n+1st object comes from category k; and $P_i(j|k)$ is the probability of displaying value j on dimension i.

Only steps (a) and (b) concern us; we are solely interested in the partitioning of objects. Therefore, we only need to compute $P(k|F)$s. In Bayesian analysis, $P(k|F)$ is a posterior probability that the object belongs to k given that it has features F. It can be expressed as follows,

$$P(k|F) = \frac{P(k)P(F|k)}{\sum_k P(k)P(F|k)},$$

where the summation in the denominator is over all categories k currently in the partitioning, including the potential new one.

Anderson derived the prior probability, $P(k)$, by making the assumption that there is a fixed *coupling probability* (c) that two objects come from the same category, and that this probability is independent of the number of objets categorised so far:

$$P(k) = \frac{cn_k}{(1-c)+cn},$$

where n_k is the number of objects assigned to category k so far; and n is the total number of objects seen so far. $P(0)$, the probability that the object comes from a entirely new category, is

$$P(0) = \frac{(1-c)}{(1-c)+cn}.$$

The coupling probability is crucial for the finding of the basic level. With large c, only one category will be created. As it gets smaller and smaller, more and more categories will be created. Anderson proposes–in a very ad hoc fashion–that, at c = .3, only the groupings associated with the greatest basic-levelness emerge.

To derive the conditional probability or likelihood term, $P(F|k)$, Anderson made the additional assumption that objects' dimensions are independent:

$$P(F|k) = \underset{i}{\ } P_i(j|k),$$

where the values j on dimensions i constitute the feature set F; and where $P_i(j|k)$ is the probability that an object from category k displays value j on dimension i. $P_i(j|k)$ is given by

$$P_i(j|k) = \frac{c_j + 1}{n_k + m}, \qquad (7)$$

where n_k is the number of objects in category k that have a value on dimension i; c_j is the number of objects in category k with the same value as the object to be classified; and m is the number of dimensions to be classified (Anderson, 1989).

Anderson has modelled Murphy and Smith's (1982, Experiment 1) category structure, with Corter and Gluck's coding (see Figure 17 in Chapter 4), but added another dimension–a label dimension–per level of abstraction. For $c > .96$, all items are put in one category; for $.8 > c > .95$, only the high-level clustering emerges; for $.4 > c > .8$, the model fluctuates between the high- and middle- level clusterings; for $.4 > c > .2$, it extracted only the middle-level clustering; for $.2 > c > .05$, it usually grouped the items in the intermediate clusters, and sometimes in singleton categories; and, for $c < .05$, the rational algorithm only extracted singleton sets. Hoffmann and Ziessler's experiments (including Hoffmann & Ziessler, 1983, Hierarchy I) were also successfully modelled.

We will not include rational analysis in our numerical simulations because it is not a basic-levelness metric per se; in the best of worlds, it only identifies the level with highest basic-levelness.

3.1.4 MDL and compression[14]

Minimum Description Length (MDL) is a method that uses partitioning of data to compress them (Chater & Pothos, 1999; Pothos & Chater, 1998, 1999). The amount of compression that a particular partitioning achieves is the difference in bits between a raw and a compressed description of the same data set. Different levels of a taxonomy correspond to different partitionings of the same data set. In Pothos and Chater (1998, 1999), the suggestion is made that the maximal compression of the data set is accomplished at the basic-level.

With r objects in a data set, there are $s = \frac{r(r-1)}{2}$ possible pairwise similarities between the objects. There are $A = \frac{s(s-1)}{2}$ possible binary relationships (inequalities) between the similarities. The representation of these inequalities requires A bits of information (one bit per similarity relationship). A bits of information describe the data set before partitioning. If D_i bits of information describe one partition of the same set, then A - D_i, measures the compression efficiency of this partition. The basic level should maximise this difference if this level achieves the maximal compression. Compression is therefore a measure of the utility of the basic level.

We now derive D_i , the encoding of a partitioning. The number of possible partitions of r items into n clusters, *Part(r,n)*, is

$$\sum_{v=0}^{n} (-1)^v \frac{(n-v)^r}{(n-v)!v!}$$

[14] Pothos and Chater (1999) present their model as a version of Rosch et al.'s (1976) differentiation model (see *3.2 Similarity and the differentiation model*), but we believe it bears more commonalties with the utility tradition than with the similarity one.

To encode this partitioning, we need $\log_2[Part(r,n)]$ bits of information. We then compute *u*, the combinatorics of all within-cluster similarities with all between-cluster similarities. The scheme assumes that within-cluster pairwise similarities *s*(*i*, *j*) are all greater than any between-cluster pairwise similarities, *s*(*k*, *l*). However, this does not always hold and *e* counts the number of times the assumption is violated (*e* can vary between 0 and *u* and be encoded on a maximum of $\log_2(u+1)$ bits). There are $C_e^u = \frac{u!}{(u-e)!e!}$ possible ways of selecting *e* errors among the combinatorics of relationships *u*. A total of $\log_2(u+1) + \log_2(C_e^u)$ bits encode the errors.

Remember that *A* bits specify all possible binary relationships between pairwise similarities, whereas *u* specifies those constrained by the clustering. *A* - *u* counts the relationships left outside the clustering. *A* - *u* bits encode them.

The compression of information offered by one partitioning of the data is *A* - D_i, where D_i = $\log_2[Part(r,n)]$ + $[\log_2(u+1) + \log_2(C_e^u)]$ + (*A* - *u*.). Hence *A* - D_i is equal to

$$u - \{\log_2[Part(r,n)] + \log_2(u+1) + \log_2(C_e^u)\}.$$

We will illustrate this with the taxonomies in Figure 1. First, the parameters *u* and *e* must be calculated (*r* is simply the number of objects). Note that the category tree is the same in both cases: two high-level categories, each divided into two mid-level categories, each divided into two low-level categories, each containing two exemplars. It is the feature definitions of the categories of these category structures that differ; this will play a role in phase two of the computations. *r* = 16 (2 exemplars * 8 low-level categories). How many binary inequalities are constrained by the assumption that all within cluster pairs of similarities are greater than all between cluster ones at the high-, middle- and low-levels of categorisation?

At the high-level, we have two clusters with eight objects each. We have a total of 56 within-cluster pairwise similarities (28 pairs per cluster * 2 clusters), and 64 between-cluster pairwise similarities (8 objects in the first cluster * 8 objects in the second). Thus, we have $u = 3584$ (56 within-cluster pairs * 64 between-cluster pairs). At the middle-level, we have four clusters each containing four objects. A total of 24 within-cluster pairs (6 pairs per cluster * 4 clusters), and 96 between-cluster pairwise similarities (4 objects in a first cluster * 4 objects in a second cluster * 6 permutations). And $u = 2304$ (24 within-cluster pairs * 96 between-cluster pairs) for the middle-level clustering. At the low-level, we have eight clusters with two objects each. This implies 8 (1 pair per cluster * 8 clusters) within-cluster pairs and 112 (2 items in a first cluster * 2 items in a second cluster * 28 permutations) between-cluster pairwise similarities, for a grand total of $u = 896$ (8 within-cluster pairs * 112 between-cluster pairs).

The tree branching cost term, $\log_2[Part(r,n)]$, is equal to about 15, 27.36, and 31 bits, at the high-, middle-, and low- levels of abstraction, respectively. Second, this description must be corrected for similarity errors. The taxonomies of Figure 1 must be treated separately from now on. How many of these constraints are false in the top and bottom taxonomies of Figures 1? Remember that Pothos and Chater (1998, 1999) assumed that one similarity can be either greater or smaller than–but that it cannot be equal to–another. This is false of many pairwise similarities here. For example, in the bottom taxonomy of Figure 1, the similarity (defined à la Tversky–see section *3.2.1 Contrast model*) between *acdeox* (*small com*) and *afghrw* (*large lar*) is the same as that between the *acdeox* (*small com*) and the *bijksx* (*small wam*). Following Pothos (1999), we will only count $S_{\text{within}} > S_{\text{between}}$ as an error, but not $S_{\text{within}} = S_{\text{between}}$ which will be treated as $S_{\text{within}} < S_{\text{between}}$. Therefore not a single constraint derived from the categorisation-level partitionings in the bottom taxonomy of Figure 1 description is incorrect. The worst high-level within similarity is *S*(*acdeox*, *afghrw*) (share 1 feature, contrast on 5) is equal to all high-level between similarities; the worst middle-level within similarities share 4 features and differ on 2, e.g. *S*(*acdeox*, *acdepw*), which is better than all middle-level between similarities; and the worst low-level similarity is *S*(*acdeox*, *acdeow*) (share 5 features, contrast on 1) is greater than all low-level between similarities. So compression is equal to 3569 bits at high-level, 2273.64 bits at the mid-level, and 865 bits at the low-level. The middle level advantage for Murphy and Smith (1982, Experiment 1) is not predicted. However, Pothos and Chater (1998, 1999) have modelled the same experiment with Corter and Gluck's coding, and predicted a middle level advantage (see section *4.5 A cautionary note about coding*).

Given the top taxonomy of Figure 1, however, mistakes are made with the simple category tree description. Trouble arises at the high-level: The worst high-level within pairwise similarity is, for example, *S*(*acgk*, *acil*) (1 shared feature; 3 contrasting features), and the best high-level between similarity is *S*(*acgk*, *begk*) (2 shared features; 2 contrasting features). No error is made at the mid- and low- levels. The worst mid-level within similarity is *S*(*acgk*, *achl*) (2 shared features; 2 contrast features) and the best mid-level between similarity is *S*(*acgk*, *adik*) (2 shared features; 2 contrast features). And the worst low-level within similarity is *S*(*acgk*, *acgl*) (3 shared features; 1 contrast feature); this is just as good as the best low-level between similarity, *S*(*acgk*, *achk*). The information required to correct all the high-level pairwise similarity mistakes is 809.68 bits. The net compression indexes are thus 2759.32, 2276.64, and 865.00, respectively, at the high-, mid-, and low- levels of abstraction. This is not too far from the pattern of response times observed by Hoffmann and Ziessler (1983, Hierarchy I).

When there is little feature overlap, MDL is essentially dependent on u, the combinatorics of within and between category similarities (or the category tree structure–see *Preamble*). This combinatorics grows with level of generality and so compression is greater when fewer categories are considered, irrespective of how redundant the features are within the categories. Compression thus has a bias for high levels of abstraction. Variations of strategy length create overlap between features and adds errors to the MDL description. However, these tend to be insufficient to counterbalance the bias for the higher levels.

3.2 Similarity and the differentiation model

So far, the models reviewed implemented the principle that the basic level is the most useful level of a taxonomy. Another principle for formal models is that of *differentiation,* or *dissimilarity*. As Rosch et al. (1976) put it, categories at the basic level "... have the most attributes common to members of the category and the least attributes shared with members of other [contrasting] categories." (p. 435) This is family resemblance applied vertically, or to embedded categories, rather than horizontally, or to contrasting categories (Rosch, 1978). The horizontal category with the greatest family resemblance index is called the *prototype* (Rosch and Mervis, 1975), and the vertical category with the greatest family resemblance index is called *basic* (Rosch et al., 1976). The first component of this family resemblance definition has been called the *specificity* (Murphy & Brownell, 1985), or the *informativeness* (Murphy, 1991a) of a category, and the second component the *distinctiveness* of a category (Murphy & Brownell, 1985; Murphy, 1991a)[15]. However,

[15] The two determinants of SLIP can loosely be mapped onto those of the differentiation model. Both strategy length and informativeness have a tendency to increase with specificity; both internal practicability and distinctiveness have a tendency to

category differentiation is not sufficiently specified to be refuted. For example, there is a polynomial on distinctiveness (or on informativeness) of degree n that fits perfectly an arbitrary RT pattern on n+1 levels of categorisation. A similar point was made by Medin (1983).

3.2.1 Contrast model

Tversky's (1977) *contrast model* is a measure of similarity between pairs of exemplars. It is a particular case of the *matching function*. The matching function states that the similarity between two exemplars is a function of the number of their shared and distinctive features. Formally the contrast model can be expressed as follow:

$$S(a,b) = \ f(A \quad B) \quad f(A \quad B) \quad f(B \quad A),$$

where a and b are two exemplars; $S(a,b)$ is the similarity of a to b; A and B are the sets of features of objects a and b, respectively; and , , and are three constants between 0 and 1. Providing that $f(A \quad B) \quad f(B \quad A)$ and that , the contrast model predicts asymmetrical similarities that is, $S(a,b) \quad S(b,a)$, as, for example, North Korea is more similar to Red China than Red China is to North Korea.

The contrast model needs to be specified further if it is to be used at all for modelling. In the literature, applications of the contrast model are limited to the special case in which $f(X)$ denotes the number of features in X (Estes, 1994). We will call this the *special contrast model*.

Tversky applied this model to a number of problems among which was the determination of the basic-level of categorisation. The basic-levelness, or–as Tversky put it himself–the category resemblance, of category L is the mean of the similarities of all the pairs of distinct exemplars of L. Formally,

---

increase with generality. It can thus argued that SLIP is one instance of the differentiation model (Schyns, 1998).

$$R(\mathrm{L}) = \frac{S(a,b)}{\binom{n}{2}},$$

with a and b distinct. $R(\mathrm{L})$ is the category resemblance of L, and n is the cardinality of L.

We run into a problem when the low-level is the identity-level that is, when low-level categories contain a single exemplar. To deal with this problem, a straightforward variation of Tversky's measure–the average of the similarities of all possible pairs, including object X and itself–can be used. Formally,

$$R(\mathrm{L}) = \frac{S(a,b)}{n(n+1)/2}.$$

However, this special contrast model cannot have a mid-level advantage. As we get away from the top level of abstraction, the numerator increases (this is simply a feature of category hierarchies–categories get more and more similar as they become more and more specific) and the denominator decreases. Thus the most specific level of categorisation will always have the largest category resemblance score.

3.2.2 Context model

Estes (1994) has transformed Medin and Schaffer's (1978) influential context exemplar model of categorisation (see Nosofsky, 1986 and Lamberts, 1994, for developments) into a measure of basic-levelness. In this model, a multiplicative rule computes the similarity $S(a,b)$ between any two exemplars of a category. A match between the corresponding attributes of two objects is assigned value 1, and a mismatch is assigned ${}_D$, a dissimilarity parameter (${}_D$ varies between 0 and 1, corresponding to the saliency, or attentional weight of the considered attribute). To compute the similarity of any two exemplars, the local similarities of their component attributes are multiplied with one another.

A ratio provides an index of basic-levelness of an exemplar: The numerator sums the within-category similarity of one exemplar with all other exemplars of this target category–including itself. The denominator sums similarities extending the comparisons to the exemplars comprised in the related higher level category. The index of basic-levelness of a category is the average of such ratios, across all exemplars of the category. The basic-levelness of a level of categorisation is the average basic-levelness of all its categories.

We will illustrate this with the taxonomies from Figure 1. The bottom taxonomy is defined by a total of six dimensions. The measure of similarity between any pair of exemplars (i.e., *S(a,b)*) is thus a six-factor multiplication. Exemplars *acdeow* and *acdeox*, for instance, share values on five dimensions: *a*, *c*, *d*, *e*, and *o*; and they differ on a single dimension's value (*w* vs. *x*). Following Estes (1994), we will use a single mismatch parameter: . So, the six-factor multiplication defining the similarity between *acdeow* and *acdeox* is: 1 * 1 * 1 * 1 * 1 * = . Similarly, we obtain *S(acdeow, acdeow)* = 1, *S(acdeow, acdeox)* = *S(acdeow, acdepw)* = , *S(acdeow, acdepx)* = ^{2}, *S(acdeow, afghqw)* = *S(acdeow, afghrw)* = ^{4}, *S(acdeow, afghqx)* = *S(acdeow, afghrx)* = *S(acdeow, bijkw)* = *S(acdeow, bijktw)* = *S(acdeow, blmnuw)* = *S(acdeow, blmnvw)* = ^{5}, *S(acdeow, bijksx)* = *S(acdeow, bijktx)* = *S(acdeow, blmnux)* = *S(acdeow, blmnvx)* = ^{6}.

Applying the same multiplicative rule to the top taxonomy of Figure 1 we get *S(acgk, acgk)* = 1, *S(acgk, acgl)* = *S(acgk, achk)* = , *S(acgk, achl)* = *S(acgk, adik)* = *S(acgk, adjk)* = *S(acgk, begk)* = ^{2}, *S(acgk, adil)* = *S(acgk, adjl)* = *S(acgk, begl)* = *S(acgk, behk)* = *S(acgk, bfik)* = *S(acgk, bfjk)* = ^{3}, and *S(acgk, behl)* = *S(acgk, bfil)* = *S(acgk, bfjl)* = ^{4}.

The next step consists in computing the sum of the similarities of one exemplar of category X (any exemplar) to every other exemplar of this category. For high-, mid-, and low-level categories of the bottom taxonomy of Figure 1 these are $1+2\ \ +\ \ {}^{2}+2\ {}^{4}+2\ {}^{5}$, $1+2\ \ +\ \ {}^{2}$, and $1+\ \ $, respectively; and, for high-, mid-, and low-level categories of the top taxonomy of Figure 1 these are $1+2\ \ +3\ {}^{2}+2\ {}^{3}$, $1+2\ \ +\ \ {}^{2}$, and $1+\ \ $, respectively.

Finally, these summed similarities are divided by the summed similarities that extend the comparisons to the exemplars comprised in the related higher level category. For example: The probability that exemplar *acgk* in the top taxonomy of Figure 1 will be categorised as a *ril* is $\frac{1+2\ \ +3\ {}^{2}+2\ {}^{3}}{1+2\ \ +4\ {}^{2}+6\ {}^{3}+3\ {}^{4}}$, as a *kas*, $\frac{1+2\ \ +\ \ {}^{2}}{1+2\ \ +3\ {}^{2}+2\ {}^{3}}$, and as a *lun*, $\frac{1+\ \ }{1+2\ \ +\ \ {}^{2}}$. If, for instance, $\ \ =.3$ (this is the value used by Estes to model Corter, Gluck & Bower's, 1988, results), these probabilities become .896, .878, and .769, respectively. This is not too far from what Hoffmann and Ziessler (1983, Hierarchy I) observed.

And the probability that exemplar *acdeow* in the bottom taxonomy of Figure 1 will be categorised as *hob* is $\frac{1+2\ \ +\ \ {}^{2}+2\ {}^{4}+2\ {}^{5}}{1+2\ \ +\ \ {}^{2}+2\ {}^{4}+6\ {}^{5}+4\ {}^{6}}$, as *bot*, $\frac{1+2\ \ +\ \ {}^{2}}{1+2\ \ +\ \ {}^{2}+2\ {}^{4}+2\ {}^{5}}$, and as *com*, $\frac{1+\ \ }{1+2\ \ +\ \ {}^{2}}$. With $\ \ =.3$, these probabilities are .993, .988, and .769, respectively. Here, the model does not account for the observed middle-level preference. Estes showed, however, that it can account for the observed middle level preference in Corter, Gluck and Bower's taxonomy (1988; this taxonomy is isomorphic to Hoffmann & Ziessler's, 1983, Hierarchy II). The context model thus fulfils the basic requirement of possessing the capacity of predicting a middle-level preference.

The context model also has a bias for the higher-levels of a taxonomy. The measure is a ratio of two polynomials where the numerator differs from the denominator by only one term. With increasing levels of abstraction in a taxonomy, the numerator increases and the ratio approaches 1. The bias for the high level can be overcome by increasing the similarity between the target exemplars and those of lower level contrast categories.

3.3 Preparation model

Murphy and Smith (1982) proposed a different kind of model to explain basic-levelness that they called the *preparation model*. It is not a fully articulated formal model of basic-levelness. What is interesting about it is that it breaks with the utility and the similarity traditions. Moreover it is an embryo of a response-time model of basic-levelness (not entirely unlike SLIP) with a signal detection flavour. The model is schematised in Figure 6.

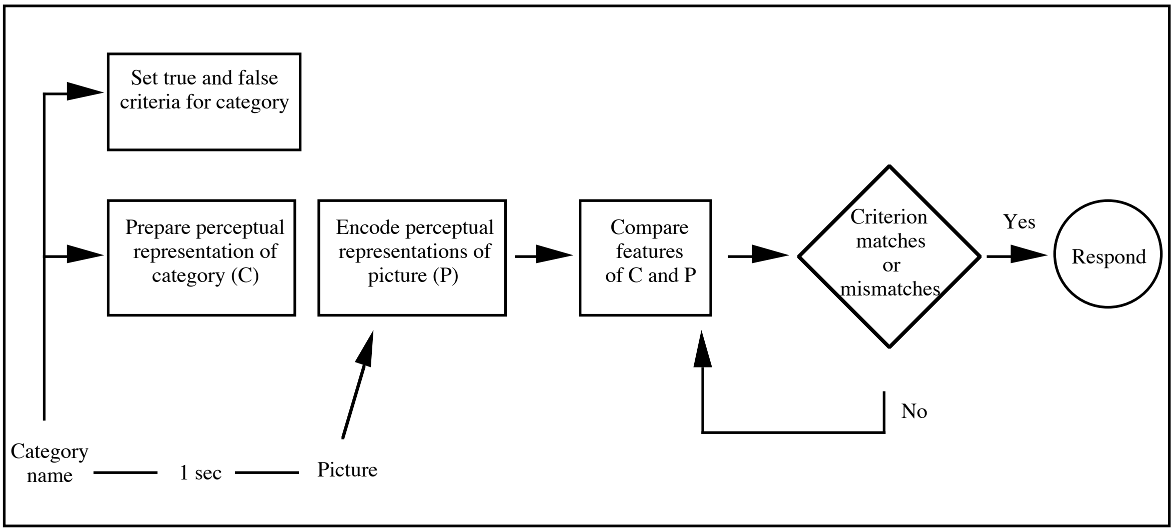


Figure 6. Box diagram of the preparation model. Adapted from Murphy & Smith (1982, Figure 4).

When the category is named, the participant gets ready for the forthcoming picture in two ways: (1) he activates a representation of the category; and (2) he sets a criteria on the number of matches and mismatches that will be needed to respond true and false, respectively. As soon as the picture appears, the perceiver starts to compare its features to those of the representation of the category, keeping track of matches and mismatches. When one of the pre-set criteria is reached, the participant responds either true, or false.

The superiority of the basic and subordinate categories over superordinate ones would result from people usually not having a single perceptual representation for a superordinate; two (or more) representations must thus be activated when the target category is a superordinate. The presence of these additional representations implies that additional matches will be needed on average. As a result RTs are lengthened compared to the case when the category is at a lower level, and a single perceptual representation is activated. This is in essence Jolicoeur, Gluck and Kosslyn's (1984) explanation. We address this explanation in the next section.

The advantage of the basic over the subordinate categories would occur because participants set different true and false criteria for categories at the two levels. Murphy and Smith assume that the criteria is set so as to maximise discrimination between the target category and contrasting ones. For example, take the taxonomy represented in the bottom of Figure 1. If the category presented was *rel* criteria are set so as to maximise discrimination from *bot*, since that is the closest "false" contrasting category. Because subordinates overlap more with their contrast categories than do basics, the true threshold should be set higher for subordinates; the greater the criterion, the longer the feature-comparison process, and therefore the longer the RT.

3.4 Part-based accounts

A commonality of all the reviewed models so far is that they are determined by category feature-structures. Provided that this abstract structure is unchanged, the content of these attributes is of no importance: it can be *shapes*, *colours*, *textures*, *shapes* and *colours*, and so on. To conclude this section on models of basic-level performance, we will examine the influential content-based or, more specifically, part-based accounts (e.g., Jolicoeur, Gluck & Kosslyn, 1984; Biederman, 1987).

We have already told part of the story (see section *1.3.3 Importance of shapes for basci-levelness*) but we believe it is worth repeating here. You will remember that Rosch et al. (1976) found a large and reliable increase in similarity of the overall look of objects from basic level to superordinate categories, and a significant–but significantly smaller–increase from basic to subordinate. They also found that averages of basic-level objects (i.e., some sort of primitive average interpolation between standardised objects belonging to the same basic-level categories) were the most inclusive average objects that were readily identifiable (see also Rosch, 1978). This suggests that shape is an important factor of basic-levelness for natural categories.

One determinant of shape is part structure. Tversky and Hemenway (1984) found–for a broad range of natural categories including both objects and organisms–a sharp increase of listed part-features from the superordinate to the basic level (e.g., *handle* and *blade* for "knife"; *peel* and *pulp* for "banana"), but little rise from the basic to the subordinate level. So it seems that parts–rather than their spatial relationships–are a crucial factor of basic-levelness.

In parallel to Tversky and Hemenway's research, Jolicoeur, Gluck, & Kosslyn (1984) took Rosch et al.'s (1976) claim one step further: not only are objects first recognised at their shape-based basic categories, *on average*, they are necessarily first recognised at their shape-based *entry point* categories. Entry point categories are usually at the basic-level but not always; for example, Jolicoeur, Gluck and Kosslyn showed that this was not the case for atypical objects (e.g., a penguin is accessed at the subordinate category "penguin" rather than the basic category "bird"). To access categories above the entry point, such as Rosch's superordinates, a search in the semantic tree is necessary after the entry point identification (note: this is equivalent to saying that a superordinate category is defined by the disjunction of entry point categories), and to access categories below the entry point, such as Rosch's subordinates, additional perceptual information is required. Thus the entry point is the point of contact between perception and semantic memory.

Both Jolicoeur, Gluck and Kosslyn's (1984) theory and Tversky and Hemenway's (1984) empirical findings have been extremely influential among the object recognition community (e.g., Neisser, 1987; Biederman, 1987; Biederman and Gerhardstein, 1993; Edelman, 1998). But, perhaps, it is in Biederman's hands that is best known.

Biederman developed a theory of object recognition called *recognition by components* (Biederman, 1987). This theory supposes that all objects can be decomposed into under 50 geometrical primitives or *geons* (Biederman, 1998). Geons are defined by visual properties that are invariant through most views and most orientations such as parallel lines and surface intersections. They include shapes such as cube, sphere, cylinder, etc. Once an object has been decomposed into geons, and once the structure of these geons (e.g., a typical house = a wedge ON TOP OF a cube) has been described, object identification ensues *at the basic level*. For Biederman, objects are first identified at the basic level and then they are identified at other levels. The superiority of basic and subordinate categories would come from the fact that no unique geon-structure representation exists for superordinate categories. The superiority of basic categories over subordinate categories results from the fact that the latter–but not the former–"[...] have a high degree of overlap in their components and in the relations among these components [...]" (Biederman, 1987, p. 143), and it takes more time to distinguish them. In other words, objects are recognised more rapidly at the basic level than at the subordinate level because the geon-based basic level categories are more *dissimilar* to one another than the geon-based subordinate level categories.

McMullen and Jolicoeur (1992) suggested that this "additional processing" required at the subordinate level could include determining spatial relations between geons. And that objects could be categorised at the basic level following only geon identification. This was latter corroborated by empirical findings (Hamm & McMullen, 1998).

To sum up: Biederman made two separate claims: (1) basic (or entry point) categories are necessary halts before accessing related higher or lower level categories, and (2) these basic categories are necessarily defined by parts. We believe that both are incorrect.

It has been shown several times that you can get a basic level effect in all kinds of taxonomies without parts (Corter, Gluck & Bower, 1988). Furthermore, Murphy (1991a) has shown that in a taxonomy with parts, you can have a superiority effect at a level of abstraction not defined by parts (see *1.3.3 Importance of shape for basic-levelness*).

As for Biederman's other claim, Thorpe, Fize, and Marlot (1996; see also Thorpe, Gegenfurtner, Fabre-Thorpe & Bülthoff, 1999), for example, have shown that people can decide whether new natural scenes presented for a very short presentation time (28 ms) contains an animal or not (this is a superordinate categorisation). In fact, people do this very efficiently; they are 93% accurate. More importantly for our argument, people could not identify the basic name of the animal in almost all cases. This demonstrates that basic-level identification is not necessary for superordinate recognition of natural objects. It does not imply that the superordinate category animal has a greater basic-levelness than its basic subdivisions. For the superordinates the participants had only two known choices whereas for the basic-level categories their responses were unrestricted, and thus much more uncertain or difficult.

3.5 Summary

We will not compare the performance of all the reviewed models in the following chapters. Instead we will concentrate on Jones's (1983) category feature-possession, Corter and Gluck's (1992) category utility, Pothos and Chater's (1998, 1999) compression measure, and Medin and Schaffer's (1978; modified by Estes, 1994) context model.

We have left aside Rosch et al.'s (1976) cue validity and Tversky's (1977) contrast model because they cannot predict the classic advantage for an intermediate level. Fisher's (1986) COBWEB measure has been discarded because it is based on Corter and Gluck's category utility, and makes roughly the same predictions. We have also excluded Anderson's (1989, 1990) rational analysis model because it does not provide a metric of basic-levelness. Finally, Murphy and Smith's (1982) preparation model as well as Biederman's (1987) recognition-by-component theory are too crude to allow comparisons with the other basic-level models.

Chapter 4. Numerical simulations of published experiments[16]

In this chapter, we will compare the predictive performance of SLIP with those of Jones's (1983) category feature-possession, Corter and Gluck's (1992) category utility, Pothos and Chater's (1998) compression measure and Medin and Schaffer's (1978) context model. The database comprises 21 (all, as far as we know) basic-level experiments. We will first present the experiments before turning to their simulations with the models. In the following sections, the experiments are organised according to the factor of SLIP they test–strategy length or internal practicability. For variations of internal practicability, we will examine the experiments that found faster access at an intermediate, low, and high level of categorisation. We will then describe the only two experiments that explicitly tested strategy length, the second computational factor of SLIP. Finally, we will examine taxonomies with disjunctions of attributes, or mixtures of conjunctions and disjunctions.

4.1 Variations of internal practicability determines basic-levelness

4.1.1 Faster access at an intermediate level

One of the most influential experiments on the basic-level is that of Murphy and Smith (1982, Experiment 1). It is influential because most subsequent experiments on the basic level used the same procedure. Their participants were initially taught the artificial taxonomy represented at the top of Figure 7. (We have normalised the notation of information in taxonomies, substituting letters of the alphabet for the actual features.

[16] This chapter expands on Gosselin and Schyns (1997, 1999).

This enables direct comparisons of the taxonomies; the mapping between the letters and what they signify is provided in the figures. Underneath the category names, we give the abstract optimal strategies fed to SLIP in the shorthand notation described earlier. At the bottom of each taxonomy, the abstract feature constitution of all exemplars is given.)

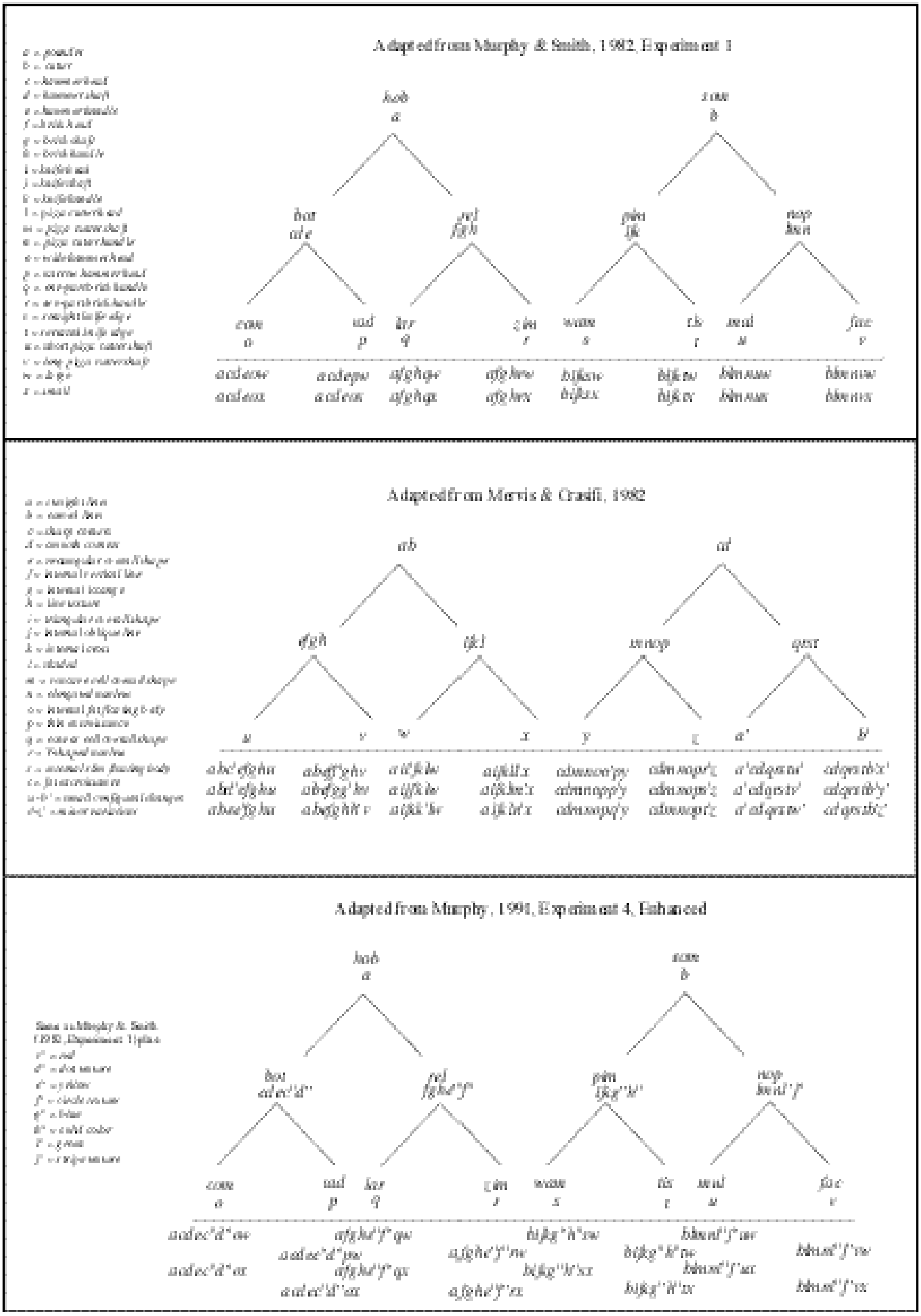


Figure 7. Taxonomies for all experiments with varying redundancy that exhibited an advantage for an intermediate level of categorisation. From top to bottom: Murphy & Smith (1982, Experiment 1–see also Murphy, 1991, Experiment 4, Simple, for a replication); Murphy & Crasifi (1982); Murphy (1991, Experiment 4, Enhanced). Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy. An index for these abstract features is also provided.

Murphy and Smith used 16 artificial tools. Four of these artificial tools are shown in Figure 8. Their tools were either *pounders*, or *cutters* (higher level); they had non-overlapping *handles*, *shafts*, and *heads* (which defined the middle level); and they had one more non-overlapping feature such as *narrow* and *wide* heads (at the lower level). Each low-level category contained a *large* and a *small* tool exemplar (*size* is thus a nondiagnostic dimension). In a later testing phase, participants were shown a category name followed by a stimulus. Subjects' task was to verify as quickly as possible whether the name and stimulus matched.

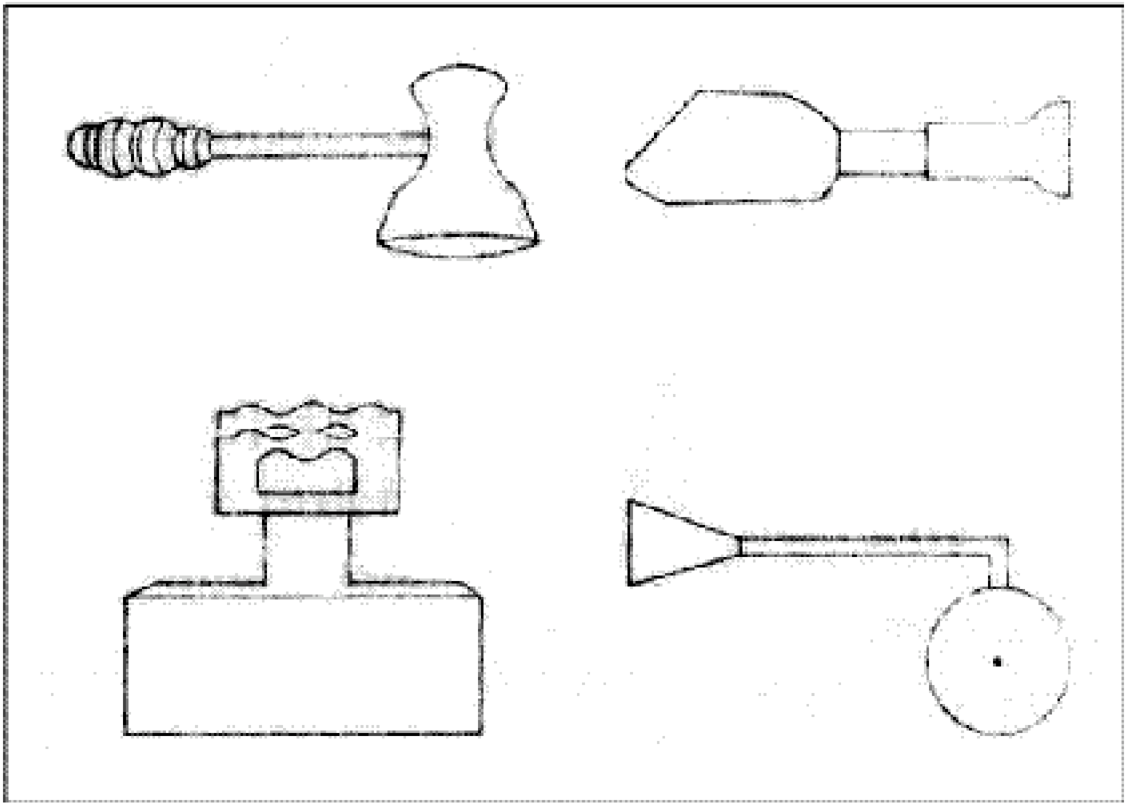


Figure 8. Four artificial tools used by Murphy & Smith in their Experiment 1. Scanned from Murphy & Smith (1982, Figure 1).

As illustrated in Figure 7, mid-level categories have the highest practicability. Table 6a shows that they were verified faster, and the high-level categories slowest. Murphy (1991a, Experiment 4, Simple) replicated these results.

In fact, the highest practicability of the middle level is also responsible for its faster access in Mervis and Crasifi (1982) and Murphy (1991a, Experiment 4, Enhanced).

Mervis and Crisafi (1982) used 24 abstract artificial objects similar to the ones shown in Figure 9.

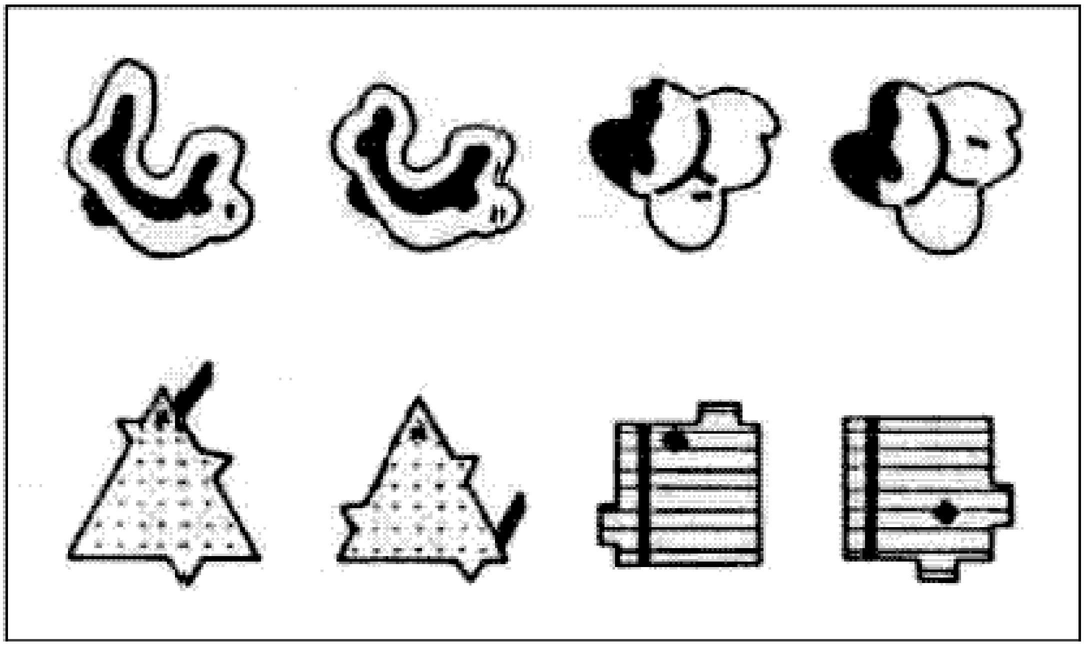


Figure 9. Eight sample objects used by Mervis & Crisafi. Scanned from Mervis & Crisafi (1982, Figure 1).

Each low-level category contained three highly similar exemplars. The superordinate categories were defined by a value on two redundant dimensions: *curvature* (either *straight* or *curved*) and *angularity* (either *sharp* or *smooth corners*). The basic categories were defined by a set of redundant values on four dimensions: the *overall shape* (*triangle*, *square*, *fat cell*, and *slim cell*) and three additional internal characteristics (e.g., *line texture*, *black stripe*, and *diamond*). The subordinate categories were defined by one *configural change* dimension. This taxonomy is the middle one in Figure 7.

In Murphy (1991, Experiment 4, Enhanced), two dimensions (i.e., *colour* and *texture*: either *red dots*, *yellow circles*, *green stripes*, or *blue solid colour*) were added to the artificial tools of Murphy and Smith (1982, Experiment 1) at the middle level of categorisation. This category hierarchy is illustrated at the bottom of Figure 7.

Our numerical simulations also include natural taxonomies. We assumed that the features subjects listed reflected their representations (see Rosch & Mervis, 1975). In addition, following Tversky and Hemenway (1984) and Tanaka and Taylor (1991), we assumed that one feature was never listed for two contrasting categories (see 4.1.1.1 Rationale for using listed features to approximate the structure of natural objects). Five natural taxonomies had a greater redundancy at the intermediate level:

In Rosch et al. (1976), Experiment 7, subjects had to verify the name of 18 objects at three levels of categorisation. These objects belonged to six non-biological taxonomies: *musical instruments*, *fruit*, *tool*, *clothing*, *vehicle*, and *furniture*. Three pictures per category structure were carefully selected. In a previous experiment (Experiment 1), subjects had to list the attributes of these categories. The mean number of added features at the lowest-level was 1.85, at the mid-level was 5.55 , and at the highest level was 3.5 (these numbers arise from Rosch et al., 1976, Table 2, non-biological taxonomies, raw tallies; we rounded these averages to integers for the simulations).

Tanaka and Taylor's (1991) subjects were taught the names of 16 natural animals at three levels of categorisation (e.g., animal, dog, Beagle). They were either bird experts (and dog novices), or dog experts (and bird novices). In Experiment 1, Tanaka and Taylor found that novices listed approximately 8, 12, and 7 new features for the higher, middle, and lower levels of categorisation, respectively (we have extracted these figures from

Tanaka and Taylor, 1991, Figure 1, and then we have rounded them to the nearest integers). In Table 6a we give the mean verification RTs for bird and dog novices together.

Johnson and Mervis's (1997, Experiment 1, Songbirds condition) participants had to list the features of songbirds at four levels of abstraction (superordinate, basic, subordinate, and sub-subordinate). The participants were either novices, tropical freshwater fish experts, intermediate songbird experts, or advanced songbird experts (we have pooled the data obtained from the tropical freshwater experts with that obtained from the novices because no significant difference between the two groups was found). Their advanced songbird experts listed 1.75, 5, 6.02, and 3.75 for the superordinate, basic, subordinate and sub-subordinate levels, respectively. For the intermediate songbird experts, these numbers were 1, 4.87, 4.28, and 2.47, for the same levels. For the novices and the tropical freshwater fish experts, the numbers were 1.08, 2.47, 0.23, and 0.02. (For the simulations, we have multiplied the average number of features by 100 to end up with a integer value with sufficiently fine discrimination.)

In all these experiments, feature redundancy was therefore the sole determinant of basic-levelness. Table 6a reveals that basic-levelness was a direct function of the average number of redundant attributes at each level.

Table 6: Observations as well as numerical predictions of feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP for 21 published basic-level experiments. For each taxonomy, the greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

Table 6a: Variations of internal practicability determines basic-levelness. Faster access at an intermediate level.

Source	Model	Level H - 3	H – 2	H - 1	Highest (H)
Murphy & Smith, Exp. 1	Observation		723 ms	678 ms	879 ms
	Possession		1	3	3
	Utility		0.453	0.725	0.688
	Compression		865 bits	2277 bits	3569 bits
	Context		.769	.988	.993
	SLIP		1.667 attempts	1.25 attempts	1.667 attempts
Murphy, Exp. 4, Simple	Observation		862 ms	811 ms	980 ms
	Possession		1	3	3
	Utility		0.453	0.725	0.688
	Compression		865 bits	2277 bits	3569 bits
	Context		.769	.988	.993
	SLIP		1.667 attempts	1.25 attempts	1.667 attempts
Mervis & Crisafi, Exp. 1	Observation		3rd	1st	2nd
	Possession		1	4	2
	Utility		0.609	1.094	1.063
	Compression		5992 bits	12916.59	18985.00
	Context		.769	.996	.999
	SLIP		1.75 attempts	1.273 attempts	1.556 attempts
Murphy, Exp. 4, Enhanced	Observation		1,132 ms	854 ms	955 ms
	Possession		1	5	1
	Utility		0.640	1.100	0.938
	Compression		865 bits	2277 bits	3569 bits
	Context		.769	.999	1.000
	SLIP		1.75 attempts	1.167 attempts	1.75 attempts
Rosch et al., Exp. 7	Observation		659 ms	535 ms	591 ms
	Possession		2	6	4
	Utility		1.030	1.701	1.874
	Compression		0 bit	85 bits	185 bits
	Context		.917	1.000	1.000
	SLIP		1.714 attempts	1.333 attempts	1.5 attempts
Tanaka & Taylor, Novice	Observation		777.5 ms	677.5 ms	745.5 ms
	Possession		7	12	8
	Utility		2.387	3.517	3.934
	Compression		0 bit	85 bits	185 bits
	Context		1.000	1.000	1.000
	SLIP		1.588 attempts	1.385 attempts	1.543 attempts
Johnson & Mervis, Songbird, Novice	Observation	~2100 ms	~1950 ms	~1600 ms	~1900 ms
	Possession	2	23	247	108
	Utility	16.323	32.519	60.778	59.429
	Compression	0 bit	865 bits	2277 bits	3569 bits
	Context	.506	.572	.958	.978
	SLIP	1.990	1.886 attempts	1.212 attempts	1.557 attempts
Johnson & Mervis, Songbird, Intermediate	Observation	~1725 ms	~1600 ms	~1550 ms	~1800 ms
	Possession	247	428	487	100
	Utility	65.551	115.417	137.706	128.219
	Compression	0 bit	865 bits	2277 bits	3569 bits
	Context	.951	.999	1.000	1.000
	SLIP	1.673	1.493 attempts	1.443 attempts	1.853 attempts
Johnson & Mervis, Songbird, Advanced	Observation	~1600 ms	~1625 ms	~1500 ms	~1750 ms
	Possession	375	602	500	175
	Utility	86.434	149.056	165.302	167.574
	Compression	0 bit	865 bits	2277 bits	3569 bits
	Context	.989	1.00	1.00	1.00
	SLIP	1.630	1.466 attempts	1.535 attempts	1.808 attempts

4.1.1.1 Rationale for using listed features to approximate the structure of natural objects

With artificial objects, we can use the construction features to build the category structures. What about natural objects? No one knows the true construction features of natural objects. There is some empirical evidence that the features listed by participants reflect quite well the construction features of objects. In his Experiment 1A, Murphy (1991a) asked some subjects to list features for the artificial objects used in Murphy, Experiment 3. He found a mean of 1 feature at the higher level of categorisation, a mean of 5.75 features added at the middle level of categorisation, and a mean of 0.87 features added at the lower level of categorisation (cf. the "true" numbers of added features were 1, 3, and 1, at the higher, middle, and lower levels of categorisation, respectively). Additional support is given by Mervis and Crisafi (1982, Experiment 2). Adults listed an average of 1.47 added features at the superordinate level, of 4.06 added features at the basic level, and of 0.82 added features at the subordinate level (cf. the "true" number of added features were, 2, 4, and 1, at the superordinate, basic, and subordinate levels of categorisation, respectively).

The good news is that Rosch et al. (1976), Tanaka and Taylor (1991), and Johnson and Mervis (1997) conducted verification as well as feature-listing experiments with natural objects; the bad news is that none of them reported these listed features, only the average number of added features per level of abstraction.

All is not lost: We know from Tversky and Hemenway (1984) as well as from Tanaka and Taylor (1991) that relatively few added features listed for any given category overlapped with those listed for contrasting categories (e.g., 2/25 in the fruit hierarchy of Tversky and Hemenway, 1984). "The relatively high percentage of non overlapping features indicated that subjects listed features with respect to some implicit contrast set, which appeared to be objects that shared the same level of abstraction (Tversky and Hemenway, 1984)." (Tanaka and Taylor, 1991, p. 464). Here we will assume that none overlapped.

The only piece of information now missing that is needed to reconstruct the taxonomy is the distribution of features among categories of a given level of abstraction. We will assume here that it was a uniform distribution (i.e., the number of added non overlapping features was constant within level of categorisation). This assumption has very few consequences for the considered basic-levelness measures.

4.1.2 Faster access at the lower level

Murphy and Smith (1982, Experiment 3) used eight of the artificial tools from Experiment 1 (see Figure 8), and added eight new tools to produce a total of sixteen. Their artificial tools were either *large*, or *small* (higher level); they were either *pounders*, *cutters*, *scrapers*, or *stirrers* (middle level); and they had non-overlapping features *handles*, *shafts*, and *heads* (lower level). Each low-level category contained two highly similar tools (e.g., *serrated* and *straight-edge knifes*). Figure 10 illustrates the abstract organization of the features. It also shows that the lower-level categories were more practicable because they had more redundant attributes. Table 6b reveals that these categories were accessed faster than categories at the other levels.

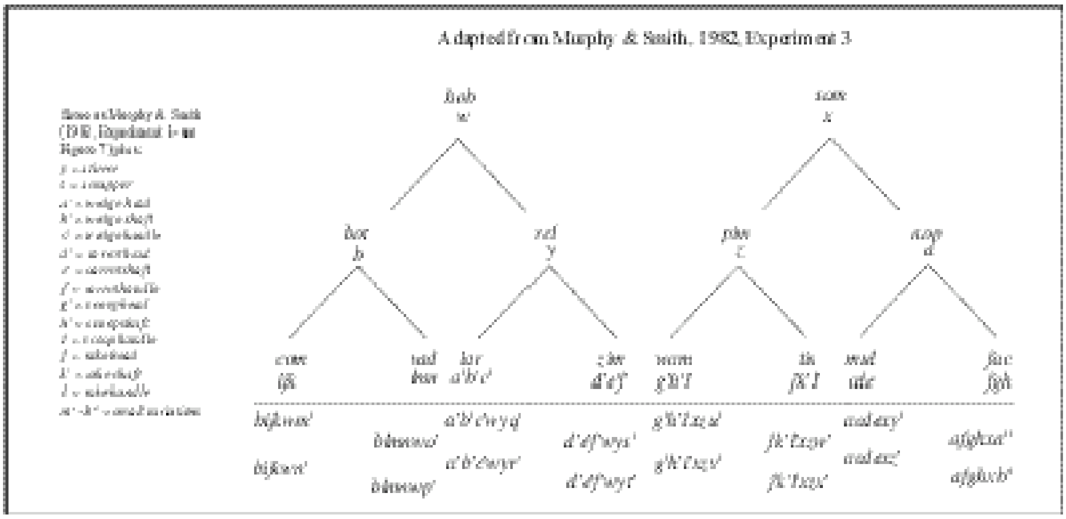


Figure 10. Taxonomy of Murphy & Smith (1982, Experiment 3), the only experiment with varying redundancy that exhibited an advantage at the lower level of categorization. Underneath the category names, we provide the optimal strategies fed to SLIP. At the bottom of the taxonomy, the abstract feature constitution of all exemplars is given. An index for these abstract features is provided left of the taxonomy. The feature constitution of all exemplar is giving underneath the taxonomy.

Tanaka and Taylor's (1992, Expert) is a variation on this theme: they used expertise to "add" redundant features at the lower level and thus speed up its access (remember section *1.3.1 Tests of the differentiation model*). Their subjects listed approximately 8, 10, and 10 new features for the superordinate, basic, and subordinate levels of categorisation, respectively (we have extracted these figures from Tanaka and Taylor, 1991, Figure 1, and then we have rounded them into integers). Compare this with 8, 12, and 7 for the superordinate, basic, and subordinate levels in their Novice condition (see previous section). They found that the basic and subordinate categories were equally fast and the superordinate categories the slowest (Table 6b gives the mean RTs of bird and dog experts).

Table 6b: Variations of internal practicability determines basic-levelness. Faster access at the lower level.

		Level		
Source	Model	H - 2	H - 1	Highest (H)
Murphy & Smith, Exp. 3, Size	Observation	574 ms	882 ms	666 ms
	Possession	3	1	1
	Utility	0.483	0.428	0.561
	Compression	865 bits	2277 bits	3569 bits
	Context	.974	.984	.991
	SLIP	1.25 attempts	1.667 attempts	1.667 attempts
Tanaka & Taylor, Expert	Observation	621.5 ms	623.0 ms	728.5 ms
	Possession	10	10	8
	Utility	2.526	3.258	3.870
	Compression	0 bits	85 bits	185 bits
	Context	1.000	1.000	1.000
	SLIP	1.474 attempts	1.474 attempts	1.556 attempts

4.1.3 Faster access at the higher level

In his Experiment 5, Murphy used eight of Murphy and Smith's 16 artificial tools. He added a set of unique values on four dimensions (*colours*, *textures*, *edges*, and *size cues*) to the high-level categorisations. Figure 11 shows that this level becomes more practicable and Table 6c reveals that it was indeed accessed faster. Note: low-level categories contained only one item here.

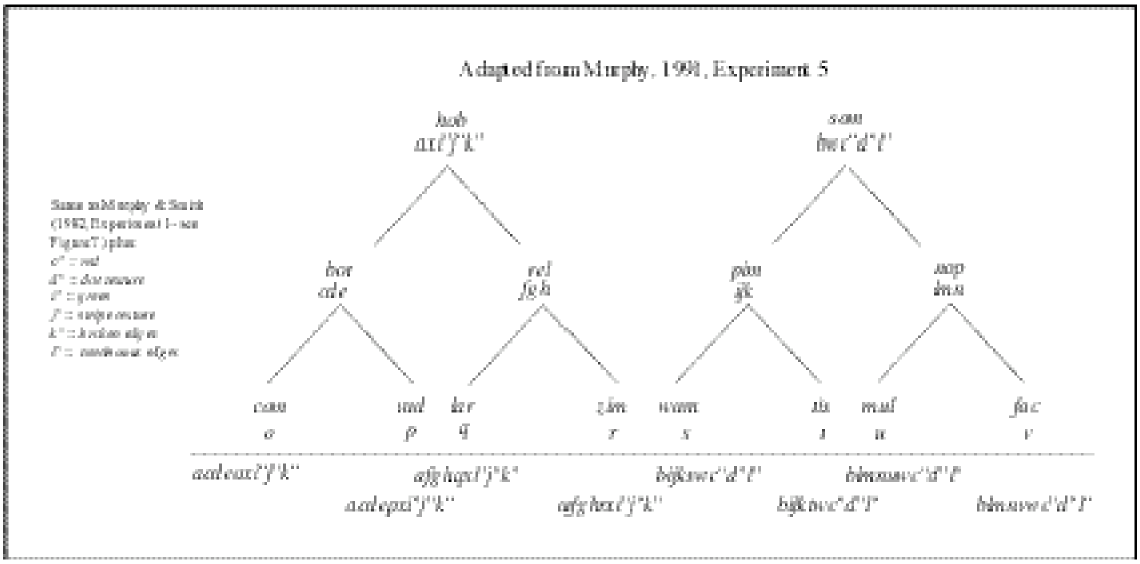


Figure 11. Taxonomy of Murphy (1991, Experiment 5), the only experiment with varying redundancy that exhibited an advantage at the

higher level of categorization. Underneath the category names, we provide the optimal strategies fed to SLIP. At the bottom of the taxonomy, the abstract feature constitution of all exemplars is given. An index for these abstract features is provided left of the taxonomy. The feature constitution of all exemplar is giving underneath the taxonomy.

Table 6c: Variations of internal practicability determines basic-levelness. Faster access at the higher level.

		Level		
Source	Model	H - 2	H - 1	Highest (H)
Murphy, Exp. 5	Observation	1,072 ms	881 ms	854 ms
	Possession	1	3	5
	Utility	.703	1.281	1.688
	Compression	0 bits	85 bits	185 bits
	Context	.769	.988	1.000
	SLIP	1.8 attempts	1.5 attempts	1.286 attempts

4.2 Variations of strategy length determines basic-levelness

In all the experiments reviewed so far, the length of categorisation strategies was constant–only one feature test was required in each case. Variations of strategy lengths were first tested in Hoffmann and Ziessler (1983, Hierarchy I). They used 16 artificial objects similar to "PacMan ghosts" (see Figure 12) organised in the top taxonomy of Figure 13. Strategy length was 1 at the high- and middle- levels, but 2 at the low-level. At the top level, the objects were defined by a *shell* (either *curved* or *rectangular*), and at the middle level by an *interior shape* (either *square*, *triangle*, *star*, or *circle*). To identify an object at the low categorisation level, however, the combination of a *shell* value and of a *bottom edge* value (*broken vertical lines*, *triangular*, *rectangular*, or *circular saw teeth*) is required. Two objects with different non diagnostic textures were members of each low-level category. Participants accessed the high- and mid- level categories equally fast, and were slower for low-level categories (see Table 6d).

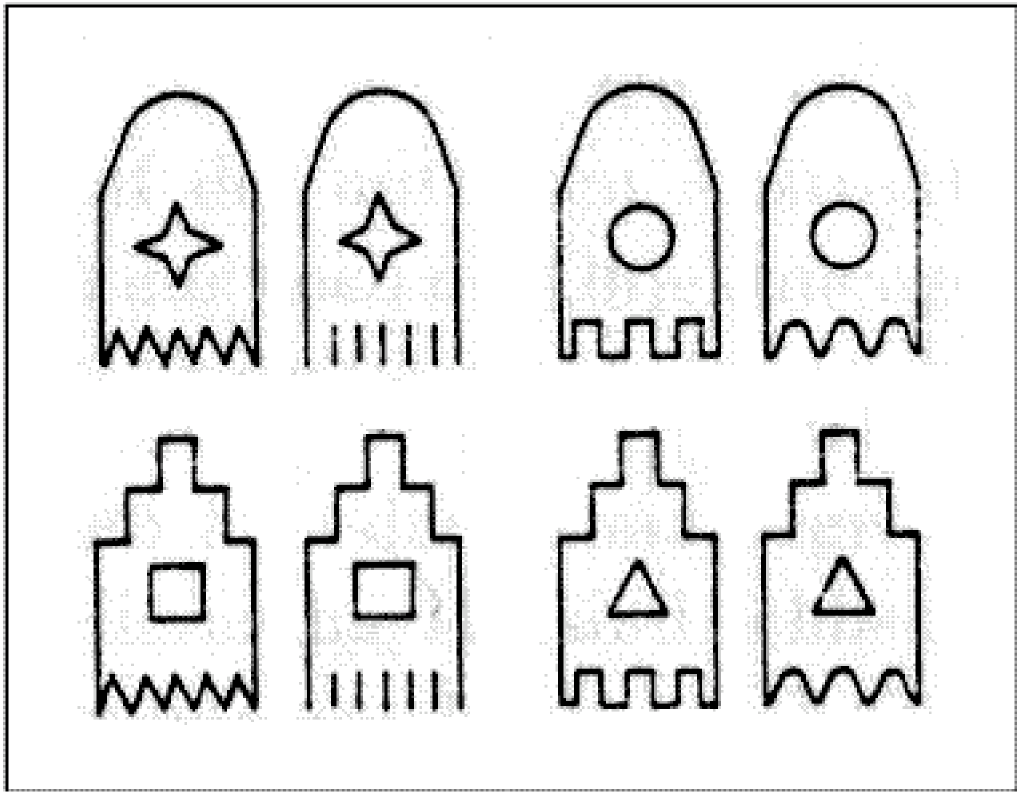


Figure 12. Complete set of "PacMan ghosts" used by Hoffmann & Ziessler in their experiments, leaving aside the two nondiagnostic textures. Scanned from Hoffmann & Ziessler (1982, Figure 2).

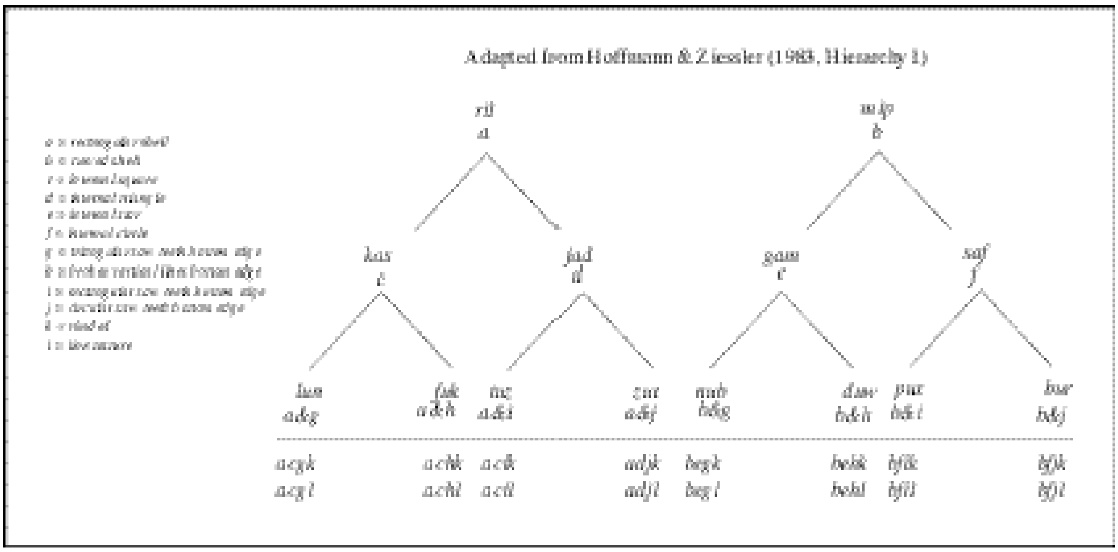


Figure 13. Abstract taxonomies of all experiments with varying strategy length. At the top: Hoffmann & Ziessler (1983, Hierarchy I); at the bottom: Gosselin & Schyns (1998a). Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy. An index for these abstract features is also provided.

Table 6d: Variations of strategy length determines basic-levelness.

		Level		
Source	Model	H - 2	H - 1	Highest (H)
Hoffmann & Ziessler, Hier. I	Observation	~700 ms	~500 ms	~500 ms
	Possession	1	1	3
	Utility	0.25	0.375	0.375
	Compression	865 bits	2277 bits	2759 bits
	Context	.769	.878	.896
	SLIP	3.2 attempts	1.6 attempts	1.6 attempts

4.3 Disjunctions and mixtures

We do not believe that there are many–if any at all–natural categories which have disjunctive strategies (see also Smith & Medin, 1981). However, some artefact concepts are clearly disjunctive (e.g., a *strike* in baseball is either a *called*, or a *swinging strike*), and several basic-level experiments have examined disjunctive categories.

4.3.1 Simple disjunctions

Figure 14 illustrates the Hierarchy II of Hoffmann and Ziessler (1983). They used the 16 objects from their Hierarchy I. At the middle level, the objects were defined by an *interior shape*; at the bottom level, by the conjunction of a *bottom edge* and an *interior shape*; and, at the top level, by the disjunction of two *interior shapes*. Two exemplars varying on non diagnostic internal texture belonged to each low-level category. The results (see RT in Table 6e) revealed that mid-level categories were accessed fastest, with high- and low-level categories equally slow (see also Corter, Gluck & Bower, 1988, for a replication using categories of artificial diseases and conceptual features).

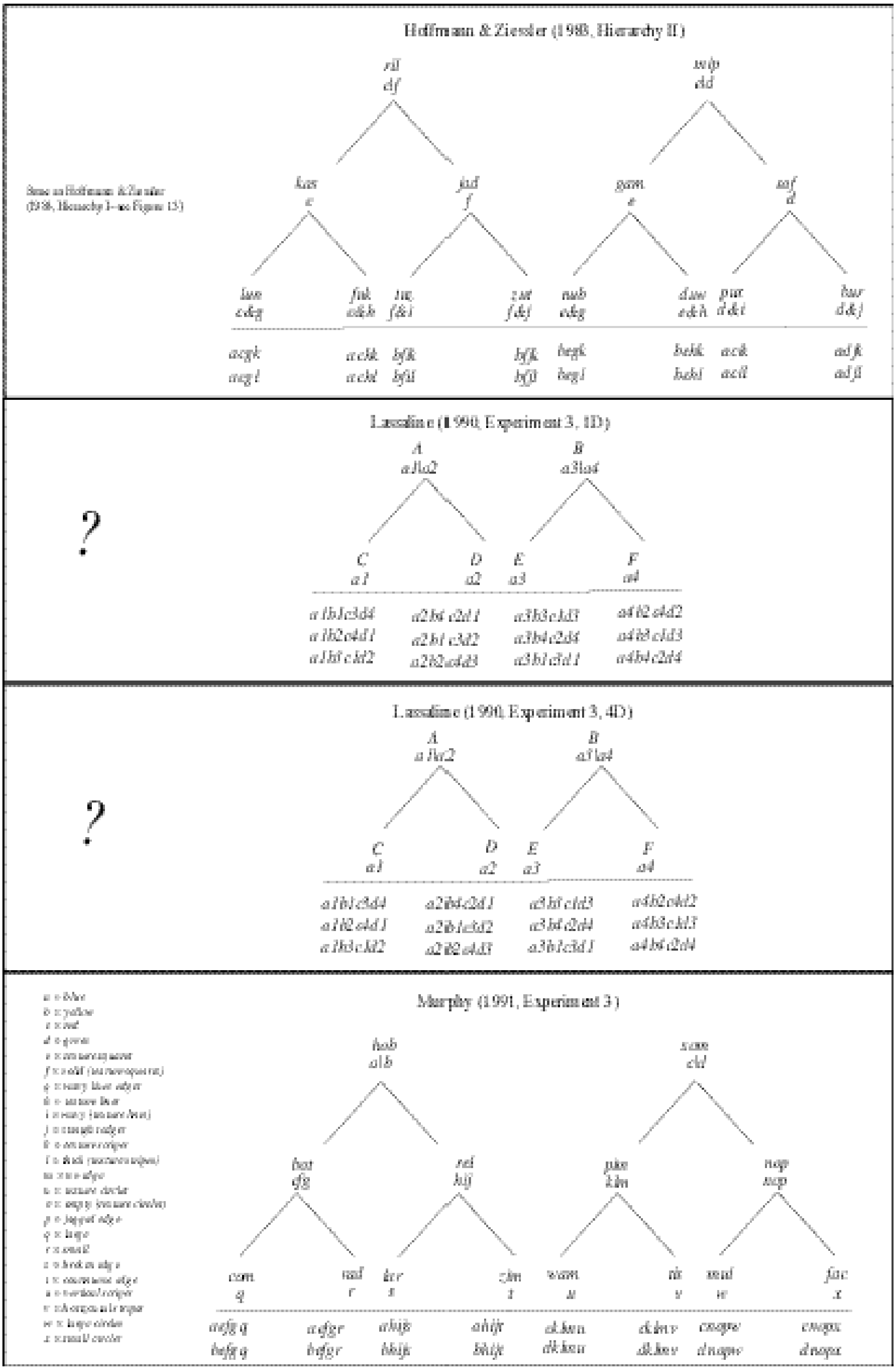


Figure 14. Taxonomies of all experiments with simple disjunctions. From top to bottom: Hoffmann & Ziessler (1983, Hierarchy II); Lassaline (1990, Experiment 3, 1D); Lassaline (1990, Experiment 3, 4D); Murphy (1991, Experiment 3). Underneath the category names, we provide the optimal

strategies fed to SLIP. The feature constitution of all exemplar is giving underneath each taxonomy. When possible, an index for these abstract features is also provided.

Table 6e: Simple disjunctions.

Source	Model	Level H - 2	 H - 1	 Highest (H)
Hoffmann & Ziessler, Hier. II	Observation	~700 ms	~500 ms	~700 ms
	Possession	1	2	2
	Utility	0.25	0.375	0.125
	Compression	865 bits	2277 bits	189 bits
	Context	.769	.960	.853
	SLIP	3.2 attempts	1.6 attempts	3.127 attempts
Corter, Gluck & Bower	Observation	3,045	2,567	3,115
	Possession	1	2	0
	Utility	0.25	0.375	0.125
	Compression	0 bits	85 bits	0 bits
	Context	.769	.960	.853
	SLIP	3 attempts	1.5 attempts	2.863 attempts
Lassaline, Exp. 3, 1-Dim.	Observation		1 st	2 nd
	Possession		1	6
	Utility		.259	.167
	Compression		119 bits	135 bits
	Context		.908	.873
	SLIP		3.694 attempts	1.818 attempts
Lassaline, Exp. 3, 4-Dim.	Observation		2 nd	1 st
	Possession		9.25	0
	Utility		.316	.048
	Compression		527 bits	389 bits
	Context		.944	.900
	SLIP		3.694 attempts	1.818 attempts
Murphy, Exp. 3	Observation	776 ms	688 ms	779 ms
	Possession	1	3	2
	Utility	0.531	0.719	0.563
	Compression	865 bits	2277 bits	3569 bits
	Context	.769	.988	.989
	SLIP	2 attempts	1.333 attempts	2.666 attempts

Lassaline (1990; also reported in Lassaline, Wisniewski, & Medin, 1992) constructed a disjunctive, two-level taxonomy using 16 artificial tools similar to those of Murphy and Smith (1982). Each one of the 16 objects was constructed by combining single values on each of four dimensions:

outer shape (corresponding to the head of Murphy & Smith's tools), *texture of a middle rectangle* (corresponding to the handle of Murphy & Smith's tools), *the texture of a small rectangle on the end of the object*, and a *shape at the end of each object*. For each subject, Lassaline randomly assigned the values of these dimensions to the letters of the two middle taxonomies of Figure 14. In two conditions (One-Dimension and Four-Dimension) of her Experiment 3, two-feature disjunctions defined the high level and a single feature defined each low-level category. In the One-Dimension condition, the features defining the low-level categories were extracted from a single dimension; in the Four-Dimension condition, the features defining the four low-level categories were extracted from four different dimensions (see Table 6e). A verification advantage was found at the low-level in the One Dimension condition, but, surprisingly, a verification advantage for the high-level was reported in the Four Dimensions condition (see Table 6e).

Murphy (1991) used 16 artificial "stamps" of various *colours*, *textures*, *types of edge*, and *sizes* (see Figure 15). His taxonomy was very similar to Murphy and Smith's (1982, Experiment 1).

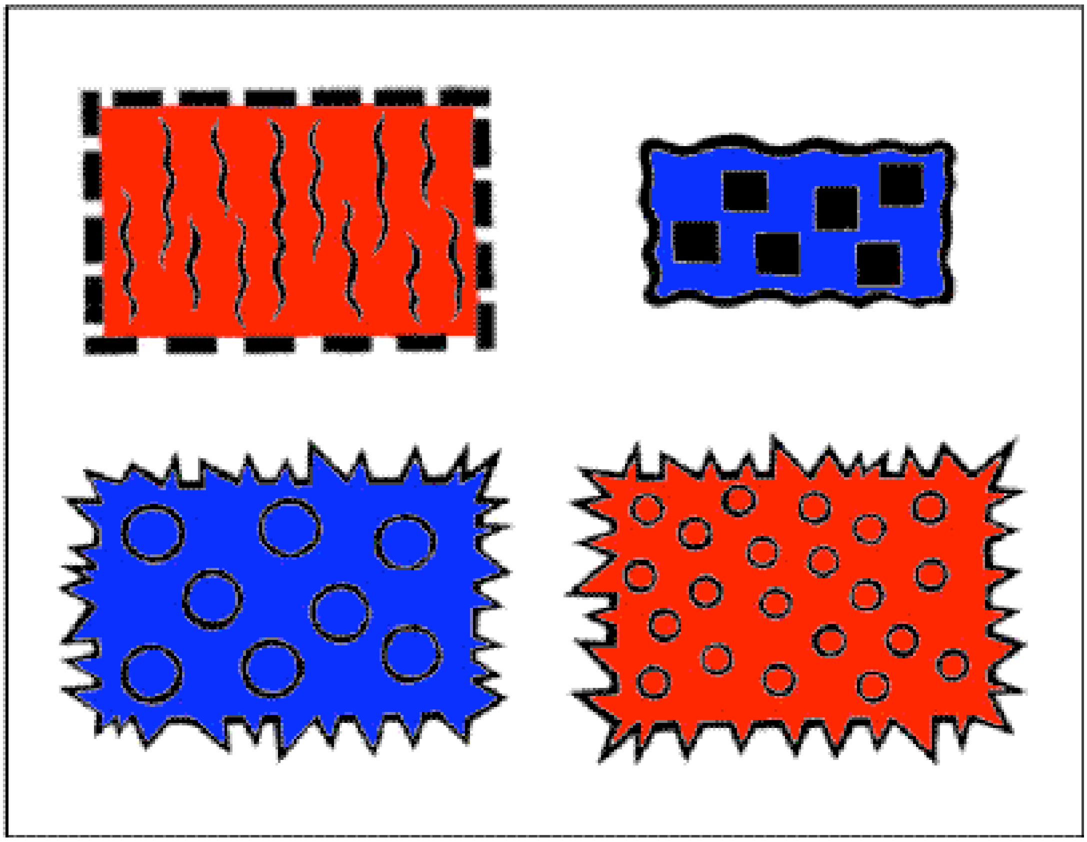


Figure 15. Sample artificial "stamps" Murphy used in his Experiment 3. Scanned from Murphy (1991a, Figure 1) and then coloured.

The only difference between the two was that, as Figure 14 shows, disjunctions of features defined the higher level categories (i.e., *nop* = *blue* or *yellow*; *som* = *red* or *green*). Table 6e reveals that middle level categories were the fastest with the other two being equally slow.

4.3.2 Mixtures of disjunctions and conjunctions

In their Hierarchy III, Hoffmann and Ziessler arranged the objects of their Hierarchy I in yet another way. Conjunctions of a *shell* and a *bottom edge* defined the low-level categories; disjunctions of two two-feature conjunctions defined the mid-level; disjunctions of four two-feature conjunctions defined the high-level (see Figure 16). In this taxonomy, low-level categories were accessed faster than the mid-level, itself faster than the low-level (see RT estimations in Table 6f).

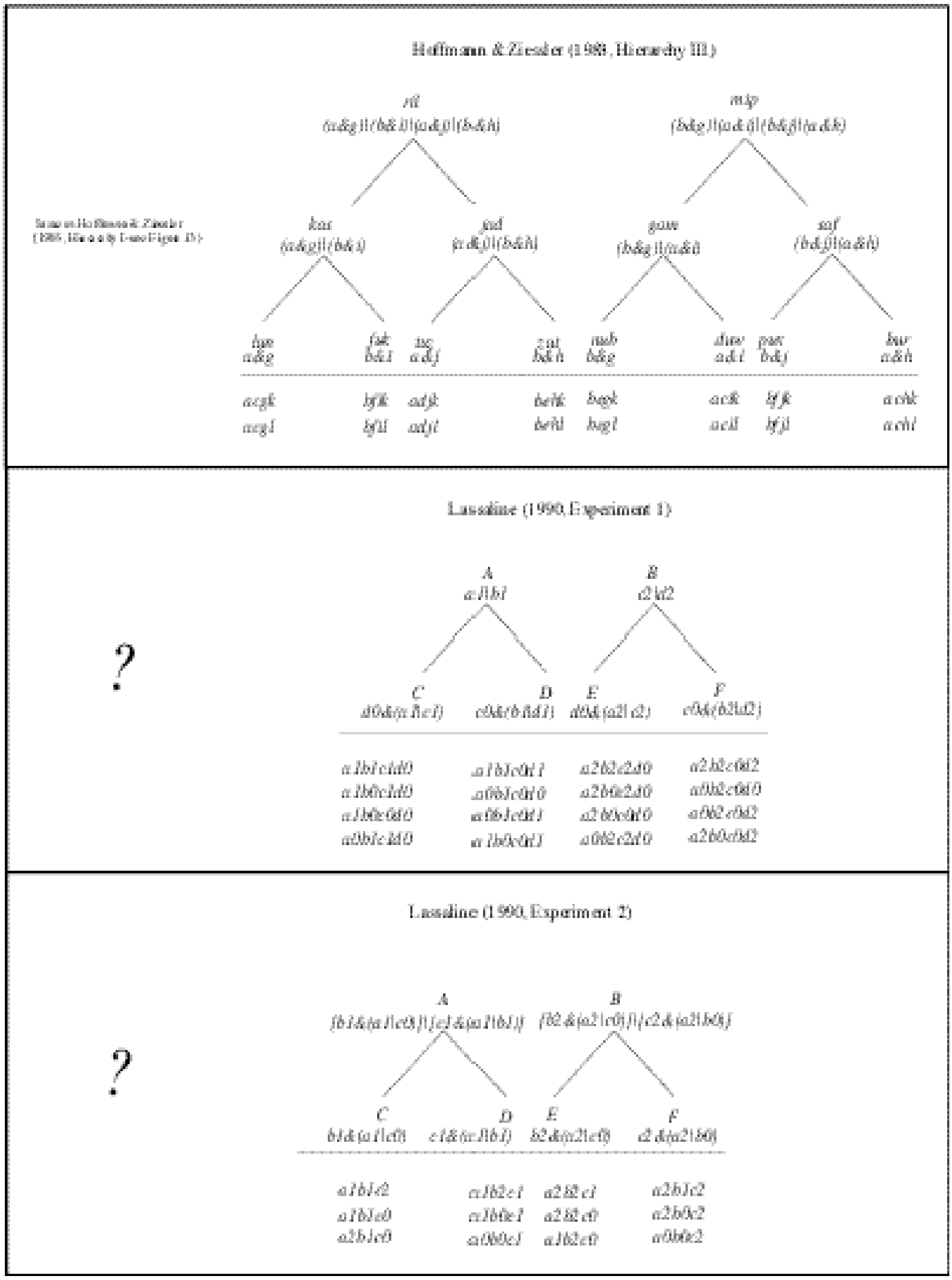


Figure 16. Taxonomies of all experiments with mixtures of disjunctions and conjunctions. From top to bottom: Hoffmann & Ziessler (1983, Hierarchy III); Lassaline (1990, Experiment 1); Lassaline (1990, Experiment 2). Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplar is giving underneath each taxonomy. When possible, an index for these abstract features is also provided.

The taxonomy of Lassaline (1990, Experiment 1) is shown in Figure 16. Two-feature disjunctions defined categories at the high level, and a conjunction of a feature with a disjunction of features defined the low level. In these conditions, the higher level was accessed faster (see Table 6f). Figure 16 also illustrates the taxonomy of Lassaline's (1990) Experiment 2. A conjunction of a feature with a disjunction of features defined the low level. A disjunction of two such strategies defined the high-level. Low-level categories were faster to verify (see Table 6f).

Table 6f: Mixtures of disjunctions and conjunctions.

		Level		
Source	Model	H - 2	H - 1	Highest (H)
Hoffmann & Ziessler, Hier. III	Observation	~700 ms	~1050 ms	~1475 ms
	Possession	3	0	3
	Utility	0.25	0.125	0
	Compression	865 bits	2 bits	-22 bits
	Context	.974	.898	.693
	SLIP	3.2 attempts	4.727 attempts	6.254 attempts
Lassaline, Exp. 1	Observation		2nd	1st
	Possession		2.5	3
	Utility		.127	.238
	Compression		1359 bits	1479 bits
	Context		.878	.924
	SLIP		4.574 attempts	3.225 attempts
Lassaline, Exp. 2	Observation		1st	2nd
	Possession		2	2
	Utility		.209	.119
	Compression		380 bits	106 bits
	Context		.870	.719
	SLIP		3.909 attempts	5.273 attempts

4.4 Comparison of the models with respect to the published experiments

We have now reviewed the main experiments on the basic-level, emphasising which factor of SLIP varied in each one of them. We now apply the models reviewed earlier (including SLIP) to the task of predicting the pattern of time to access the categories. Before turning to the results of these simulations, it is important to specify the parameters we used for each of these models: In SLIP, parameters were set to $S = .5$ and to $= .05$. Having said this SLIP is rather insensitive to changes of S (as S increases, all the average number of attempts needed to complete a strategy increase proportionally) and only comes into play with disjunctive strategies (a decrease in increases the time to complete disjunctions relative to conjunctions). Jones's category feature-possession comprises a single free parameter. Following Jones (1983), we set it to 1, but changing this parameter does not much affect category feature-possession. Medin and Schaffer's (1978) context model also has a single dissimilarity parameter which we set to .3, following Estes (1995).

Overall, SLIP predicted 74% of the experimental results, winning the competition. Second best was Jones's category feature-possession with 63%, then came Corter and Gluck's category utility with 61%, Medin and Schaffer's (modified by Estes) context model with 45%, and Pothos and Chater's compression measure with 38% (see Table 7). Monte-Carlo simulations revealed that neither the compression measure ($p < .29$, *ns*.), nor the context model ($p < .06$, *ns*.) perform significantly better than the chance model (i.e., the chance model randomly selects the ranks of the levels within each experiment); all the other basic-levelness measures significantly outperform the chance model (i.e., in the worst case $p < .001$).

Table 7: Percentage of nominal data from 21 published basic-level experiments explained by feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. The internal practicability, simple conjunction, and mean scores flanked by a star are significantly above chance levels ($p < .02$).

	Strategy length	Internal practicability	Simple conjunction	Mixture	Mean
Possession	33%	72% *	46%	57%	63% *
Utility	100%	49% *	69% *	100%	61% *
Compression	67%	28%	31%	100%	38%
Context	67%	36%	38%	100%	45%
SLIP	100%	74% *	54%	100%	74% *

It is instructive to examine the models specifically for their predictions with respect to variations of feature internal practicability, strategy length, simple conjunction, and mixture experiments. Table 7 summarises the breakdown. For experiments involving only variations of redundancy, the predictions of each model mirror its overall scores. SLIP scores a 74%, followed by category feature-possession with 72%, then by category utility with 49%, then by context model with 36%, and trailed by compression with 28%. Monte Carlo simulations have shown that SLIP ($p < .001$), category feature-possession ($p < .001$), and category utility ($p < .019$) significantly outperform the chance model. For the only experiment involving variations of strategy length, SLIP and category utility both explain 100% of the data, followed by compression and context model with 67%, and by feature-possession with 33%. Due to the scarcity of data, no Monte Carlo simulation was performed here. For the simple conjunction experiments, category utility is first with a score of 69%, then SLIP with 54%, feature-possession with 46%, context model with 38%, and finally compression with 31% of data predicted. Monte Carlo simulations have shown that only category utility significantly outperforms the chance model ($p < .011$). For the mixture experiments, all models explained 100% of the data except category feature-possession which explained 57%. Again, due to the scarcity of data, no Monte Carlo simulation was performed here.

In sum, we have compared the predictions of SLIP with those of other models of speed of access to categories using data drawn from 21 classic basic level experiments. It appears that SLIP did the best job with an overall prediction score of 74%.

4.4.1 Further predictions of SLIP

SLIP also makes new predictions which go beyond those possible in other models of basic-levelness. For example, it predicts positive linear relationships between strategy length and RTs and negative linear relationships between redundancy and RT. Unfortunately, Hoffmann & Ziessler (1983, Hierarchy I) used two different strategy lengths, and a line can always pass perfectly through two points. For redundancy, however, using all published three-plus-level taxonomies, the mean correlation is r = -.851 (r varies between -1.000 and -.666; 3/11 achieved significance at the $p < .05$ level). This is a high fit, considering the difficulty involved in evaluating the exact number of redundant features subjects used (Schyns, Goldstone & Thibault, 1997).

4.5 A cautionary note about coding

Even though coding is critical for numerical simulations, it is hardly ever discussed. This short section is meant as a coding "case study". We will compare Corter and Gluck's (1992) widely used coding of Murphy and Smith's (1982, Experiment 1) artificial tools with our own. This comparison will highlight two coding rules of thumb that we have used throughout this dissertation.

Murphy and Smith described their category structure as follows: "[I]f one considers hand tools to consist of a handle, a shaft and a head then each of the [four] basic tools were designed to be distinct from the others in each part [....] [We coded this as three obvious dimensions–*handle*, *shaft*, and *head*–that can take four values each (or 12 features).] To form the superordinates, the hammer and the brick were grouped together to produce [...] pounders, while the knife and pizza cutter were grouped together to form [...] cutters. [We coded this as one dimension–*type of tool*–with two values (or two features).] Each of the four basic tools was differentiated into two subordinates in the following ways: (1) the hammer had a wide or narrow head; (2) the brick had a single or a two-part handle; (3) the pizza cutter had a long or a short shaft [...]; and (4) the knife's edge was serrated or straight. [This was coded as one dimension–*part modifier*–that can take eight values or, alternatively, as eight features.]" (p. 3). We should add that each subordinate category contained two artificial tools: a large one and a small one. Thus, we coded each of Murphy and Smith's artificial tools by a set of values on six dimensions (or by 24 binary features).

Corter and Gluck coded them by a vector of values on four dimensions: *handle* (*single brick handle*, *two-part brick handle*, *hammer handle*, *pizza cutter handle*, and *knife handle*), *shaft* (*long pizza cutter*, *short pizza cutter*, *brick*, *hammer*, or *knife shaft*), *head* (*wide hammer*, *narrow hammer*, *brick*, *pizza cutter*, *serrated knife edge*, or *straight knife edge head*), and *size* (*large* or *small*). This corresponds to 19 binary features.

Figure 17 illustrates the resulting taxonomies, and Table 8 shows that Corter and Gluck's coding makes the predictions of SLIP, the context model, and compression fit better the observations, but that it makes no difference for the other models.

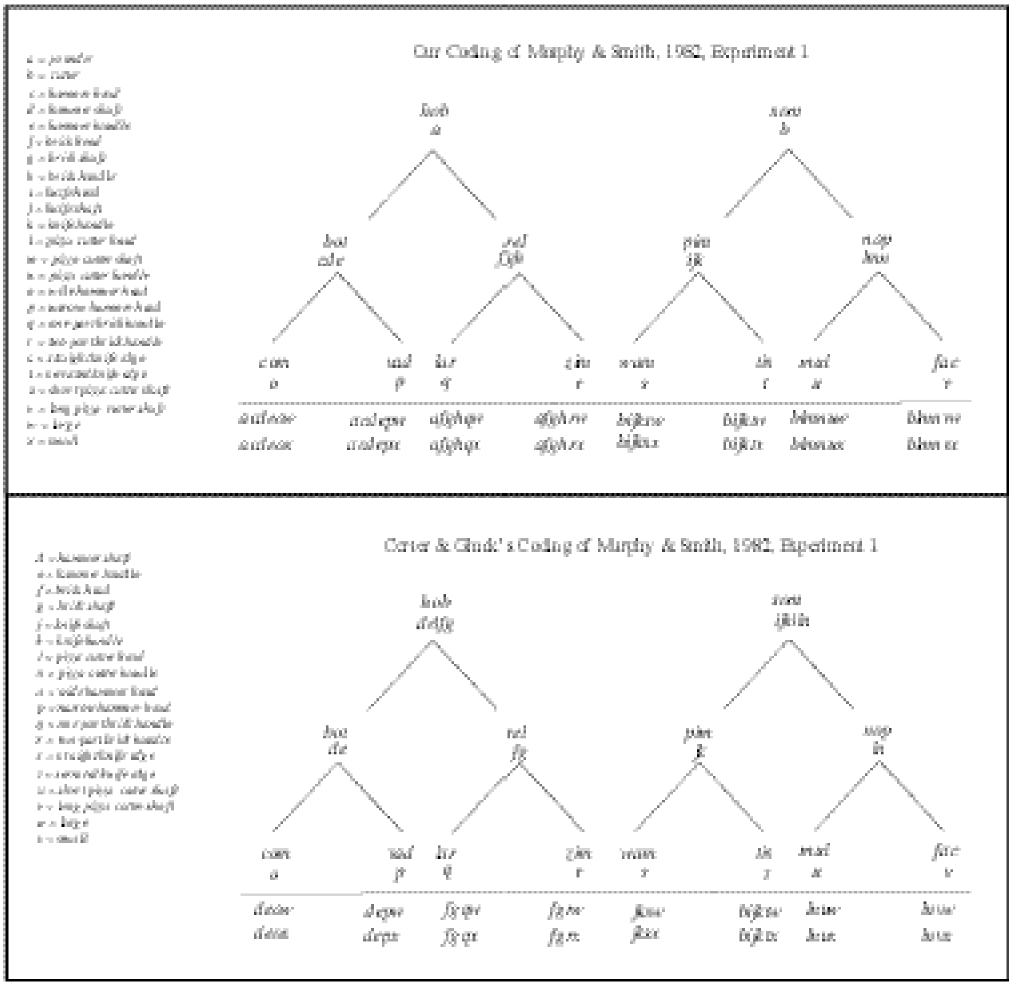


Figure 17. Taxonomies resulting from Corter & Gluck's as well as our coding of Murphy & Smith's (1982, Experiment 1) artificial tools. Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy. When possible, an index for these abstract features is also provided.

Table 8: Numerical simulations for Corter and Gluck's and our coding of Murphy and Smith's (1982, Experiment 1) artificial tools.

Model	Coding	Level		
		lower	middle	Higher
Observation		723 ms	678 ms	879 ms
Possession	Our	1	3	3
	Corter & Gluck's	1	2	2
Utility	Our	0.453	0.725	0.688
	Corter & Gluck's	0.30	0.47	0.31
Compression	Our	865 bits	2277 bits	3569 bits
	Corter & Gluck's	865 bits	2277 bits	1437 bits
Context	Our	.769	.988	.993
	Corter & Gluck's	.769	.960	.926
SLIP	Our	1.667 attempts	1.25 attempts	1.667 attempts
	Corter & Gluck's	1.6 attempts	1.333 attempts	2.414 attempts

There are two differences between these codings: (1) we used three basic part dimensions: *head* (e.g., *hammer head*), *shaft*, and *handle*, as well as one subordinate part dimension: *part modifier* (e.g., *wide* and *narrow hammer heads*), whereas Corter and Gluck integrated the modified parts in the *head* (e.g., *wide hammer* and *narrow hammer heads*), *shaft*, and *handle* dimensions; and (2) we used a superordinate *type of tool* dimension, whereas Corter and Gluck did not.

Our first rule of thumb is:

Use the intended category feature-structure, within reason.

It happens that we discussed points (1) and (2) with Murphy (1998), and he assured us that the *intended* coding was the one we use, not the one used by Corter and Gluck.

But is (1) reasonable? We believe that it is the only reasonable choice. Let us go back to the hammer example. For Corter and Gluck, *narrow hammer head* is just as dissimilar to *wide hammer head* as it is to say *pizza cutter head*. Even though hammers can have two *heads*–a *narrow* and a *wide* one–these *hammer heads* are much more similar to each other than they are to the *brick*, the *pizza cutter*, and the *knife heads*. We acknowledge this by having a *hammer head* value in a *head* dimension as well as *wide* and *narrow hammer head* values in a *part modifier* dimension.

As for (2) (i.e., using a superordinate *type of tool* dimension, unlike Corter and Gluck), it is less clear that it is the only reasonable choice. It could be argued–as we suspect Corter and Gluck as well as many others would–that *cutter* and *pounder* are not values on a perceptual dimension but that the superordinate "pounder" and "cutter" categories can only be accessed through some kind of semantic process after the recognition of the tools at the entry point of recognition, i.e. "hammer", "knife", "brick", and "pizza cutter" (see 3.4 Part-based accounts). After having recognised an object as a "hammer" (on the basis of its value on the *handle*, *shaft*, or *head* dimension), for instance, one could come to the *semantic* realisation that all hammers are pounders, and, hence, verify that the object is, in fact, a "pounder". This brings us to our second rule of thumb:

Assume that what is available to people is used by them.

Without a doubt, Murphy and Smith's participants could have extracted a perceptual *type of tool* dimension to define "cutter" and "pounder". For example, *thickness*: pounders are *thick* and cutters are *thin*. Similarly Gibson (1986) has argued that *affordances* (what a thing affords another thing; here: what cutters and pounders afford humans is *cutting* and *pounding*) are picked up directly in the world. We thus stick with (2).

Chapter 5. Empirical tests of SLIP

The following sections further examine the empirical validity of the two constraints of SLIP. So far, experiments on the basic level have been mainly motivated by empirical considerations instead of a rigorous model. As pointed out earlier, strategy length has never been tested as such. In Hoffmann and Ziessler (1983, Hierarchy I), strategy length is confounded with level of abstraction: the most inclusive level has the shortest strategy, and the least inclusive level has the longest strategy. A similar problem affects feature redundancy. Even though we showed earlier that many basic level experiments changed feature redundancy, no systematic study of this factor has been carried out. The-second-first-third-and-fourth model (i.e., a model that predicts that the highest level will have the second greatest basic-levelness, the second highest, the greatest, the third highest, the third greatest, and the fourth highest, the fourth greatest) accounts for 53% of all published data (cf. 38% for compression). A Monte Carlo simulation revealed that this model outperforms the chance model ($p < .008$). This shows that the published data set is biased. This is particularly important in differentiating between the models' behaviour because, as we have mentioned earlier, most basic-levelness measures have a bias for higher levels of categorisation (see Chapter 4).

Another problem of the reviewed experiments concerns their stimuli. One must achieve balance between control and ecological validity (e.g., Humphrey & Bruce, 1989; Bruce & Green, 1990). We have seen examples of the worst of both worlds here: On the one hand, the basic-level experiments that used artificial objects had excellent control but very poor ecological validity (e.g., Hoffmann & Ziessler's, 1982, PacMan ghosts; and Murphy's, 1991, stamps); on the other hand, the experiments that used natural objets had good ecological validity but very poor control (e.g., Rosch et al., 1976; Tanaka and Taylor, 1991).

The following experiments have been designed to overcome these shortcomings. All of them excepting one used computer-synthesised artificial 3D objects or artificial scenes to tightly control feature composition and preserve ecological validity. The first five experiments examine the two constraints of SLIP in verification. Experiment 1 isolates the effect of strategy length on basic-levelness, Experiments 2A and 2B test the effect of feature redundancy (or internal practicability), and Experiment 3 examines the interactions between the two factors. Experiment 4 examines more precisely the predictions of SLIP concerning strategy length. The last four experiments study the two computational constraints in naming. Experiment 5A isolates the effect of strategy length, and Experiment 5B that of internal practicability. Experiment 6 takes a look at the time course of length 1 and 2 strategy completion. Finally, Experiment 7 examines the effect of *robustness* (i.e., the idea that an approximate categorisation is better than none) on the order of feature test in length 2 strategies.

5.1 Introduction to Experiments 1-4[17]

The following four experiments were largely influenced by Murphy & Smith's (1982, Experiment 1). You will remember that their participants were initially taught the top artificial taxonomy of Figure 7. In a later testing phase, participants were shown a category name followed by an artificial tool similar to the ones in Figure 8. Subjects' task was to verify as quickly as possible whether the name and stimulus matched.

---

[17] Experiments 1, 2A, and 3 will appear in Gosselin and Schyns (in press), and Experiment 4 has appeared in Gosselin and Schyns (1998a).

5.2 Experiment 1

In SLIP, strategy length is the minimum number of required tests on features to access a category. Experiment 1 isolates this factor and examines how a variation of strategy length at different levels of a hierarchy influences their basic-levelness. As pointed out earlier, strategy length was shown to influence basic-levelness in Hoffmann and Ziessler (1983, Hierarchy I). However, this experiment did not dissociate strategy length from level of abstraction. That is, the highest level categories had the shortest strategy and the lowest level the longest. To overcome this problem, we designed an experiment which dissociates level of categorisation and speed of access. In the HIGH_FAST taxonomy, shorter strategies access the high-level categories faster. In LOW_FAST, the opposite applies: shorter strategies access the low-level categories faster. In both conditions, the longer strategies arose from overlap between the attributes (geons, Biederman, 1987) of categories. SLIP predicts that shorter strategies are completed faster, irrespective of categorisation levels. That is, a faster access to the high-level of HIGH_FAST, and to the low-level of LOW_FAST.

5.2.1 Method

5.2.1.1 Participants

Twenty University of Glasgow students with normal or corrected to normal vision were paid to participate in the experiment.

5.2.1.2 Stimuli

Stimuli were computer-synthesised chains of four geons similar to those of Tarr, Bülthoff, Zabinski and Blaz (1997). We designed the stimuli with the Form Z three-dimensional modelling software on a Macintosh computer. Five geons (i.e., Biederman's geometric elements) defined the categories of the HIGH_FAST taxonomy. One different geon defined each one of three high-level categories. Each one of six possible low-level categories was further specified by one of the two remaining geons. The top taxonomy of Figure 18 illustrates this.

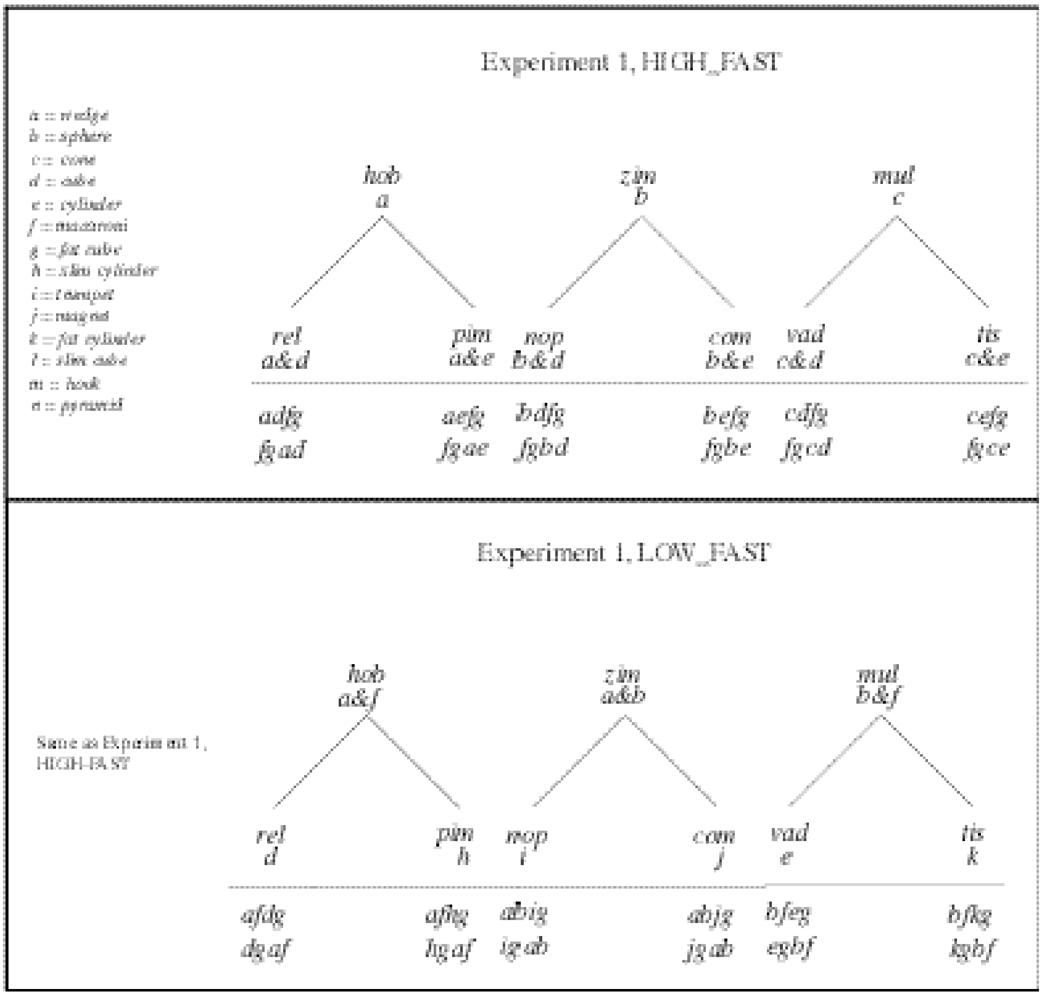


Figure 18. Taxonomies of Experiments 1, HIGH_FAST and LOW_FAST. Only strategy length was manipulated here. Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy. An index for these abstract features is also provided.

In this taxonomy, strategy length equals 1 for the higher-level categories. This means that only one feature needs to be tested to access categories at this level. Strategy length equals 2 at the lower-levels, because these categorisations require two feature tests[18]. The overlap of features across lower-level categories produced the longer conjunctive strategies.

To create the experimental stimuli, we substituted the letters in Figure 18 with their corresponding geometric elements. To these two geons, we added two supplementary geons that served as fillers. Fillers were identical across objects and so could not be used to distinguish them. We created two exemplars per low-level category by changing the location of the diagnostic geons in the chain (see Figure 19 for examples).

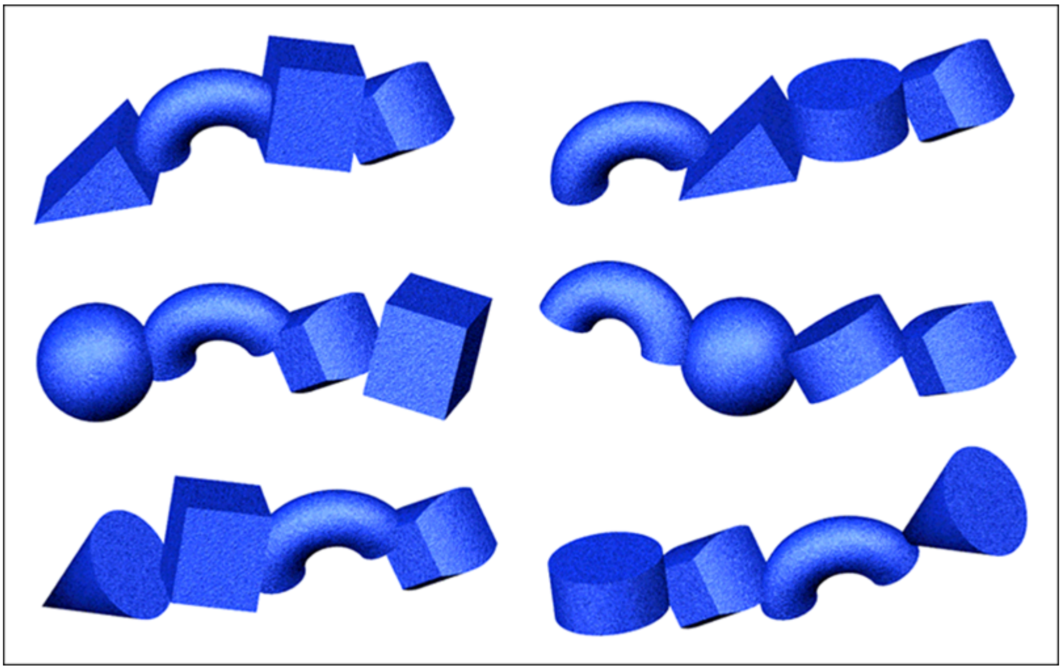


Figure 19. Sample computer-synthesised objects used in Experiment 1, HIGH_FAST (one exemplar per low-level category).

[18] Note that participants could have used two-term disjunctions instead of two-term conjunctions for high-level categories in the bottom taxonomy of Figure 18. For example, *Strat'*(X, *hob*) = *d* | *h* is extensively equivalent to *Strat*(X, *hob*) = *a&f*. To avoid this, we instructed participants to use solely the learned strategies during the experiment.

Nine geons defined the LOW_FAST taxonomy. A unique combination of two geons (sampled from a set of three) defined each one of three top-level categories (see Figure 18, bottom taxonomy). High-level strategies had length 2 because a combination of two geons defined categories at this level. A unique diagnostic geon further specified the categories at the low level. However low-level categories had length 1 strategies because a single feature test on a diagnostic geon determined membership. Figure 8 shows the LOW_FAST taxonomy. We added one filler to generate six four-geon chains. From these, we created two exemplars per category (see Figure 18, bottom taxonomy).

5.2.1.2.1 A note on these two taxonomies

There is one important difference between the two taxonomies of Figure 18: in the bottom one, the high- and low-level strategies are independent (e.g., *Strat*(*X*, *rel*) = [{"does X possess *d*?"}] is independent of *Strat*(*X*, *hob*) = [{"does X possess *a*?"} & {does X possess *f*?"}]) but in the top one they are not (e.g., *Strat*(*X*, *rel*) = [{"does X possess *a*?"}] is included in *Strat*(*X*, *hob*) = [{"does X possess *a*?"} & {does X possess *d*?"}]). It is impossible by the definition of a hierarchy for low-level strategies to be included in high-level ones. However, we can easily make high-level strategies independent of low-level ones (see Figure 20). We chose not to here because we believe that overlap is a crucial property of real-world strategies. For example, to identify your blue Porsche 911 in a parking lot also comprising a blue Toyota Tercel and a lime Porsche 911, you must examine both the *colour* and the *shape* of the cars; whereas to identify any Porsche 911 in this same parking lot, only the *shape* of the cars has to be examined.

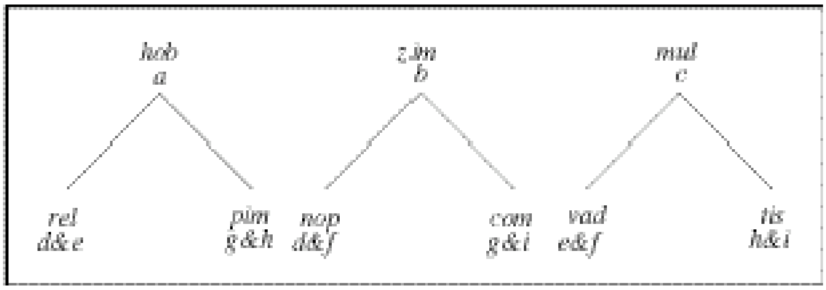


Figure 20. An alternative to the Experiments 1, HIGH_FAST, taxonomy. Here the high- and low-level strategies are independent.

5.2.1.3 Procedure

The procedure closely followed that of Murphy and Smith (1982). In a learning phase, participants were evenly split between the learning of the HIGH_FAST and LOW_FAST taxonomies. We were not interested in strategy learning; we were interested in how people use known strategies. We thus instructed participants to learn the names and the defining geon(s) of nine categories (see the specific names and corresponding geon combinations in Figure 18). Participants saw their taxonomy on a sheet of paper; this learning phase was not constrained in time.

We tested participants knowledge of the taxonomy by asking them to list the features associated with each category name. Criterion was reached when participants could list twice in a row, without any mistake, the attributes defining each category. Corrective feedback was provided.

When subjects knew the taxonomy, a category verification task measured categorisation time at each level. The experiment was ran with the SuperLab software on a Macintosh PowerPC 7200. Each trial began with the presentation of a category name. Subjects pressed one keyboard key to see the list of all learned definitions on the screen (each definition corresponded to a set of geons per category). Participants had to identify the list corresponding to the previously shown category name. This ensured that subjects had accessed the category representations. SLIP is a theory about how strategies are matched to distal objects, not about strategy remembrance. After a 200 ms delay, an object appeared on the screen. Subjects had to decide as fast as they possibly could whether or not the named category and object matched by pressing the "yes" or "no" computer keyboard keys. We recorded response latencies. Note that low-level categories are more numerous than high-level categories. We normalised the number of positive and negative trials with the constraint of equating the number of trials per level.

5.2.2 Results and discussion

We performed the analysis of RTs on the correct positive trials (error rate = 6.5%) that were within two standard deviations from the means (an additional 4.9% of the responses were discarded). Table 9 reports the mean RTs at the low- and high-levels for the two taxonomies tested (see Observation in Table 9).

Table 9: Mean RTs and standard deviations (between brackets) for the positive trials of Experiment 1 as well as predictions for feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. For each taxonomy, the greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

		Level	
	Model	Lowest	Highest
Exp. 1, HIGH_FAST	Observation	1,256 ms [405]	896 ms [323]
	Possession	2	3
	Utility	.195	.222
	Compression	0 bit	30 bits
	Context	.769	.625
	SLIP	6.4 attempts	3.2 attempts
Exp. 1, LOW_FAST	Observation	948 ms [258]	1,240 ms [305]
	Possession	1	3
	Utility	.25	.333
	Compression	0 bit	30 bits
	Context	.769	.783
	SLIP	3.2 attempts	6.4 attempts

A two-way (GROUP x STRATEGY LENGTH) ANOVA of the RTs with repeated measures on one factor (STRATEGY LENGTH) revealed a main effect of STRATEGY LENGTH, $F(1, 18) = 77.08$, $p < .0001$, (mean length 1 strategies = 922 ms verification time [standard deviation = 291 ms]; mean length 2 strategies = 1248 ms verification time [standard deviation = 355 ms]), meaning that participants systematically verified length 1 strategies faster than length 2 strategies, irrespective of the considered level (low vs. high) (e.g., Keppel, 1991). All participants verified the categories associated with length 1 strategies faster. Neither the interaction between GROUP and STRATEGY LENGTH, $F(1, 18) = .84$, *ns*, nor the main GROUP effect, $F(1, 18) = .02$, *ns*, were significant. The error rate was low overall and not correlated with RT ($r = -.17$, *ns*), ruling out a speed-accuracy trade-off.

Remember that SLIP predicts that length 1 strategies should be completed faster than length 2 strategies, irrespective of categorisation level (see SLIP in Table 9 for numerical predictions). The data reported here confirms that strategy length, rather than categorisation level, determines the basic-levelness of a category.

5.2.2.1 About generalisation

Experiments with complex computer-synthesised objects rather than, say, ASCII characters, often use a fixed set of objects (although see Cutzu & Edelman, 1998). In our case, this means that all participants learned the exact same strategies and that these strategies were applied to the exact same objects. Is our data nonetheless generalisable to other sets of features and objects?

To minimise any prima facie bias, we have selected 15 geons about equally distant on Biederman's four geon dimensions[19] (i.e., the mean number of different dimension values is 2.21 with a standard deviation of 0.39). These dimensions have been shown to be rather independent and equally salient in a large number of experiments (see Biederman, 1987, 1990; Biederman & Gerhardstein, 1993).

Moreover, Experiment 1, LOW_FAST has been replicated in Experiment 3, SL_DOWN, with a different set of geons; and Experiment 1, HIGH_FAST by part of Experiment 4 with objects defined by *colour*, *texture*, and *shape*. This does suggests that our results are not stimuli-set effects.

5.3 Experiment 2A

Practicability refers to the ease with which features identify a category at any level of a taxonomy. A category has high practicability whenever many of its defining features are uniquely diagnostic of this category, and it has low practicability when a single feature defines the category. If this factor influences the basic-levelness of a category, then it should apply equally to all levels of a taxonomy.

In Experiment 2A, all strategies had length 1 but the high and low levels differed in practicability. In the HIGH_FAST condition, high-level strategies had greater practicability than low-level strategies. The opposite applied to the LOW_FAST condition, with low-level strategies having higher practicability. SLIP predicts that categories with higher

---

[19] These four geon dimensions are: *straight* vs. *curved* axis; *straight* vs. *curved* cross-section; sides: *parallel* (constant cross-section), *diverging*, *expanding*, *contracting* (as a lemon or an American football), and *converging and diverging* (as a bow-tie); ends (for non-parallel-sided geons): *truncated* (straight), *pointed* ('L' vertex at end), and *convex rounded* (Biederman, 1998).

practicability will be verified faster, irrespective of their level in the taxonomy.

5.3.1 Method

5.3.1.1 Participants

Twenty students from University of Glasgow with normal or corrected vision were paid to participate in the experiment.

5.3.1.2 Stimuli

Stimuli were similar to those of Experiment 1: four-geon chains synthesised with the Form Z three-dimensional modelling software on a Macintosh computer. The HIGH_FAST condition used 10 diagnostic geons. Three different geons defined each one of two high-level categories; one different geon further defined each low-level category (see Figure 21, top taxonomy). We generated two exemplars per category by changing the location (either rightmost or leftmost of the chain) of the three geons defining the high-level categories.

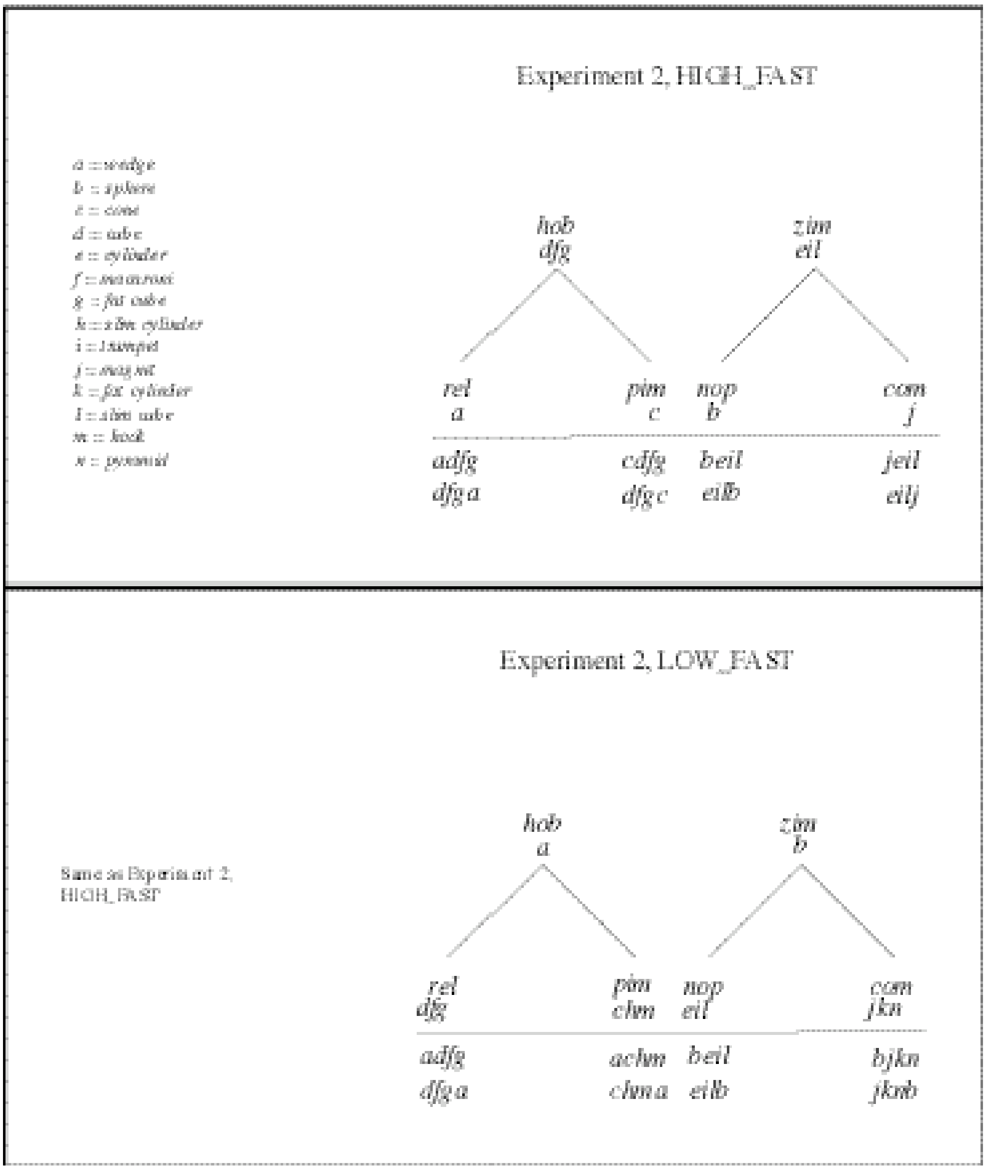


Figure 21. Taxonomies of Experiments 2, HIGH_FAST and LOW_FAST. Only internal practicability (redundancy) was manipulated here. Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy. An index for these abstract features is also provided.

The LOW_FAST taxonomy comprised fourteen diagnostic geons. A single diagnostic geon defined each one of two high-level categories, and three different geons further defined each one of four low-level categories. As before, we created two category exemplars by changing the location (either far right or far left of the object) of the triplets defining the low-level categories. Practicability is greater for high-level categories in HIGH_FAST and for the low-level categories in LOW_FAST. These levels have more unique features associated with them.

5.3.1.3 Procedure

The procedure followed in all respects that of Experiment 1: Participants were randomly assigned to the HIGH_FAST and LOW_FAST conditions. They were taught their respective taxonomy before entering a verification task where we measured speed of access to the two levels of categorisation. The experiment was ran with the SuperLab software on a Macintosh PowerPC 7200. Each one of 280 trials consisted in the initial presentation of a category name followed by an object. Participants had to decide as fast as they possibly could whether the two matched, and we recorded response latencies.

5.3.2 Results and discussion

We analysed only the correct positive trials RTs (error rate = 5.4%) within two standard deviations from the means (an additional 2.4% of the responses were discarded). Table 10 shows the mean RTs at the low and high-levels for HIGH_FAST and LOW_FAST.

A two-way (GROUP x PRACTICABILITY) ANOVA on the RTs with repeated measures on one factor (PRACTICABILITY) revealed a significant GROUP x PRACTICABILITY interaction, $F(1, 18) = 5.53$, $p < .05$, as well as two significant simple effects: GROUP(HIGH_FAST) by LEVEL, $F(1, 18) = 61.50$, $p < .001$ (only one subject responded faster for the high-level categories, $p < .011$–e.g., Siegel, 1956), and GROUP(LOW_FAST) by LEVEL, $F(1, 18) = 67.20$, $p < .001$ (two subjects responded faster for the high-level categories, $p < .055$). The error rate was low overall and not correlated with RTs ($r = .05$, *ns*), ruling out a speed-accuracy trade-off explanation.

Table 10: Mean RTs and standard deviations (between brackets) for the positive trials of experiment 2 as well as predictions for feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. For each taxonomy, the greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

		Level	
	Model	Lowest	Highest
Exp. 2, HIGH_FAST	Observation	788 ms [265]	660 ms [284]
	Possession	1	3
	Utility	.375	.500
	Compression	0 bit	5 bits
	Context	.769	.988
	SLIP	3.2 attempts	2.286 attempts
Exp. 2, LOW_FAST	Observation	740 ms [227]	774 ms [292]
	Possession	3	1
	Utility	.624	.500
	Compression	0 bit	5 bits
	Context	.974	.984
	SLIP	2.286 attempts	3.2 attempts

In sum, SLIP predicted that strategies with greater practicability should yield faster categorisation decisions, irrespective of categorisation level (see SLIP in Table 10 for numerical predictions). The results of Experiment 2 confirmed the prediction.

5.4 Experiment 2B

Experiment 2A suffers from the same generalisation limitation as Experiment 1. Experiment 2B was designed to overcome this. It replicates Experiment 2A using randomly generated letter strings instead of 3D computer-synthesised objects. In other words, ecological validity is sacrificed for control. Just to refresh your memory: All strategies had length 1 but the high and low levels differed in practicability. In the HIGH_FAST condition, high-level strategies had greater practicability than low-level strategies; the opposite applied to the LOW_FAST condition, with low-level strategies having higher practicability. SLIP predicts that categories with higher practicability will be verified faster, irrespective of their level in the taxonomy.

5.4.1 Method

5.4.1.1 Participants

Twenty four students from University of Glasgow with normal or corrected vision were paid to participate in the experiment.

5.4.1.2 Stimuli

Stimuli were 10-letter strings. The HIGH_FAST condition used 10 diagnostic letters. A three-letter sub-string defined each one of two high-level categories; and a one-letter sub-string further defined each low-level category. This is illustrated in the top taxonomy of Figure 21 (see 5.3 Experiment 2A). However, here "what you see is what you get": each letter represents itself or, at least, another letter. For each subject, the diagnostic letters were randomly selected. Each one of eight 10-letter templates was created by putting this three-letter sub-string randomly at one of the eight possible positions inside the 10-letter string and then by putting the one-letter sub-string at any remaining position except at the extremities (this ensured that the number of configurations was the same for three-letter and one-letter sub-strings). The other positions in the 10-letter strings were filled with nondiagnostic letters (there are 16 of those here) randomly selected. For example, **kamnopqdfg** and **sxmdfgqayk** are two possible *rels*–and *hobs*–in the top taxonomy of Figure 21.

The LOW_FAST condition involved fourteen diagnostic letters. A one-letter sub-string defined each one of two high-level categories, and three-letter sub-strings further defined each one of four low-level categories (see bottom taxonomy of Figure 21). As in the HIGH_FAST condition, each exemplar was generated by placing a three-letter sub-string randomly at one of the eight possible locations inside the 10-letter string and then by randomly putting an appropriate one-letter sub-string at any remaining position except at the endings. Randomly selected nondiagnostic letter fillers (there are 12 of those) occupied the 10-letter string template's empty slots.

Practicability is greater for high-level categories in the HIGH_FAST condition and for the low-level categories in the LOW_FAST condition because more unique features are associated with the top- and bottom-level categories, respectively. SLIP predicts a faster verification performance for categories with higher practicability (high in HIGH_FAST and low in LOW_FAST) irrespective of the level of the taxonomy considered.

5.4.1.3 Procedure

This experiment was controlled by a Silicon Graphics Computer running a home-made C program. The procedure followed that of Experiments 1 and 2A closely: Participants were randomly assigned to the HIGH_FAST and LOW_FAST conditions. They were taught their respective taxonomy before being measured on the categorisation speeds of its levels. Each one of 280 trials consisted of the initial presentation of a category name followed by an 10-letter string. Participants had to decide as fast as they possibly could whether the two matched and we recorded response latencies.

5.4.2 Results and discussion

We analysed only the correct positive trials RTs (error rate = 8.5%) within two standard deviations from the means (an additional 4.2% of the responses were discarded). Table 11 shows the mean RTs at the low and high-levels for the HIGH_FAST and for the LOW_FAST taxonomies.

Table 11: Mean RTs and standard deviations (between brackets) for the positive trials of Experiment 2B as well as predictions for feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. For each taxonomy, the greyshade indicates the order of Predicted or of observed basic-levelness, with the lightest being the greatest.

		Level	
	Model	Lowest	Highest
Exp. 2, HIGH_FAST	Observation	1113 ms [341]	1034 ms [324]
	Possession	1	9
	Utility	.375	.500
	Compression	0 bit	5 bits
	Context	.769	.988
	SLIP	14.225	11.641
Exp. 2, LOW_FAST	Observation	867 ms [144]	949 ms [164]
	Possession	3	7
	Utility	.624	.500
	Compression	0 bit	5 bits
	Context	.974	.984
	SLIP	11.641	14.225

A two-way (GROUP x PRACTICABILITY) ANOVA on the RTs with repeated measures on one factor (PRACTICABILITY) revealed a main effect of practicability, $F(1, 22) = 21.21$, $p < .001$ (mean verification time = 950 ms for high practicability strategies [standard deviation = 234 ms]; 1031 ms for low practicability strategies [standard deviation = 253 ms]). Out of 24 participants, only 6 did not respond faster to the greater practicability categories; a sign test showed that this is significant, $p < .0114$. Neither the GROUP x PRACTICABILITY interaction, $F(1, 22) = .01$, *ns*, nor the main GROUP effect, $F(1, 22) = 2.51$, *ns*, was significant. The error rate was low overall and was positively correlated with RT ($r = .53$, $p < .001$), ruling out a speed-accuracy trade-off.

In sum, Experiment 2B replicated Experiment 2A. This lends further support to our conclusion that internal practicability (and not accidental characteristics of a stimuli set) drives verification time.

5.5 Experiment 3

Experiments 1, 2A, and 2B revealed that the two determinants of SLIP (strategy length and internal practicability) can independently determine the basic-levelness of any level of a taxonomy. Experiment 3 explores how these two factors interact to determine performance. There are many possible interactions to investigate and we will not investigate them all. Instead, we will examine three scenarios that change the fastest level by modifying either strategy length or internal practicability.

EQUAL will be our neutral scenario. Strategies at the high and low-levels have an equal length of 1 and the same constant practicability. SLIP predicts that categorisation speeds should be equal across levels. In the SL_DOWN scenario, we will produce faster categorisations at the lower level by augmenting the length of the strategies that access the high-level categories. This scenario uses Experiment 1, LOW_FAST, taxonomy with a different set of geons. In the IP_UP scenario, we will keep the difference of strategy length just discussed, but the high-level will now be fastest because the practicability of the low level will be decreased. In sum starting from an EQUAL access to two levels of a taxonomy, a change of strategy length in SL_DOWN produces faster categorisations at the low level. From this SL_DOWN taxonomy, a decrease in the internal practicability of the low level in IP_UP produces faster categorisation at the high level.

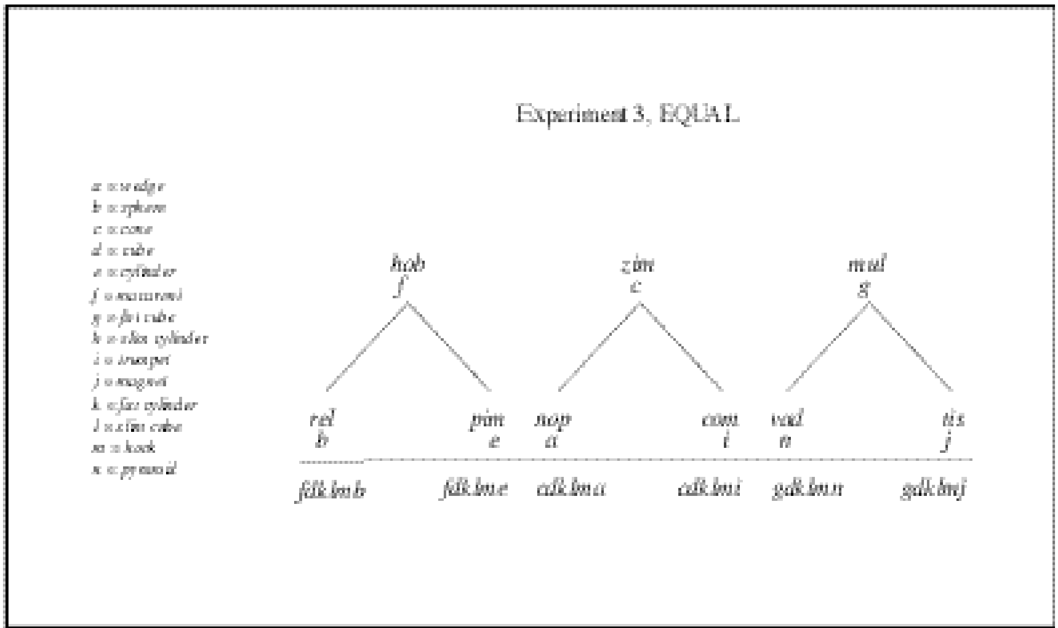
5.5.1 Method

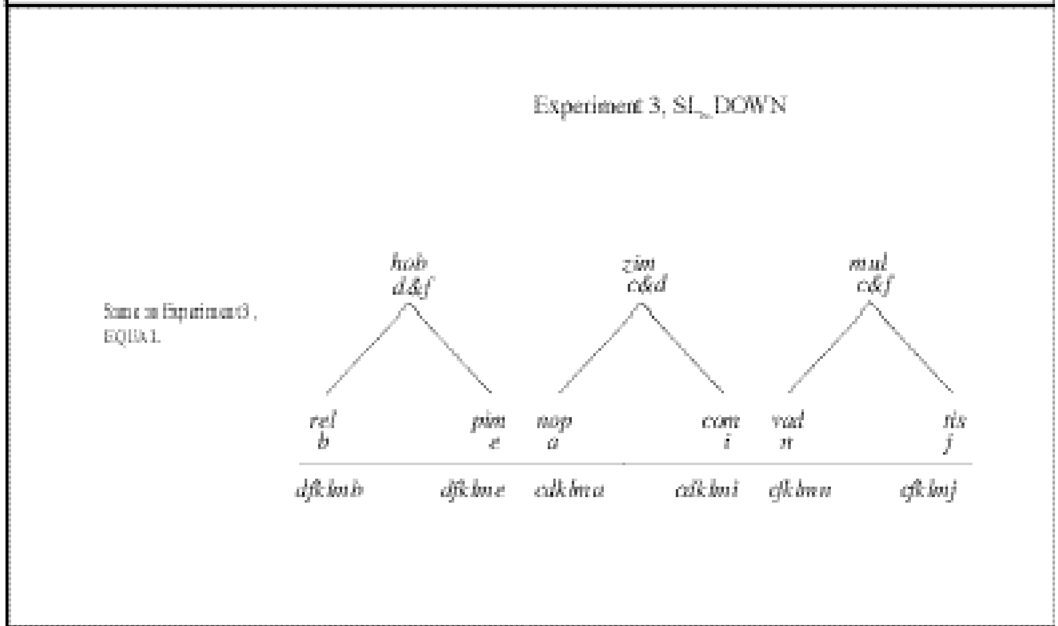
5.5.1.1 Participants

Thirty students from University of Glasgow with normal or corrected vision were paid to participate in the experiment.

5.5.1.2 Stimuli

Stimuli were similar to those of experiments 1 and 2A: geon chains designed with the Form Z 3D-object modelling software. Nine diagnostic geons entered the composition of categories in the EQUAL, SL_DOWN, and IP_UP conditions. In EQUAL, one geon defined each one of the nine categories of the taxonomy (see the top taxonomy of Figure 22).


Experiment 3, EQUAL
a = wedge
b = sphere
c = cone
d = cube
e = cylinder
f = macaroni
g = flat cube
h = slim cylinder
i = trumpet
j = magnet
k = fat cylinder
l = slim cube
m = hook
n = pyramid
hob
f
zim
c
mul
g
rel
b
pim
e
nop
a
com
i
vad
n
tis
j
fdklmb fdklme cdklma cdklmi gdklmn gdklmj


Experiment 3, SL_DOWN
Same as Experiment 3, EQUAL
hob
d&f
zim
c&d
mul
c&f
rel
b
pim
e
nop
a
com
i
vad
n
tis
j
dfklmb dfklme cdklma cdklmi cfklmn cfklmj

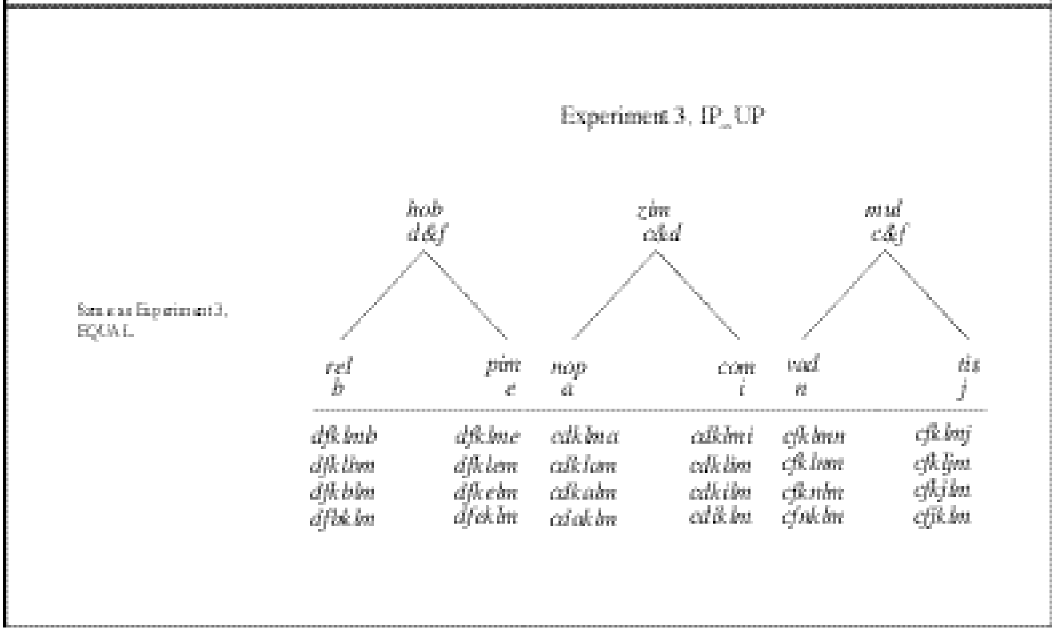

Experiment 3, IP_UP
Same as Experiment 3, EQUAL
hob
d&f
zim
c&d
mul
c&f
rel
b
pim
e
nop
a
com
i
vad
n
tis
j
dfklmb dfklme cdklma cdklmi cfklmn cfklmj
dfklbm dfklem cdklam cdklim cfklnm cfkljm
dfkblm dfkelm cdkalm cdkilm cfknlm cfkjlm
dfbklm dfeklm cdaklm cdiklm cfnklm cfjklm

Figure 22. Taxonomies of Experiment 3, EQUAL, SL_DOWN, and IP_UP (from top to bottom). Strategy length and internal practicability interacts here. Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy. An index for these abstract features is also provided.

To this defining geon, we added four fillers to form six six-geon chains (see Figure 23). We placed the geons defining the high-level categories at the far left of the chains, and those defining the low-level categories at the far right (see the top taxonomy of Figure 22).

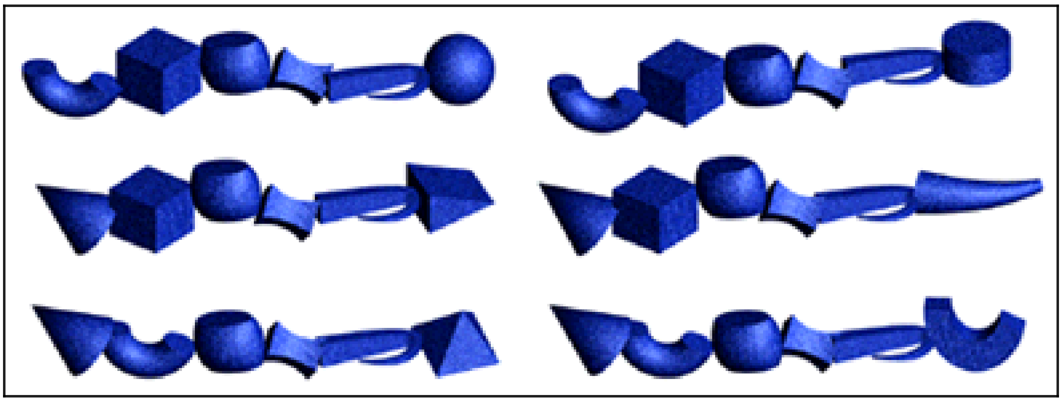


Figure 23. Computer-synthesised objects used in Experiment 3, EQUAL.

In SL_DOWN, a unique combination of two geons defined each top-level category. The addition of one different geon further specified each lower-level category. We produced six six-geon chains by adding three fillers. We placed the geon pairs defining the high-level categories at the far left of the chains, and those defining the low-level categories at the far right (see the middle taxonomy of Figure 22). These chains also served to construct the exemplars of IP_UP. Here, we generated four exemplars per category by changing only the location in the chain of the single geon defining the low-level categories (one of the four rightmost positions in the six-geon chains–see the bottom taxonomy of Figure 22).

5.5.1.3 Procedure

The procedure was identical to that of Experiments 1 and 2A. Participants were randomly assigned to one of three conditions (EQUAL, SL_DOWN, and IP_UP). Following a learning of their taxonomy, participants performed 240 verification trials. The experiment was run with the SuperLab software on a Macintosh PowerPC 7200. Each trial consisted in the presentation of a category name followed by an object. Participants had to decide whether these matched and we measured response latencies.

5.5.2 Results and discussion

We performed the analysis of RTs on the positive, correct trials (error rate = 2.3%) that were within two standard deviations from the means (4.6% of the trials were discarded). Table 12 shows the mean RTs. A two-way (GROUP x LEVEL) ANOVA with repeated measures on LEVEL revealed a significant interaction between GROUP and LEVEL, $F(2, 27) = 11.85$, $p < .001$, and two significant simple effects of GROUP(SL_DOWN) by LEVEL, $F(1, 27) = 10.58$, $p < .003$ (only one subject responded faster for the high-level categories, $p < .011$), GROUP(IP_UP) by LEVEL, $F(1, 27) = 13.09$, $p < .001$ (two subjects verified the low-level categories faster, $p < .066$). The last main effect is not significant: GROUP(EQUAL) by LEVEL, $F(1, 27) = .04$, *ns*. The error rate was low overall and was positively correlated with RT ($r = .31$, $p < .05$), ruling out a speed-accuracy trade-off.

SLIP predicted all the results observed in Experiment 3 (see SLIP in Table 12 for numerical predictions). Participants categorised equally fast at both levels in EQUAL. Increasing the strategy length of the higher level in SL_DOWN induced faster categorisations of the lower level. Note that Experiment 3, SL_DOWN, replicated Experiment 1, LOW_FAST, with a different set of geons. Diminishing practicability at the lower level in IP_UP then made the high level faster. The two computational factors of SLIP predicted speed of categorisation in these taxonomies.

Table 12: Mean RTs and standard deviations (between brackets) for the positive trials of Experiment 3 as well as predictions for feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. For each taxonomy, the greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

		Level	
	Model	Lowest	Highest
Exp. 3, EQUAL	Observation	672 ms [212]	680 ms [225]
	Possession	1	5
	Utility	.176	.260
	Compression	0 bit	30 bits
	Context	.769	.783
	SLIP	1.714 attempts	1.714 attempts
Exp. 3, SL_DOWN	Observation	920 ms [295]	1,058 ms [374]
	Possession	1	5
	Utility	.250	.333
	Compression	0 bit	30 bits
	Context	.769	.783
	SLIP	1.714 attempts	3.429 attempts
Exp. 3, IP_UP	Observation	928 ms [576]	775 ms [506]
	Possession	1	5
	Utility	.250	.333
	Compression	0 bit	30 bits
	Context	.769	.783
	SLIP	6.857 attempts	3.429 attempts

5.6 Experiment 4[20]

You will remember that when internal practicability is constant SLIP reduces to a Pascal density function with a mean equal to strategy length divided by this constant internal practicability (see section *2.3.2 The Pascal density function*). Put otherwise: SLIP predicts a positive linear relationship between strategy length and RTs. Unfortunately all basic-

[20] Experiment 4 was published in Gosselin and Schyns (1998a).

level experiments examining the effect of strategy length on basic-levelness have used only two strategy length values (Hoffmann & Ziessler, 1983; our Experiment 1), and a line can always pass perfectly through two points. Experiment 4 was designed primarily to test this prediction. Participants learned the category structure of Figure 24. In this category hierarchy, membership to different levels requires optimal strategies of different lengths: length 1 for high-level categories, length 2 for mid-level categories, and length 3 for low-level categories. Note also that the cardinality of the sets of redundant tests of these strategies is equal to 1 at all level. SLIP predicts that no matter what type of feature (i.e., *geon*, *colour*, or *texture*) is used the diagnostic structure of the categories will determine the basic level.

Another goal of Experiment 4 was to replicate Experiment 1, HIGH_FAST, with a three-level taxonomy (rather than two) and objects varying on three dimensions (rather than objects composed of features).

5.6.1 Method

5.6.1.1 Participants

Thirty students from University of Glasgow with normal or corrected vision were paid to participate in this experiment.

5.6.1.2 Stimuli

Our eight stimuli (see Figure 25) filled the whole space defined by three binary dimensions: geon (G) (either cylinder or pyramid), colour (C) (either red or green), and texture (T) (either smooth or rough). One binary dimension was added at every categorisation level of the category hierarchy of Figure 24. Objects were constructed with the Form Z three-dimensional modelling software on a Macintosh computer.

Forty-eight items were constructed. An item consisted of the presentation of a category name followed by an object. They were generated with the constraints that their number across levels of categorisation had to be equal, and that their number across categories at a given level of categorisation had to be equal. Moreover, the number of positive items (i.e., items for which participants had to respond "yes") was equal to the number of negative items (i.e., items for which participants had to respond "no") for any given category.

5.6.1.3 Design

The addition of binary dimensions at levels of categorisation was counterbalanced across participants using a Latin square to ensure that each binary dimension was added to the higher, middle, and lower levels of categorisation equally often. Thus, our three experimental conditions were: CTG, GCT, and TGC (the first, second, and third binary dimensions correspond to the higher, middle, and lower levels of categorisation, respectively).

5.6.1.4 Procedure

The procedure of this experiment followed closely that of Murphy (1991). The experiment was controlled by a Silicon Graphics computer running a home-made computer program written in C.

The experiment was divided into three phases: a learning phase, a test of learning phase, and a critical phase.

During the learning phase, participants had to learn the nonsense names of 14 categories, as well as their defining feature(s). For example, a participant from the CTG condition might have had to learn–among other things–that a "hob" was *green*, *rough*, and *cylindrical*, that a "zim" was *green* and *rough*, and that a "tis" was *green*. The nonsense names were randomly assigned to categories for each participant.

Participants were given their categories' defining feature hierarchy on a sheet of paper (see Figure 24). (Most participants took about one hour to finish the learning phase.)

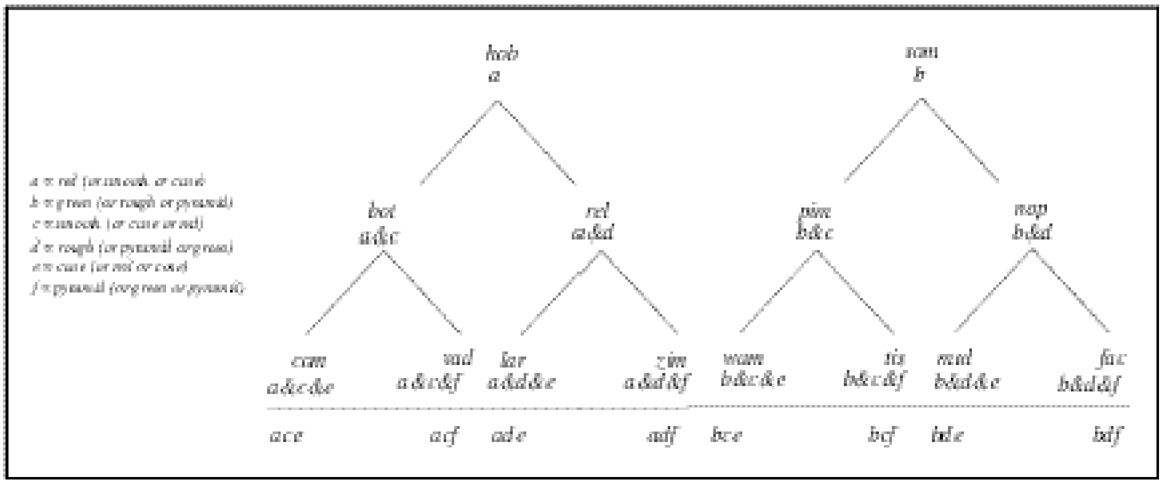


Figure 24. Taxonomies of Experiment 4, CTG, TGC, and GCT. Only strategy length varies here. Underneath the category names, we provide the optimal strategies fed to SLIP. The feature constitution of all exemplars is given underneath each taxonomy. An index for these abstract features is also provided.

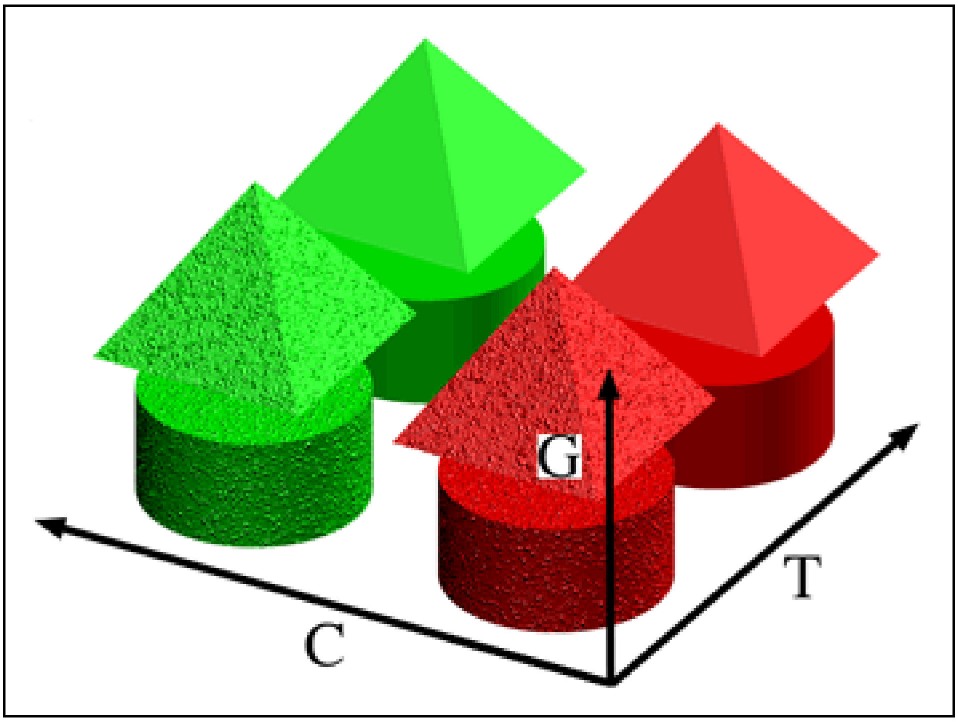


Figure 25. All computer-synthesised objects used in Experiment 4 with their three dimensions: C (*colour*), G (*geon*), and T (*texture*).

After participants reported having learned the categories, they were given a learning test. The experimenter asked them to give the defining features associated with every category name. The order in which the category names were given to every participant was randomly generated. Participants had to recall–without a single mistake–all the defining features of all the categories twice in a row. Corrective feedback was given. (Most participants took about ten minutes to complete the test of learning phase.)

During the critical phase of the experiment, 48 items were presented 5 times to participants. The order of the 48 items were randomly generated within each block. An item began with the presentation of a category name. Participants had to recall the defining feature(s) of the associated category. As soon as they had remembered the appropriate defining feature(s), they pressed one of two response keys, and, 200 ms later, an object was presented to them. They had to decide–as fast as they could without making too many mistakes–whether or not it was a member of the shown category by pressing on the "yes" key or the "no" key. Participants responded with their right and left indexes. Half of the participants pressed on the "yes" key with the left index. The time it took them to respond was recorded. No corrective feedback was provided. (Participants took approximately 25 minutes to complete the critical phase.)

5.6.2 Results and discussion

We only analysed positive items (50% of the items) correctly answered (5.7% of the responses were discarded) and within two standard deviations of the mean (an additional 0.05% of the responses were rejected). The means of the remaining RTs are shown in Table 13 by group and level of abtraction.

Table 13: Mean RTs and standard deviations (between brackets) for the positive trials of Experiment 4, CTG, GCT, and TGC. For each taxonomy, the greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

Group	Level		
	Lowest	Middle	Highest
CTG	1396 ms [421]	1169 ms [369]	861 ms [223]
GCT	1123 ms [526]	919 ms [262]	778 ms [242]
TGC	1034 ms [798]	948 ms [658]	818 ms [384]
Mean	1184 ms [582]	1012 ms [430]	819 ms [283]

We performed a two-factor (GROUP x LEVEL) ANOVA with repeated measurements on one factor (LEVEL): Neither the interaction between GROUP and LEVEL, $F(4, 54) = 0.89$, *ns.*, nor the main effect GROUP, $F(2, 27) = 1.07$, *ns.*, was significant. However, the main effect LEVEL was significant, $F(2, 54) = 12.93$, $p < .0001$. A regression test (e.g., Hays, 1988) on the main effect LEVEL revealed a significant linear component, $r = .999$, $F(1, 87) = 10.31$, $p < .01$, and no significant curvilinear component, $F(1, 87) = 0.02$, *ns.*. As predicted by SLIP, irrespective of the type of information, there was a positive linear relationship between response latencies and strategy length (see Table 14 for numerical predictions). The error rate was low overall and was positively correlated with RT ($r = .28$, $p < .001$), ruling out a speed accuracy trade-off explanation.

Table 14: Mean RTs for the positive trials of Experiment 4, overall, as well as predictions for feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. For each taxonomy, the greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

		Level		
Source	Model	H - 2	H - 1	Highest (H)
Exp. 4, Overall	Observation	1,184 ms	1,012 ms	819 ms
	Possession	1	1	3
	Utility	0.188	0.25	0.25
	Compression	0 bits	73 bits	149 bits
	Context	.769	.769	.769
	SLIP	4.5 attempts	3 attempts	1.5 attempts

5.7 Comparison of models of basic-levelness with respect to Experiments 1 to 4

This completes the presentation of all the verification experiments of this dissertation. We will now compare the performance of the various basic-level models at predicting the results of these experiments.

SLIP predicts all the qualitative results of these experiments. Category feature-possession is second best with 63% of the data explained, followed by utility and compression with 58%, and trailed by the context model with 37% (see Table 15). Monte Carlo simulations showed that only SLIP significantly outperforms the chance model ($p < .001$).

Table 15: Percentage of nominal data from Experiments 1 to 4 explained by feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. Only the scores flanked by a star are significantly above chance level ($p < .001$).

	Strategy length	Internal practicability	Mean
Possession	45%	88%	63%
Utility	36%	88%	58%
Compression	55%	63%	58%
Context	18%	63%	37%
SLIP	100% *	100% *	100% *

The decomposition of these global scores into strategy length and internal practicability scores is: For the conditions testing only strategy length (Experiment 1, HIGH_FAST and LOW_FAST, as well as Experiment

3, EQUAL and SL_DOWN, and Experiment 4, Overall), the compression measure predicts 55% of the data, category feature-possession, 45%, category utility, 36%, and the context model, 18%. Again Monte Carlo simulations revealed that only SLIP significantly outperforms the chance model ($p < .001$). This confirms the argument made earlier that all models have so far neglected strategy length as a specific factor of basic level performance. This is a serious problem because attributes do overlap between categories in the real-world, and so strategy length is an important factor of recognition outside the laboratory.

For the conditions testing only practicability (Experiment 2, HIGH_FAST and LOW_FAST, and Experiment 3, EQUAL and IP_UP), category feature-possession and category utility predicts 88% of the data, compression measure and context model, 63% (see Table 15). (Note that we have included Experiment 3, EQUAL, in the break down into strategy length and internal practicability; it is an extreme case of both.) Monte Carlo simulations demonstrated that SLIP, category feature-possession and category utility significantly outperform the chance model ($p < .001$). This corroborates our earlier claim regarding category feature-possession and category utility taking the internal practicability into account.

5.8 Implications for part-based basic-level accounts

You will remember that in most recognition theories objects are first identified at the basic level from their parts and then they are identified at other levels (e.g., Biederman's, 1987, recognition-by-component theory; see Chapter 3). We will say here that these theories postulate a *strong Hardwired Bias for parts*. However, the observed correlation between proliferation of parts and basic-levelness (Tversky & Hemenway, 1984) could result from a *Contingent Diagnosticity of parts* for the task. This is what most formal theories of basic-levelness predict, including SLIP.

Murphy (1991a, b) tested these two rival hypotheses with artificial category hierarchies. He found that basic-levelness was a function of the structure of information in these taxonomies rather than their content–corroborating Contingent Diagnosticity and falsifying Hardwired Bias. More specifically, Murphy tried to show that parts were neither *necessary* (parts are necessary if, when a taxonomy does not have parts collected at one level, it will not display basic-level phenomenon), nor *sufficient* (parts are sufficient if, when a taxonomy has parts collected at one level, that level will tend to display basic-level phenomena) for basic-level performance. He has been criticised for having used unnatural objects (Tversky & Hemenway, 1991). Our Experiments 1, 2A, 3, and 4 used realistic objects and also gave support to the Contingent Diagnosticity hypothesis. Experiments 1, 2A, and 3 showed that parts were not *sufficient*: their taxonomies all had parts collected at all levels and these did not display equal basic-levelness (except for Experiment, 3, EQUAL). Experiment 4 showed that parts are not *necessary*: both in the TGC and the CTG taxonomies, levels defined by non-part information had the greatest basic-levelness.

5.9 Is basic-levelness really influenced by taxonomies?

An intriguing observation is that we could rewrite this dissertation disregarding completely the notion of taxonomy, only considering the computational factors of SLIP (this will, in fact, be exploited in Experiments 5A and 5B). Features can be redundant within a category and overlapping between categories without the need of a hierarchy to be explicitly represented in the memory of the categoriser. The categoriser would only need to explicitly represent its categories in memory, without an explicit representation of the hierarchical dependencies. SLIP would make exactly the same predictions, whether or not hierarchical dependencies are explicitly represented. This is because the relationship between the computational principles of SLIP and taxonomic knowledge is asymmetrical: realistic taxonomies could hardly exist without feature overlap and redundancy, whereas the principles of SLIP do not imply an explicit representation of taxonomic knowledge. Does the taxonomy itself influence performance in real world categorisations? At this stage, we do not have the elements of an answer, but it is clear that this issue deserves further considerations. It is brought into sharp focus when principles of knowledge organisation that subsume hierarchies can explain data that are supposed to arise from a hierarchical organisation of knowledge.

5.10 Introduction to Experiments 5A, 5B, and 6

So far we have shown that, in a verification tasks, strategy and internal practicability could–alone and together–determine basic-levelness. But we have stressed in the Preamble and Chapter 1 the importance of using multiple empirical basic-level indexes of performance; basic-levelness is a global measure of performance. The main objective of Experiment 5A, 5B, and 6 is to demonstrate that SLIP's predictive power extends beyond verification to naming. After verification, naming is the most commonly used index of performance. Experiment 5A and 5B isolates the effect of strategy length and internal practicability, respectively, whereas Experiment 6 examines the effect of strategy length on the time course of categorisation.

5.11 Experiment 5A

You will remember that strategy length is the minimum number of required tests on features to access a category. Experiment 5A was designed to isolates its effect in a two-alternative-forced-choice naming task. Unlike in Experiment 1, we did not use two two-level taxonomies to dissociate level of abstraction and strategy length; instead we used the simplified category structure illustrated in Figure 26. It has no taxonomic organisation: it is made of two partly overlapping categories with different strategy lengths. But as we have pointed out above (see section *5.9 Is basic-levelness really influenced by taxonomies?*), taxonomies do not play any particular role in our SLIP framework. Only strategy length and internal practicability matter.

Very few basic-level experiments have studied naming performance with artificial taxonomies. The only exceptions are Murphy and Smith (1982) and Hoffmann and Ziessler (1982); both used naming speed as their basic-levelness measure. Naming speed is arguably very similar to verification time. A more natural naming basic-levelness measure is frequency of use (this is Brown's original proposal). Experiment 5A estimates both these naming indexes of performance, and SLIP predicts both of them (see section *5.11.1.4 Predictions*).

5.11.1 Method

5.11.1.1 Subjects

Ten paid University of Glasgow students with normal or corrected vision participated in this experiment.

5.11.1.2 Stimuli

Objects were designed with the Form Z three-dimensional object modelling software on a Macintosh computer. We used a total of five four-geon sets. The *JON* (or *BOB*) category was defined by the conjunction of features *c* and *d*, and the *BOB* (or *JON*) category by the unique feature *a*. One filler was added to all geon sets (i.e., feature *b*). Another filler (i.e., *e* for half the UNMAMB_SL1 and UNMAMB_SL2 conditions, and *f* for the remaining) was added to each UNAMB geon set. Furthermore, half the UNAMB_SL1 sets possessed feature *c*, and the other half feature *d* (see Figure 26). From each four-geon set, two exemplars similar to those used in Experiments 1, 2A, and 3 (see Figures 19 and 23) were extracted. It is important to realise that AMB objects satisfy both the JON and BOB definitions.

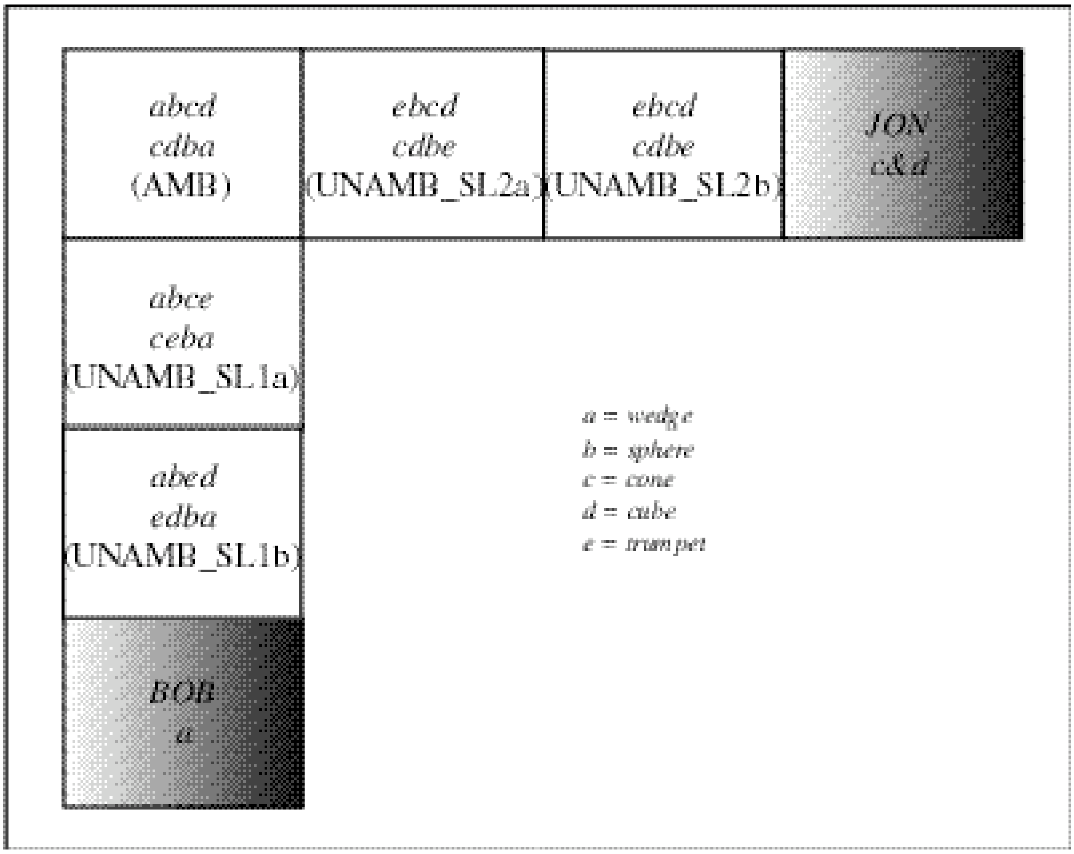


Figure 26. Only strategy length (SL) varies in Experiment 5A. In the dark boxes, underneath the category names, we provide the optimal strategies fed to SLIP. In the light boxes, we give the feature structures of the exemplars of the different experimental conditions (UNAMBiguous and AMBiguous). An index for these abstract features is also provided.

5.11.1.3 Procedure

During a learning session, participants were shown the defining features of *BOB*s and of *JON*s. Half the participants learned to associate the length 1 strategy with the name "BOB" and the length 2 strategy with "JON"; half learned the other associations. We verified that subjects had correctly learned the categories by asking them to give the defining features of *JON*s and *BOB*s.

The experiment ran on a Macintosh 7500 PowerPC and used the SuperLab experiment software.

Participants were instructed that during the testing phase, they would sometimes be presented unambiguous *JON*s (either UNAMB_SL2, or UNAMB_SL1) and *BOB*s (either UNAMB_SL1, or UNAMB_SL2) and sometimes ambiguous objects (AMB). For the unambiguous objects, their task was to name them as quickly as possible without making too many mistakes. For the critical ambiguous objects, their task was to give the first name that poped to their mind (note that mistakes are impossible in this case). Subjects responded by pressing one of two keys on the computer.

Subjects were submitted to a total of 120 UNAMB items (each individual exemplar was presented 15 times), half of which were *BOB*s; they were also submitted to just as many AMB items. The whole experiment lasted less than 30 minutes.

5.11.1.4 Predictions

In the last chapters we have given several examples of how to compute speed of access predictions in our framework (the only difference between verification and naming being the value of S; see section *2.3.4 Naming*). All items confounded, SLIP predicts that length 1 names should be associated with shorter response times than length 2 strategies. Exact predictions are shown in Table 16.

SLIP also predicts that the names associated with length 1 strategies will be used 75% of the time for the AMB objects. This expected percentage was derived by first computing the internal practicability of the sets of redundant features: = .375 (i.e., $C_j(1\text{-}S\text{+}SR_j)$ = .5*(1-1+1*.75) = .375). Then by applying the scheme described in section *2.3.4 Naming*: i.e., .375 / [.375 + .375] + {.375 / [.375 + .375]}*{.375 / [.375 + .375]} = .75.

5.11.2 Results and discussion

Subjects chose the length 1 names 72.5% of the time in AMB items (see Table 16). The difference between the subjects' preference for length 1 and length 2 categories is significantly different from zero, $Z = 8.091$, $p < .001$ (Hogg & Tanis, 1988). So it appears that length 1 names are chosen more often than length 2 ones in naming. Furthermore, the difference between this observed preference for length 1 strategies (i.e., 72.5%) and SLIP's prediction (i.e., 75%) is not significantly different from zero, $Z = 0.632$, *ns*. This supports SLIP's predictions concerning percentage of use.

The mean RTs for length 1 and length 2 names–AMB and UNAMB conditions confounded–are 814 ms and 1056 ms, respectively (see Table 16). The difference between these two mean latencies is significant, $t(10) = 1.90$, $p < .029$. In fact, nine subjects out of 10 responded faster in SL1 than SL2 cases, $p < .011$. As predicted by SLIP, it thus takes less time to name a category associated with a length 1 strategy than one associated with a length 2 strategy.

Table 16: Percentage of times the name associated with the smallest strategy length was chosen in the ambiguous cases with standard deviations (between brackets) and speed of access of all correct cases confounded with standard deviations (between brackets) as well as predictions for feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), and SLIP. The greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

Model	Measure	Strategy Length	
		1	2
Observation	Preference	73% [32%]	27% [32%]
	RT	814 ms [426]	1055 ms [550]
Possession	Preference	2	3
	RT	idem	Idem
Utility	Preference	.053	.717
	RT	Idem	idem
Compression	Preference	?	?
	RT	idem	Idem
Context	Preference	?	?
	RT	idem	idem
SLIP	Preference	75%	25%
	RT	2.667 attempts	5.333 attempts

In the UNAMB cases, subjects made significantly more errors for length 2 categories (26.6%) than for length 1 categories (2.8%), $Z = 3.542$, $p < .001$[21]. This is consistent with the results of Experiments 1 and 4 that both revealed positive correlations between strategy length and errors.

The data reported here confirms again that strategy length determines the basic-levelness of a category.

5.12 Experiment 5B

You will remember that internal practicability, the second determinant of SLIP, refers to the ease with which feature sets identify a category. A category has high practicability whenever many of its defining features are uniquely diagnostic of this category. Experiment 5B was designed to isolate internal practicability in a two-alternative-forced-choice naming task. Instead of a full-blown taxonomy, we used two partly overlapping categories with different practicabilities (see Figure 27).

5.12.1 Method

5.12.1.1 Subjects

Ten paid University of Glasgow students with normal or corrected vision participated in this experiment.

5.12.1.2 Stimuli

The objects were designed with the Form Z three-dimensional object modelling software on a Macintosh computer. We used a total of three four-geon sets. The *JON* (or *BOB*) category was defined by the redundant set of features *b*, *c*, and *d*, and the *BOB* (or *JON*) category by

[21] The experiment-wise error level can be computed using the following formula: $1 \quad (1 \quad _i)$, where i spans all individual error levels. Thus the overall error is equal to 1 - [(1 - .001) * (1 - .029) * (1 - .001) * (1 - .001)] = .042 in Experiment 5A.

feature *a*. One filler was added to the UNAMB sets (i.e., feature *e*). The UNAMB_IP1 exemplars possessed two more fillers (i.e., *f* and *g*) (see Figure 27). Note that the AMB objects satisfy both the *JON* and *BOB* strategies. From these geon sets two examplars similar to those used in Experiments 1, 2A, 3, and 5A were extracted.

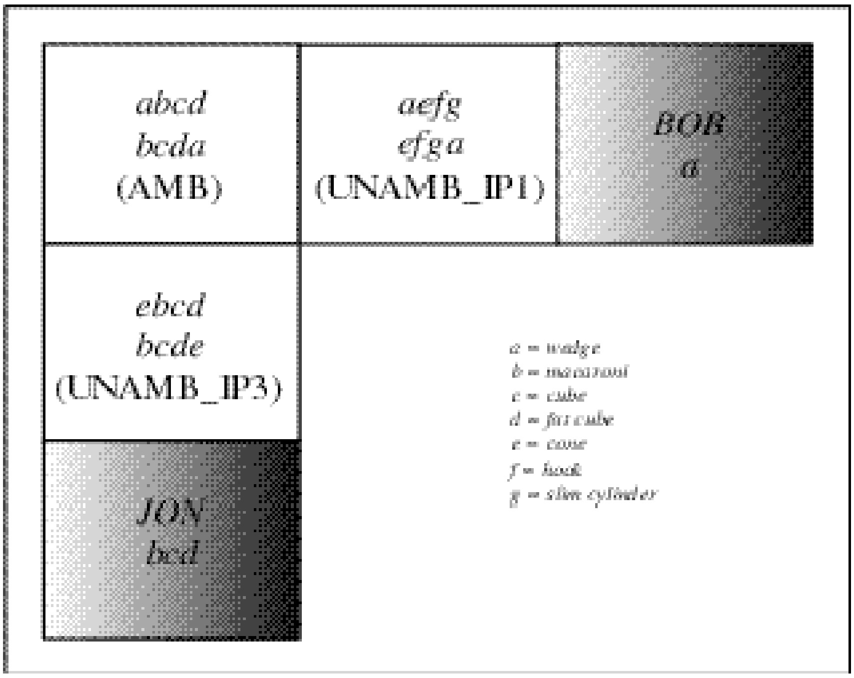


Figure 27. Only internal practicability (IP) varies in Experiment 5B. In the dark boxes, underneath the category names, we provide the optimal strategies fed to SLIP. In the light boxes, we give the feature structures of the exemplars of the different experimental conditions (UNAMBiguous and AMBiguous). An index for these abstract features is also provided.

5.12.1.3 Procedure

The procedure was identical to that of Experiment 5a. During a learning session, participants were shown the defining features of *BOB*s and of *JON*s. This is illustrated in Figure 27. We made sure that the subjects knew the definitions.

The experiment ran on a Macintosh 7500 PowerPC and used the SuperLab experiment software.

Participants were instructed that during the testing phase, they would sometimes be presented unambiguous *JON*s (either UNAMB_IP3, or UNAMB_IP1) and *BOB*s (either UNAMB_IP1, or UNAMB_IP3) as well as ambiguous objects (AMB). For the UNAMB objects, they would have to name them as quickly as possible without making too many mistakes; for the critical AMB objects, they would have to give the first name that comes to their mind.

Subjects were presented 80 UNAMB object items (each individual exemplar was presented 20 times), half of which were BOBs; they were also submitted to just as many AMB items. The whole experiment lasted less than 25 minutes.

5.12.1.4 Predictions

SLIP predicts that length 1 names should be associated with shorter response times than length 2 strategies all items confounded (see Figure 17 for exact predictions).

It also predicts that the names associated with high practicability will be used 75% of the time in the AMB items. To derive this percentage the categories' internal practicabilities have to be computed: the high practicability categories have a = .375 (i.e., $C_j(1\text{-}S+SR_j)$ = .5*(1-1+1*.75) = .375) and the low practicability ones a = .125 (i.e., $C_j(1\text{-}S+SR_j)$ = .5*(1-1+1*.25) = .125). Then Equation 5 is applied (see 2.3.4 Naming), that is $\frac{X}{\quad_i \; _i}$ = [.375 / (.125 + .375)] = .75.

5.12.2 Results and discussion

For the AMB items, participants chose the most redundant category 76.6% of the time (see Table 17). The difference between their preference for the most redundant and least redundant names is significantly different from zero, $Z = 9.525$, $p < .001$. Thus it seems that highly practicable names are used more often than less practicable ones in naming. The difference between this observed preference for highly practicable strategies (i.e., 76.6%) and SLIP's prediction (i.e., 75%) is not significantly different from zero, $Z = 0.330$, *ns.*. This confirms SLIP's frequency of use prediction.

The mean RT for more practicable and less practicable names, AMB and UNAMB conditions confounded, are, respectively, 619 ms and 695 ms (see Table 17). The difference between these two mean latencies goes in the expected direction but is not significant, $t(10) = 1.44$, *ns.*. This is probably due to a large variance in the denomination data, and more subjects would probably reveal a significant difference. In fact, nine subjects out of 10 responded faster in IP3 cases than in the IP1 cases, $p < .011$. Based on our sign test and the direction of the means' difference, we conclude with some confidence that it takes less time to name a category associated with high practicability than one associated with low practicability.

Table 17: Percentages of times the name associated with the largest internal practicability was chosen in the ambiguous cases with standard deviations (between brackets) and speed of access of all correct cases confounded with standard deviations (between brackets) as well as predictions for feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), and SLIP. The greyshade indicates the order of predicted or of observed basic-levelness, with the lightest being the greatest.

Model	Measure	Internal Practicability	
		3	1
Observation	Preference	77% [16 %]	23% [16 %]
	RT	620 ms [285]	695 ms [302]
Possession	Preference	4	4
	RT	idem	Idem
Utility	Preference	.703	.039
	RT	Idem	idem
Compression	Preference	?	?
	RT	idem	Idem
Context	Preference	?	?
	RT	idem	idem
SLIP	Preference	75%	25%
	RT	2.667 attempts	8 attempts

In the unambiguous cases, subjects made few errors (error for most redundant items was 2.75% and for least redundant items 1.51%). The difference between theses values is not significant, $Z = 0.710$, *ns.*.

The data reported here confirms once more that internal practicability can determine the basic-levelness of a category.

5.13 Experiment 6

We have shown repeatedly now that strategy length determines basic-levelness. We have done so in verification tasks (Experiments 1, 3, SL_DOWN, and 4) as well as in naming (Experiment 5A). However, all these experiments looked at final products of categorisation, i.e. category speed of access or frequency of use. SLIP authorises a more precise study of the time course of categorisation. Experiment 6 was designed to investigate this. We looked at what happens when the visual input is available for processing for various durations in a two/four-alternative-forced-choice naming task. A cue appearing before the scene to be named indicated which level of categorisation to use: when it was "low" participants had to choose between two names, when it was "high" they had to choose between four.

We constructed four stimuli by combining two different luminance patterns (that we call *flat* and *hilly*) with two different chromatic patterns (labelled *grassy* and *sandy*). Subjects learned to categorise these scenes in a two-level taxonomy (see Figure 28). At the general level, they learned to separate the four scenes into "flat" and "hilly", on the basis of luminance cues. At a specific level, they learned to categorise the stimuli as either "field" (the combination of *flat* and *grassy*), "desert" (*flat* and *sandy*), "mountain" (*hilly* and *grassy*), or "dune" (*hilly* and *sandy*).

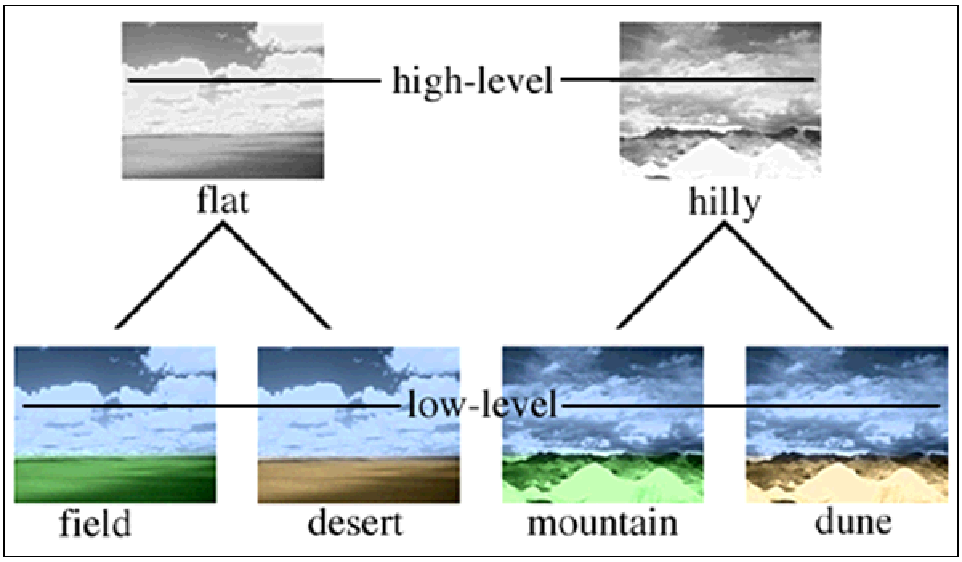


Figure 28. The four scenes used in this experiment and the corresponding general ("flat" and "hilly") and specific-level category names ("field", "mountain", "desert", and "dune") learned by all participants.

Suppose that the field picture is briefly presented on the screen, immediately followed by a mask. Subjects can make two types of responses at the general level: correctly respond "flat" (0 error, high) and respond "hilly" rather than "flat" (1 error, high), implying a misperception of the luminance information (see top of Figure 29). At the specific-level participants can make three types of response: (1) correctly respond "field" (0 error, low), respond "dune", implying a misperception of both the flat luminance and the green chromaticity of the field (2 errors, low), and (3) respond "mountain" (or "desert"), implying a misperception of only the flat luminance (or the colour information) (1 error, low) (see bottom of Figure 29).

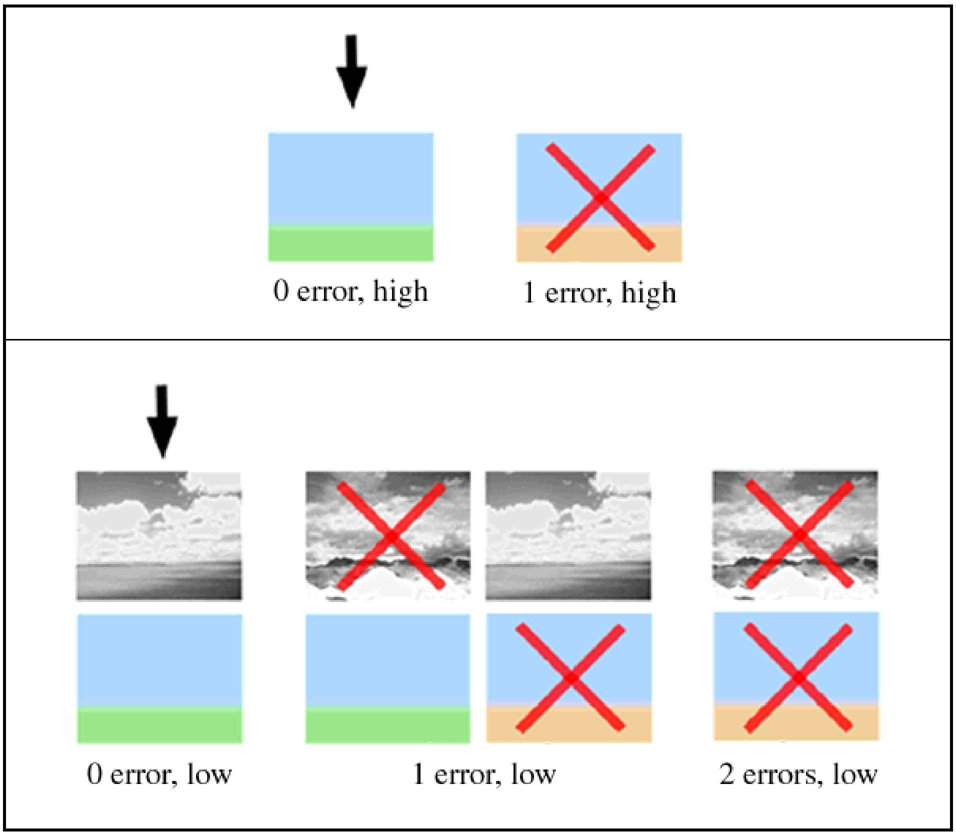


Figure 29. An illustration of the analysis of categorisation responses when a field is required to be named either at the high or low-level of abstraction.

5.13.1 Time-course predictions

SLIP makes explicit predictions concerning how often each type of these responses just reviewed should occur throughout time. For the general-level identifications there is a single way to make an error (1 error, high): no feature test has been successfully completed and the category has been guessed incorrectly. The probability of the latter is .5 because there are two high-level categories; and the probability that no feature has not been completed is one minus the probability that it has. Now, the probability that one feature test has been successfully completed at attempt *t* or before is given by the cumulative form of Equation 4 (see Chapter 2), i.e. $\overset{t}{\underset{i=n}{}} (1 \quad)^{i \;\; n} \quad {}^{n}$ with n = 1. Let us call this Cumulative distribution for Strategies of Length 1, *csl1* (by extension: *csl2*, *csl3*, …, and *csln*). It follows that the predicted distribution of 1 error, high, is .5 * (1 - *csl1*).

A no mistake verdict at the general-level (0 error, high) can result from two independent events: (1) the relevant feature test is completed and (2) the relevant feature test is not completed but a correct guess is made. The probability of (1) is given by *csl1* as we have seen. As for the probability of (2), it is equal to that of 1 error, high because the probability of making a correct guess is the same as the probability of making a wrong guess. Thus, distribution of 0 error, high, is *csl1* + .5 * (1 - *csl1*) = 1.5 * *csl1* + .5.

Let us now turn to low-level categorisation time course. Two errors, low, can only be due to having completed no relevant feature test and having guessed them all incorrectly. The probability of this last event is .25 because there are four low-level categories. And the distribution of no successful feature test for length 2 strategies is one minus the distribution of two completed feature tests minus the distribution of one–but not two–completed feature test, i.e., [1 - *csl2* - (*csl1* - *csl2*)] = (1 - *csl1*). So the distribution of 2 errors, low, is .25 * (1 - *csl1*). There are two ways a 1 error, low, can occur: either no feature test is completed (1 - *csl1*) and one dimension is correctly guessed (.5), or one feature test is completed but not two (*csl1* - *csl2*) and the other feature is wrongly guessed (.5). Thus the distribution of 1 error, low, is given by .5 * (1 - *csl1*) + .5 * (*csl1* - *csl2*) = .5 * (1 - *csl2*). Finally, three routes can lead one to make a correct low-level categorisation: (1) one could have completed the two diagnostic feature tests (*csl2*), (2) one could have completed one feature test, but not two, (*csl1* - *csl2*) and have guessed the other one correctly (.5), and (3) one could have completed no relevant feature test (1 - *csl1*) and have guessed both features correctly (.25). So SLIP predicts the following distribution of 0 error, low: $csl2 + .5 * (csl1 - csl2) + .25 * (1 - csl1) = .5 * csl2 + .25 * csl1 + .25$.

5.13.2 Method

5.13.2.1 Participants

Twenty University of Glasgow students with normal or corrected vision were paid to participate in the experiment.

5.13.2.2 Stimuli

We synthesised four distinct 450 x 350 pixels (spanning 7 per 5.4 deg) stimuli–a field, a desert, a mountain, and a dune–with the Photoshop image processing software by combining two luminance patterns with two chromatic patterns. The luminance patterns were extracted from a field (called *flat* here) and a dune (*hilly* here) photographs from the Corel Draw Photo Database; they were normalised for size and horizon level. The chrominance patterns were composed of two coloured rectangles corresponding roughly to the ground and the sky. The sky was the same blue, and the ground either green (called *grassy*) or yellow (called *sandy*). To eliminate the sharp boundary edge between the two coloured rectangles, we low-passed the patterns. A mask was created by randomly assigning to each square of a 18 x 14 grid the content of the corresponding region of one of the four scenes (e.g., Breitmeyer, 1989).

5.13.2.3 Procedure

The experiment ran on a Macintosh Power PC 7200 using a home-made program written with the Psychophysics Toolbox for MatLab (Brainard & Pelli, 1998). The subjects learned the name of the four stimuli at the specific level of a taxonomy: "field," "mountain," "desert" and "dune"; and also learned to categorise the stimuli at a general taxonomic level into "flat" vs. "hilly," on the basis of luminance cues (see Figure 28).

A learning block was completed when participants had named consecutively–and without mistake–all scenes at the high and at the low levels of categorisation (4 scenes x 2 levels of abstraction = 8 trials minimum). Within one learning block, trials followed each other in random order. A trial began with the display of the word "high" or "low," instructing subjects of the level at which the subsequent scene had to be named. Subjects then pressed a key to display the scene to categorise (presented on the screen for 1 s) immediately followed by a 450 ms mask. Subjects indicated their categorisation using one of six response keys (two for the general level, four for the specific level) before moving on to the next trial. Corrective feedback was provided.

When participants reached criterion (all subjects reached criterion after having been exposed to the minimum number of items), they were transferred to a testing phase including trials differing in two ways from those described above: First, presentation time varied being either 15, 30, 45, 60, or 75 ms. Half of the 600 test trials (4 scenes x 2 levels x 5 presentation times x 15 repetitions presented in a random order) started with the "low" cue and the other half with the "high" cue. Second, no corrective feedback was given.

5.13.3 Results and discussion

Figure 30 shows the effect of presentation time on the average proportion of responses with standard deviations. At 15 ms exposure, performance is near chance. We bestfitted linearly the predictions derived from the section *5.13.1 Time-course predictions* equations with $S = 1$ (see solid lines in Figure 30). Overall $R^2 = .98$. A chi-square goodness of fit test did not reveal any significant difference between the observed and predicted proportions ($^{2}_{obs}(24) = 7.02$, *ns.*) (e.g., Hogg & Tanis, 1988).

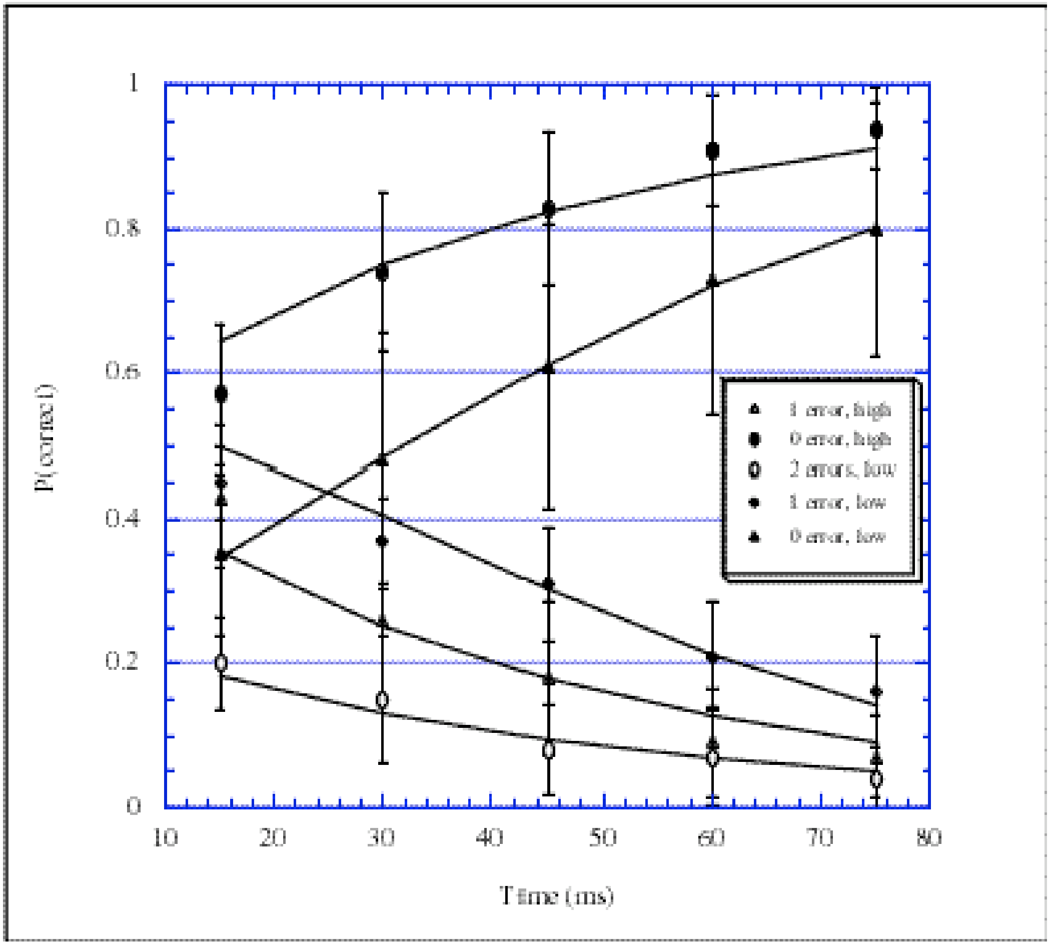


Figure 30. Average percentages of 0 error, and 1 error, high, and 0 error, 1 error, and 2 errors, low, with standard deviations and SLIP bestfits (solid lines).

The other models of basic-levelness cannot make such time course predictions, at least in their current state. However they can predict whether or not the overall speed of access for high-level categories (i.e., average speed of access for 0 error, high) will be faster than that for the low-level categories (i.e., average speed of access for 0 error, low). We do not have these speeds of access here, but it can be argued that they are inversely related to our average percentages of 0 error, high and low: The faster the strategy associated with a category can be completed, the more often it will be completed in a masked situation. On average participants made no error 80% of the time for high-level categories and 59% for low-level ones. The difference between these averages is significantly different from zero, $Z = 37.23$, $p < .001$. Table 18 shows the predictions of the various models. (For SLIP's predictions, we averaged the points of the bestfitted curves.)

Table 18: Average percentage of correct responses in function of strategy length as well as predictions by feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP.

	Average 0 Error	
Model	High	Low
Observation	80%	59%
Possession	1	3
Utility	.25	.25
Compression	1.678	3.192
Context	.769	.769
SLIP	80%	59%

It is unclear what information exactly is masked in our experiment (see Breitmeyer, 1989). Is it the availability of all the visual information which is terminated by the appearance of the mask? Or is it only the low-level visual attributes which are affected, leaving the higher-level visual information relatively unaltered? It would be worth replicating this experiment with a categorisation under-pressure task (e.g., Lamberts, 1995; McElree, Dolan, and Jacoby, 1999). Here, participants learn to respond within 200-300 ms of the onset of an auditory cue. Both the duration availability of visual input and that of processing is controlled. SLIP predicts the same pattern of errors.

5.14 Comparison of models of basic-levelness with respect to Experiments 5A, 5B, and 6

We have presented the three basic-level naming experiments of this dissertation which can be used to compare the performance of the various basic-level models. We will now proceed with this comparison.

SLIP predicts all the nominal data. Category utility follows with 83% (i.e., 5 out of 6) of the data explained, then comes category feature-possession with 17% (i.e., 1 out of 6). The context model and the compression measure cannot make any prediction for Experiments 5A and 5B. In Experiment 6, they explain respectively, 50% (i.e., 1 out of 2) and 0% of the data. Due to the scarcity of data, no Monte Carlo simulation was performed here. This pretty much corroborates our previous assessments of the models: SLIP leads the pack, followed by the category feature-possession and category utility pair, and trailed by the context model and compression measure pair.

The inability of the context model and the compression measure to model Experiments 5A and 5B illustrates a fundamental difference between these two models and the others. The context model requires at least two embedded categories–a minimum taxonomy–to be applied. You will remember that the most inclusive serves as a "standard" for the other. And at least two such embedded categories are needed for a score comparison. Experiments 5A and 5B each used two partly overlapping categories which is insufficient. The compression measure requires at least one partitioning of objects into two independent categories to be applied at all. We do not have this in Experiments 5A and 5B. And two such partitionings are needed for a minimum score comparison. In other words, these two models require taxonomies whereas the other models do not. We have discussed the implications of this in section *5.9 Is basic-levelness really influenced by taxonomies?*.

5.15 Experiment 7[22]

A perceptual prediction of SLIP is that strategies specify an order of feature testing (see Chapter 2, especially section *2.4 SLIP: a special diagnostic recognition model*). If this order is respected, then the perceptual appearance of the stimulus could change. To illustrate, a *rel* in the top taxonomy of Figure 18 (see Experiment 1) is *optimally represented* either by $Strat(X, rel) = [\{wedge\} \& \{cube\}]$, or by $Strat(X, rel) = [\{cube\} \& \{wedge\}]$. These two strategies have equal speed of access, but the order in which the two features are tested differs. Why would one adopt the first or the second strategy? In the top taxonomy of Figure 18, one strategy (i.e., $Strat(X, rel) = [\{cube\} \& \{wedge\}]$) is more *robust* in categorisation under time pressure. It is more robust because it is more likely to lead to a valid, if approximate, categorisation of the input. We know that the input is at least a *hob* if it has a *cube. Robustness* is critical in everyday categorisation: unseen features can be inferred on the basis of a categorisation (e.g., Anderson, 1991; Rosch, 1978). Experiment 7 was designed to investigate

[22] Experiment 7 was submitted as Gosselin and Schyns (2000).

this. It stands alone among the other experiments presented in this dissertation.

We used the four stimuli synthesised for Experiment 6. A learning procedure was devised to induce a different, two-level taxonomic knowledge of these four stimuli in two subject groups (labelled LUMI and CHRO, see Figure 31). At the high level, LUMI subjects learned to separate the four scenes into "flat" and "hilly", on the basis of luminance cues, whereas CHRO subjects learned to separate the same scenes into "grassy" and "sandy" on the basis of chromatic cues. At the low level, LUMI and CHRO subjects all learned to categorise the stimuli as either "field" (the combination of *flat* and *grassy*), "desert" (*flat* and *sandy*), "mountain" (*hilly* and *grassy*) or "dune" (*hilly* and *sandy*). Note that the specific categorisations are strictly identical in the two groups. The conjunctive nature of the stimuli warrants that the input scene can only be recognised as, e.g. "field," when its *flat* luminance and its *grassy* chrominance are perceived and integrated.

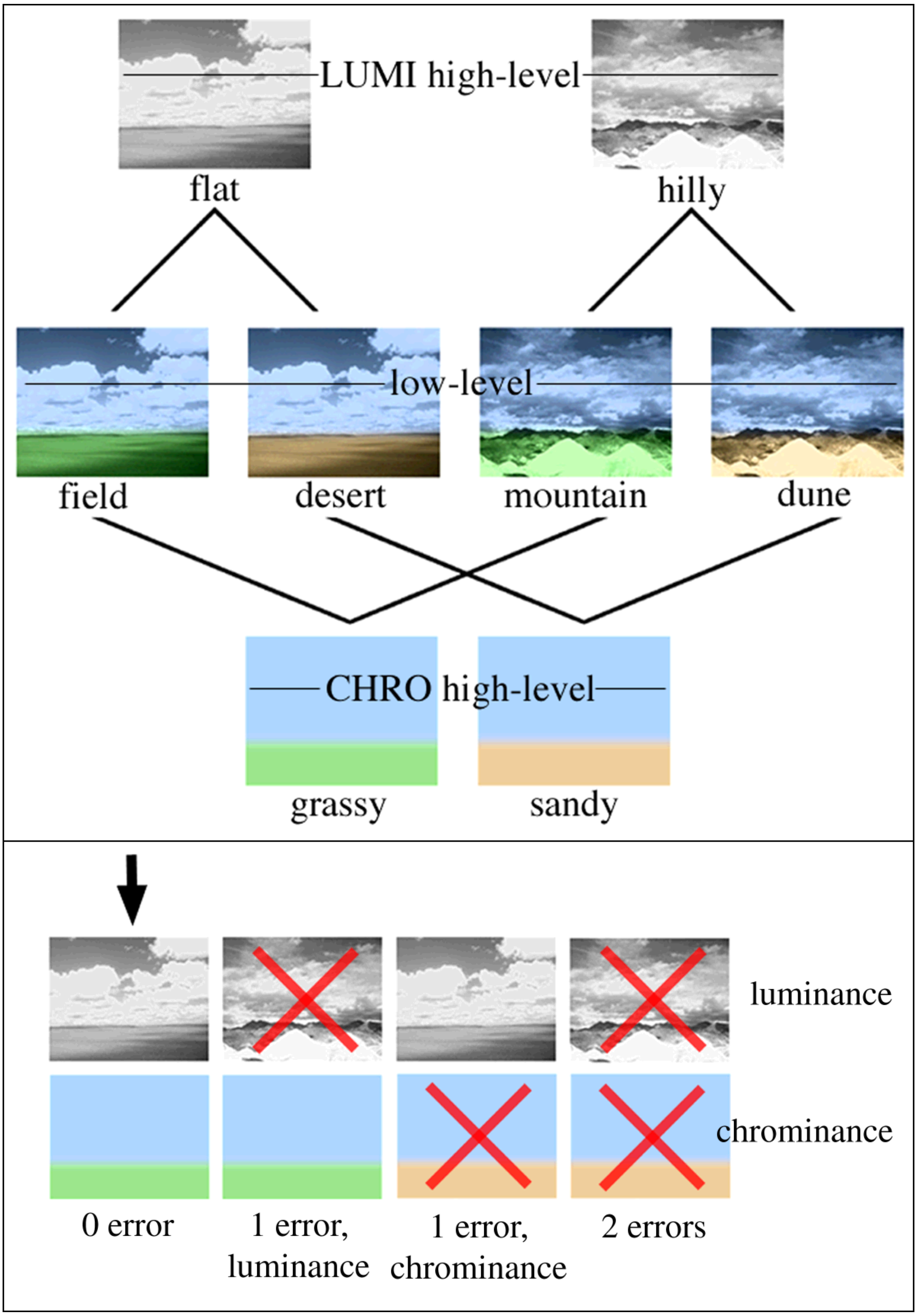


Figure 31. At the top, are the four scenes used in this experiment and the corresponding specific-level category names learned by all participants ("field", "mountain", "desert", and "dune"), surrounded by the two general-level categorisations ("flat" and "hilly") LUMI subjects learned, and those ("grassy" and "sandy") CHRO subjects learned. At the bottom, is an illustration of the analysis of categorisation errors and their implications. When a field is presented, the four possible categorisation

responses have a different implication for the perception of luminance and chromatic information.

This property can be used to ascertain whether subjects are more sensitive to the dimension defining the high, than the low level of their taxonomy, and therefore perceive the scenes according to their organisation of knowledge. Suppose that the field picture is briefly presented on the screen, immediately followed by a mask. Subjects can make four types of errors at the specific level, depending on which information they misperceive: "dune" implies a misperception of both the flat luminance and the green chrominance of the field (henceforth, 2 errors); "mountain" implies a misperception of only the flat luminance (henceforth, 1 error, luminance), whereas "desert" implies a misperception of only the green chrominance (henceforth, 1 error, chrominance). (Unfortunately, in Experiment 6, we did not distinguish the two types of 1 error responses separately.)

We predicted that the organisation of luminance and chromatic information in the LUMI and CHRO taxonomies would determine different perceptions of identical stimuli. That is, subjects placed in an identical condition of stimulation (e.g. seeing a field) and response (choosing between "field," "mountain," "desert" or "dune") would produce opposite patterns of categorisation errors (i.e. respond more often "desert" than "mountain" in LUMI, but "mountain" than "desert" in CHRO), revealing a differential sensitivity to luminance and chrominance in the groups (see Figure 31). (A similar analysis applies to all four stimuli of the experiment.)

5.15.1 Method

5.15.1.1 Participants

Twenty-four University of Glasgow students with normal or corrected vision were paid to participate in the experiment.

5.15.1.2 Stimuli

We used the quadruplet of artificial scenes of Experiment 6. That is, four 450 x 350 pixels (spanning 7 per 5.4 deg) stimuli–a field, a desert, a mountain, and a dune–synthesized with the Photoshop image processing software by combining two luminance patterns with two chromatic patterns (see top half of Figure 31). Our mask was created by randomly assigning to each square of a 18 x 14 grid the content of the corresponding region of one of the four scenes.

5.15.1.3 Procedure

The procedure was quite similar to the one of Experiment 6. We will nonetheless describe it and emphasize the major differences. The experiment ran on a Macintosh Power PC using a program written with the Psychophysics Toolbox for MatLab. Two subject groups (called LUMI and CHRO) learned the names of the four stimuli at the specific level of a taxonomy: "field," "mountain," "desert" and "dune." LUMI subjects also learned to categorise the stimuli at a general taxonomic level into "flat" vs. "hilly," on the basis of luminance cues, whereas CHRO subjects learn to categorise the stimuli at a general level into "grassy" vs. "sandy," using chromatic cues (see top half of Figure 31). (You will remember that Experiment 6 only had a LUMI subject group.)

A learning block was completed when participants had named consecutively–and without mistake–all scenes at the high and at the low levels of categorisation (4 scenes * 2 level of abstraction = 8 trials minimum), with LUMI and CHRO differing only in their high level categorisations. Within one learning block, trials followed each other in random order. A trial began with the display of the word “high” or “low,” instructing subjects of the level at which the subsequent scene had to be named. Subjects then pressed a key to display the scene which they had to categorise (presented on the screen for 1 s) immediately followed by a 450 ms mask. Subjects indicated their categorisation using one of six response-keys (two for the high level, four for the low level) before moving on to the next trial. Corrective feedback was provided.

When participants reached criterion (all subjects reached criterion after exposition to only four general- as well as four specific- levels trials, the minimum number), they were transferred to a testing phase including trials differing in three ways from those described above: First, in order to get the full spectrum of responses, presentation time varied, being either 15, 45, 75, 105, 135, 165, or 195 ms. Second, we only tested the low-level categorisations. That is, each one of the 700 test trials (4 scenes x 7 presentation times x 25 repetitions presented in a random order) started with the word “low” (Experiment 6 had “high” as well as “low” trials). This ensured that all participants were required to perform the exact same categorisations after the brief learning phase. Third, no corrective feedback was given.

5.15.2 Results and discussion

Figure 32 shows the evolution of 0 error, 1 error (luminance and chrominance, confounded), and 2 error responses for the two groups combined. At 15 ms exposure, performance is near chance; it quickly rises above chance for longer exposures. We bestfitted linearly the predictions derived from the section *5.13.1 Time-course predictions* equations with $S = 1$ (see solid lines in Figure 32). Overall $R^2 = .99$. A chi-square goodness of fit test did not reveal any significant difference between the observed and expected proportions ($^{2}_{obs}(20) = 17.03$, *ns*.). This replicates the low-level results of Experiment 6.

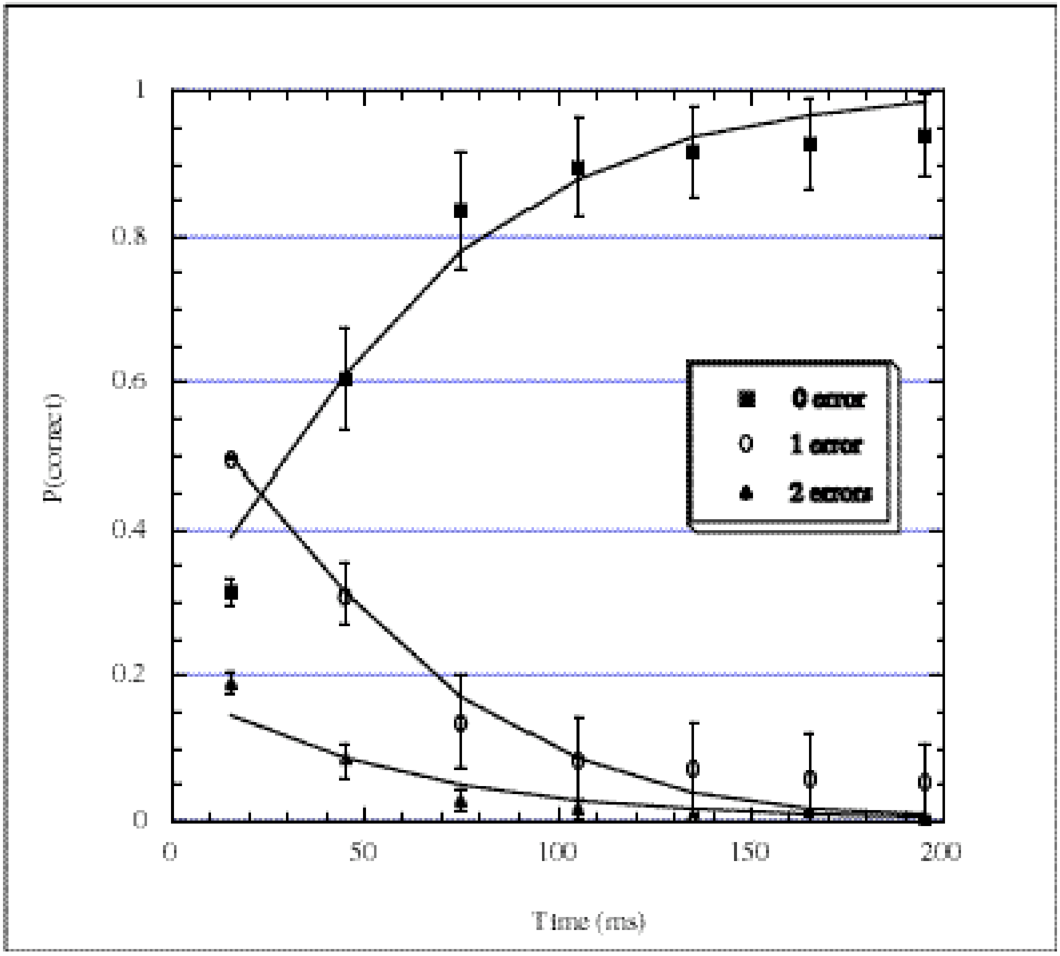


Figure 32. Average percent 0 error, 1 error (chromaticity and luminance errors confounded), and 2 errors, with standard deviations and SLIP bestfits (solid lines).

We predicted–on the basis of the robustness hypothesis–that subjects would produce opposite patterns of one-dimension errors (i.e., respond more often "desert" than "mountain" in LUMI, but "mountain" than "desert" in CHRO), revealing a differential sensitivity to luminance and chrominance. This is exactly what we observed: On average, CHRO subjects made more 1 error responses on luminance (58%) than on chrominance (42%), whereas LUMI subjects made more 1 error responses on chrominance (59%) than on luminance (41%). Significance tests on the difference score between 1 error responses on luminance and chrominance were different from zero in both groups (CHRO: mean difference = 16%; standard deviation = 13%, $Z = 4.423$, $p < .001$; LUMI mean difference = 19%; standard deviation = 21%, $Z = 5.848$, $p < .001$).

It is well established that luminance and colour are two of the main dimensions of visual processing (e.g., Livingstone & Hubel, 1988). If different categorisation strategies applied to strictly identical scenes, in strictly identical conditions of response and stimulus presentation, can produce a different order of integration of luminance and chromatic cues, then this would constitute strong evidence that categorisation strategies can determine perception. In the early days of vision research it was commonly thought that knowledge about the external world influenced its perception (Bruner & Goodman, 1947; Helmholtz, 1856); nowadays discoveries in the study of human knowledge rarely inform vision research, with the two fields drifting apart (e.g., Gordon, 1997). Our results could thus have potentially far reaching implications.

Chapter 6. General discussion

This dissertation presented SLIP, a measure of basic-level performance that implements two computational constraints on the organisation of information in taxonomies: *strategy length*, the number of feature tests necessary to place the input in one category, and *internal practicability*, the ease with which these tests can be performed. We designed SLIP to model category verification. We extended its reach to naming, and we discussed how SLIP relates to the other basic-levelness correlates (i.e., more features are listed at the basic level than at the superordinate level, with only a slight increase at the subordinate level; throughout development, basic level names are learned before those of other categorisation levels; and basic-levelness seems quite universal across domains as well as cultures). We reviewed 21 published experiments and examined how the two constraints varied in each one of them. We further examined strategy length and internal practicability in nine experiments. We used computer-synthesised artificial 3D objects or artificial scenes to tightly control feature composition and preserve ecological validity. The first five experiments examined the two constraints of SLIP in verification. Experiment 1 isolated the effect of strategy length on basic-levelness, Experiments 2A and 2B tested the effect of internal practicability, and Experiment 3 examined the interactions between the two factors. Experiment 4 verified whether strategy length is linearly related to basic-levelness, as predicted by SLIP. The last four experiments studied the two computational constraints in naming. Experiment 5A isolated the effect of strategy length, and Experiment 5B that of internal practicability. Experiment 6 looked at the time course of length 1 and 2 strategy completion. Finally, Experiment 7 examined the effect of *robustness* (i.e., the idea that an approximate categorisation is better than none) on the order of feature test in length 2 strategies.

Throughout, we compared the performance of SLIP at predicting basic-level data and that of other models. The comparisons were made on individual taxonomies or on meaningful collections of them that is, all the published taxonomies, the ones from our verification experiments, and the ones from our naming experiments. In the final section of this dissertation, we will proceed to an overall assessment of the reviewed basic-level models.

6.1 Overall assessment of the basic-level models

To our knowledge, we have examined all the formal models of basic-level performance. You will remember that these models are: Rosch et al.'s (1976) cue validity model, Tversky's (1977) contrast model, Jones's (1983) category feature-possession model, Corter and Gluck's (1992) category utility model, Fisher's (1986) COBWEB model, Anderson's (1989, 1990) rational analysis model, Medin and Schaffer's (1978; modified by Estes, 1994) context model, and Pothos and Chater's (1998, 1999) compression model (see Chapter 3).

We have rejected from further comparisons Rosch et al.'s (1976) cue validity model and Tversky's contrast model because they cannot predict the classic advantage for an intermediate level; Fisher's (1986) COBWEB measure because it is based on Corter and Gluck's category utility, and makes roughly the same predictions; and, finally, Anderson's (1989, 1990) rational analysis model because it does not provide a metric of basic-levelness. Combining the 21 published experiments (see Table 7) with the 11 taxonomies from our Experiments 1 to 6 (see Tables 15 to 18), the performance of the four remaining models (i.e., Jones's category feature possession, Corter & Gluck's category utility, Medin & Schaffer's context model, and Pothos & Chater's compression) at predicting basic-levelness is as follows: It appears that SLIP predicts 84% of the data, category utility, 63%, category feature-possession, 59%, compression, 57%, and context model, 41% (see Table 19 for a summary). Monte-Carlo simulations show that the context model ($p < .31$, *ns.*) does not perform significantly better than a chance model (i.e., a model that randomly selects the ranks of the levels within each experiment); and that all the other basic-levelness measures significantly outperform the chance model (i.e., in the worst case $p < .01$).

Table 19: Percentage of nominal data from 21 published basic-level experiments and of Experiments 1 to 6 explained by feature-possession (Jones, 1983), category utility (Corter & Gluck, 1992), compression (Pothos & Chater, 1998a), context model (Medin & Schaffer, 1978; Estes, 1994), and SLIP. The strategy length, internal practicability, simple conjunction, mixture, and mean scores flanked by a star are significantly above chance ($p < .01$).

	Strategy length	Internal practicability	Simple conjunction	Mixture	Mean
Possession	32%	74% *	46%	57%	59% *
Utility	54%	61% *	69% *	100%	63% *
Compression	49%	58% *	31%	100%	57% *
Context	30%	36%	38%	100%	41%
SLIP	100% *	81% *	54%	100%	84% *

Once more, it is instructive to examine the models specifically for their predictions of variations of feature redundancy and strategy length. Table 19 summarises the breakdown (the percentages presented here for simple conjunction and mixture experiments are the same as those presented in Table 7 in Chapter 4). For all experiments involving only variations of internal practicability, SLIP scores 81%, followed by category utility with 74%, then by category feature-possession with 61%, then by the compression measure with 58%, and trailed by the context model with 36%. All these scores are significant (i.e., in the worst case $p < .01$) except that of the context model ($p < .44$, *ns.*). The strategy length results are more interesting: For all experiments involving solely variations of strategy length, SLIP accounted for 100% of the data, category utility for 54%, compression for 49%, category feature-possession for 32%, and the context model for 30%. Of all these percentages, only SLIP's is significantly above chance ($p < .001$; for the next best model, $p < .3$, *ns.*). This confirms the argument made earlier that all models have so far neglected strategy length as a specific factor of basic level performance. This is a serious problem because attributes do overlap between categories in the real-world: what distinguishes your cellular phone, your fountain pen, your computer, your house, and other everyday objects of yours from those of your neighbours is often a conjunction of features (e.g., *colour* and *shape*).

To the extent that any model of categorisation implements computational constraints (even if these are not well specified), the conclusion is that those of SLIP are closest to those underlying the speed of access to the categories of a taxonomy. Therefore, if, for example, some animal categories are more equal than others (e.g., in verification tasks, "dog" is superior to "mammal" as well as to "Doberman"), we would say that this is because these superior animal categories have shorter strategies than the others, or strategies with greater internal practicability.

References

Adelson, B. (1985). Comparing natural and abstract categories: a case study from computer science. *Cognitive Science*, *9*, 417-430.

Anderson, J. R. (1990). *The adaptative character of thought.* New Jersey: Lawrence Erlbaum Associates, Publishers

Anderson, J. R. (1991). The adaptative nature of human categorisation. *Psychological Review*, *98*, 409-429

Anglin, J. M. (1977). *Word, object and conceptual development.* New York: Norton.

Archambault, A. and Schyns, P. G. (1999). *Ceci n'est pas une orange.* Manuscript submitted for publication.

Archambault, A., Gosselin, F., and Schyns, P. G. (in press). Natural bias for the basic-level? *Proceedings of the twenty-second annual conference of the cognitive science society*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Archambault, A., O'Donnell, C., and Schyns, P. G. (1999). Blind to object changes: when learning the same object at different levels of categorisation modifies its perception. *Psychological Science*, *10*, 249-255.

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*, 105-141.

Bendig, A. W. (1953). Twenty question: an information analysis. *Journal of Experimental Psychology*, *46*, 345-348.

Berlin, B. (1972). Speculations on the growth of ethnobiological nomenclature. *Language and Society*, *1*, 51-86.

Berlin, B. (1992). *Ethnobiological classification: principles od categorisation of plants and animals in traditional societies.* Princeton, New Jersey: Princeton University Press.

Berlin, B., Breedlove, D. E., and Raven, P. H. (1973). General Principles of classification and nomenclature in folk biology. *American Anthropology*, *75*, 214-242.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.

Biederman, I. (1990). Higher-level vision. In Daniel N. Osherson, Stephen M. Kosslyn, & John M. Hollerbach (Eds.), *An Invitation to Cognitive Science: Visual Cognition and action* (pp. 41-72). Cambridge, Massachusetts: MIT Press.

Biederman, I. (1998). Personal communication.

Biederman, I. and Gerhardstein, P. C. (1993). Recognizing depth rotated objects: evidence for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performances*, *18*, 121-133.

Biederman, I. and Schiffrar, M. M. (1987). Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 640-645.

Brainard, D. and Pelli, D. (1998). *Psychophysical Toolbox* [computer software], Santa Barbara.

Breitmeyer, B. G. (1989). *Visual masking: An integrative approach.* Oxford: Oxford University Press.

Brown, R. (1958). How shall a thing be called? *Psychological Review*, *65*, 14-21.

Brown, R. (1974). *A first language.* Cambridge, Massachussett: Harvard University Press.

Bruce, V. and Green, P. R. (1990). *Visual Perception: Physiology, Psychology and Ecology* (2nd ed.). London: Lawrence Erlbaum Associates.

Bruner, J. S. and Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*, 33-44.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. (1956). *A Study of Thinking*. New York: Wiley.

Bulmer, R. N. H. (1970). Which came first, the chicken or the egghead? In J. Pouillon & P. Maranda (Eds.), *Échanges et communications: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ième anniversaire* (pp. 1069-1091). The Hague: Mouton.

Callanan, M. A., Repp, A. M. McCarthey, M. G., and Latzke, M. A. (1994). Children's hypothesis about words meanings: Is there a basic level constraint? *Journal of Experimental Child Psychology, 57*, 108-138.

Cantor, N. and Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 12). pp. 3-52. New York: Academic Press.

Cantor, N., Smith, E. E., French de S., and Mezzich, J. (1980). Psychiatric diagnosis as prototype categorisation. *Journal of Abnormal Psychology, 89*, 181-193.

Chater, N. and Pothos, E. (1999). *Non-Parametric clustering by minimum description length.* Manuscript submitted for publication.

Christensen, E. E., Murry, R. C., Holland, K., Reynolds, J., Landay, M. J., and Moore, J. G. (1981). The effect of search time on perception. *Radiology*, *138*, 361-365.

Coley, J. D., Medin, D. L., and Atran, S. (1997). Does rank have its privilege? Inductive infereces within folkbiological taxonomies. *Cognition*, *64*, 73-112.

Conklin, H. C. (1954). *The relation of Hanuóo culture to the plant word.* Unpublished doctoral dissertation, Yale University, New Haven.

Collins, A. M. and Quinan, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour*, *8*, 241-248.

Corter, J. E., Gluck, M. A., and Bower, G. H. (1988). Basic levels in hierarchically structured categories. In *Proceedings of the XXth annual conference of the cognitive science society* (pp.118-124). New Jersey: Lawrence Erlbaum Associates, Publishers

Corter, J. E. and Gluck, M. A. (1992). Explaining basic categories: Features predictability and information. *Psychological Bulletin, 111*, 291-303.

Cutzu, F. and Edelman, S. (1998). Representation of object similarity in human vision: psychophysics and a computational model. *Vision Research*, *38*, 2229-2257.

Dewdney, A. K. (1989). *The Turing omnibus.* New York: Computer Science Press.

Diamond, J. (1966). Classification system of primitive people. *Science*, *151*, 1102-1104.

Dougherty, J. W. D. (1978). Salience and relativity in classification. *American Ethnologist*, *15*, 66-80.

Draper, N. R. and Smith, H. (1966). *Applied regression analysis.* New York: John Wiley & Sons.

Edelman, S. (1998). Shape representation by second-order isomorphism and the chorus model. *Behavioral and Brain Sciences*, *21*, 484-498.

Eimas, P. D. and Quinn, P. C. (1994). Studies on the formation of perceptually-based basic-level categories in young infants. *Child Development*, *65*, 903-917.

Estes, W. K. (1994). *Classification and Cognition*. Oxford: Oxford University Press.

Finnigan, D. (1996). *The complete juggler: all the steps from beginner to professional.* Atlanta: JUGGLEBUG.

Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning, 2*, 139-172.

Fisher, D. (1988). A computational account of basic level and typicality effects. In *Proceedings of the seventh national conference on artificial intelligence* (pp. 233-238). Saint Paul, Minnesota: Morgan Kaufman.

Garner, W. R. (1974). *The processing of information and structure.* New York: Lawrence Erlbaum Associates.

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General*, *107*, 287-308.

Garner, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, Cognition, and Development: Interaction Analyses* (pp. 1-38). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gauthier, I., and Tarr, M. J. (1997). Becoming a 'Greeble' expert: exploring the face recognition mechanism. *Vision Research*, *37*, 1673-1682.

Gibson, J. J. (1986). *The ecological approach to visual perception.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Goldstone, R. L. (1994). Influences of categorisation on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*, 178-200.

Gordon, I. E. (1997). *Theories of visual perception.* New York: John Wiley & Sons.

Gosselin, F. and Schyns, P. G. (1997). Debunking the basic level. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the cognitive science society* (pp.277-282). New Jersey: Lawrence Erlbaum Associates, Publishers.

Gosselin, F. and Schyns, P. G. (1998a). Contingency of parts in object concepts. In *Proceedings of the twentieth annual conference of the cognitive science society* (p. 1222). New Jersey: Lawrence Erlbaum Associates, Publishers

Gosselin, F. and Schyns, P. G. (1998b). 'Not-' cracker. *In Proceedings of the twentieth annual conference of the cognitive science society* (p. 1221). New Jersey: Lawrence Erlbaum Associates, Publishers

Gosselin, F., Archambault, A., and Schyns, P. G. (in press). Interactions between taxonomic knowledge, categorisation, and perception. In M. Ramscar, U. Hahn, E. Cambouropolos, & H. Pain (Eds.) *Similarity and categorisation.* Cambridge: Cambridge University Press.

Gosselin, F. and Schyns, P. G. (in press). The determinants of basic-level performance. *Proceedings of the twenty-second annual conference of the cognitive science society*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Gosselin, F. and Schyns, P. G. (1999). *Why do we SLIP to the basic-level? Computational constraints and their implementation.* Manuscript submitted for publication.

Gosselin, F. and Schyns, P. G. (2000). *New look at the interactions between between knowledge and perception.* Manuscript submitted for publication.

Halverson, J. (1992). Palaeolithic art and cognition. *The Journal of Psychology*, *126*, 221-236.

Hamm, J. P. and McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology: Human Perception and Performances, 24*, 413-426.

Hastings, N. A. J. and Peacock, J. B. (1975). *Statistical distributions.* New York: John Wiley & Sons.

Hays, W. L. (1988). *Statistics*. Chicago: Holt, Rhinhart and Winston.

Helmholtz, H. (1856-1866). *Handbuch der physiologischen optik* [Handbook of physiological optics]. Leipzig: Voss.

Hill, H., Schyns, P. G., and Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201-222.

Hoffmann, J. and Ziessler, C. (1983). Objectidentifikation in kunstlichen begriffshierarchien [Object identification in artificial concept hierarchies]. *Zeitschrift fur psychologie, 194*, 135-167.

Hofstadter, D. and the Fluid Analogies Research Group (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books.

Hogg, R. V. and Tanis, E. A. (1988). *Probability and Statistical Inference*. New York: Macmillan.

Horowitz,T. S. and Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*, 575-577.

Horton, M. S. and Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51*, 708-719.

Humphrey, G. W. and Bruce, V. (1989). *Visual cognition.* London: Lawrence Erlbaum Associates.

Hunn, E. (1977). *Tzeltal folk zoology: the classification of discontinuities in nature.* New York: Academic Press.

Johnson, K. E. and Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorisation. *Journal of Experimental Psychology: General, 126*, 248-277.

Johnson, N. I. and Kotz, S. (1969). *Discrete distributions.* Boston: Houghton mifflin company.

Jolicoeur, P., Gluck, M., and Kosslyn, S. M. (1984). Pictures and names: Making the connexion. *Cognitive Psychology, 19*, 31-53.

Jones, G. V. (1983). Identifying basic categories. *Psychologival Bulletin, 94*, 423-428.

Karmiloff-Smith, A. *Beyond modularity.* Cambridge, Massachussetts : MIT Press.

Klatzky, R. L. and Lederman, S. J. (1995). Identifying objects from a haptic glance. *Perception & Psychophysics*, *57*, 1111-1113.

Keppel, G. (1991). *Design and analysis: a researcher's handbook*. Englewood Cliffs, New Jersey: Prentice Hall.

LaBerge, D. A. (1962). A recruitement theory of simple behaviour. *Psychometrika, 27*, 375-396.

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorisation. *Journal of Experimental Psychology: Learning, Memory & Cognition, 20*, 1003-1021.

Lamberts, K. (1995). Categorisation under time pressure. *Journal of Experimental Psychology: General*, *124*, 161-180.

Landau, I. (1996). Hierarchical structure in schematic representations: Aspects of meaning in the cinematic shot. *Journal of Pragmatics, 26*, 737-766.

Lassaline, M. E. (1990). *The basic level in hierarchical classification.* Unpublished master's thesis, University of Illinois.

Lassaline, M. E., Wisniewski, E. J., and Medin, D. L. (1992). Basic levels in artificial and natural categories: are all basic levels created equal? In B. Burns (Ed.) *Percepts, Concepts and Categories: the Representation and Processing of Information.* (pp. 326-378). New York: North-Holland.

Lin, E. L., Murphy, G. L., and Shoben, E. J. (1997). The effects of prior processing episodes on basic-level superiority. *The Quaterly Journal of Experimental Psychology, 50A*, 25-48.

Lipschutz, S. (1973). *Theory and problems of probability.* New York: McGraw-Hill.

Livingstone, M. and Hubel, D. (1988). Segregation of form, colour, movement, and depth: Anatomy, physiology, and psychology. *Science*, *240*, 740-749.

Luce, D. R. (1986). *Response times: their role in infering elementary mental organisation.* Oxford: Oxford University Press.

Malt, B. C. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology*, *29*, 85-148.

Markman, A. B. and Gentner, D. (1996). Commonalties and differences in similarity comparisons. *Memory & Cognition*, *24*, 235-249.

Markman, A. B. and Wisniewski, E. J. (1997). Similar and different: the differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 54-70.

Markman, E. M. (1987). How children constrain the possible meaning of words. In U. Neisser (Ed.). *Concepts and conceptual development: ecological and intellectual factors in categorisation.* Cambridge: Cambridge University Press. (p. 255-287).

Markman, E. M. (1989). *Categorisation and naming in children: problems of induction.* Cambridge, Massachusetts: MIT Press.

Markman, E. M. and Hutchinson (1984). Children's sensitivity to constraints on word meanings: Taxonomic vs. thematic relations. *Cognitive Psychology*, *16*, 1-27.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman and Company.

McElree, B., Dolan, P. O. and Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*, 563-582.

McGill, W. J. and Gibbon, J. (1965). The general-gamma distribution and reaction time. *Journal of Mathematical Psychology, 2*, 1-18.

McMullen, P. A. and Jolicoeur, P. (1992). Reference frame and effects of orientation on finding the tops of rotated objects. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 807-820.

Medin, D. L. (1983). Structural principles of categorisation. In B. Shepp et T. Tighe (Eds.), *Interaction: perception, development and cognition* (pp. 203-230). Hillsdale, New Jersey: Erlbaum.

Medin, D. L and Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Medin, D. L., Lynch, E. B., Coley, J. D., and Atran, S. (1997). Categorisation and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology, 8*, 210-236.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.). *Concepts and conceptual development: ecological and intellectual factors in categorisation.* Cambridge: Cambridge University Press. (pp. 201-233).

Mervis, C. B. and Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development, 53*, 258-266.

Mervis, C. B. and Mervis, C. A. (1982). Leopards are kitty-cats: Object labeling by mother for their 13 months olds. *Child Development, 53*, 267-273.

Mervis, C. B. and Rosch, E. (1981). Categorisation of natural objects. *Annual review of psychology, 32*, 89-115.

Michnick Golinkoff, R., Suff-Mailey, M., Olguin, R., and Ruan, W. (1995). Young children extend novel words at the basic level: evidence fro the principle of categorical scope. *Developmental Psychology, 31*, 494-507.

Miller. G. A. (1956). The magical number seven plus or minus two: some limits to our capacity for processing information. *Psychological Review*, *63*, 81-97..

Miller, G. A. and McGill, W. J. (1952). A statistical description of verbal learning. *Psychometrika, 17*, 369-396.

Moore, C. and Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Cognition, 67*, 45-71.

Morris, M. W. and Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition, 18*, 407-418.

Murphy, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin, 91*, 174-177.

Murphy, G. L. (1991a). Parts in objects concepts: Experiments with artificial categories. *Memory & Cognition, 19*, 423-438.

Murphy, G. L. (1991b). More on parts in object concepts: Response to Tversky & Hemenway. *Memory & Cognition, 19*, 443-447.

Murphy, G. L. (1997). Personal communication.

Murphy, G. L. and Smith, E. E. (1982). Basic level superiority in picture categorisation. *Journal of Verbal Learning and Verbal Behaviour, 21*, 1-20.

Murphy, G. L. and Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*, 70-84.

Murphy, G. L. and Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorisation. In Koen Lamberts, David R. Shanks, et al. (Eds.), *Knowledge, concepts and categories: studies in Cognition* (pp. 93-131). Cambridge, Massachussett: MIT Press.

Neisser, U. (1987). From direct perception to conceptual structure. In U. Neisser (Ed.). *Concepts and conceptual development: ecological and intellectual factors in categorisation* (p. 11-24). Cambridge: Cambridge University Press.

Newport, E. L. and Bellugi, U. (1978). Linguistic expression of category in a visual-gestural language: A flower is a flower is a flower. In E. Rosch & B. B. Lloyd (Eds.), *Semantic factors in cognition* (pp. 137-168). Hillsdale, NJ: Erlbaum.

Norman, G. R., Brooks, L. R., Coblentz, C. L., and Babcock, C. J. (1992). The correlation of feature identification and category judgments in diagnostic radiology. *Memory & Cognition*, *20*, 344-355.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorisation relationship. *Journal of Experimental Psychology: General*, *115*, 39-57.

Oliva, A. and Schyns, P. G. (1997). Coarse blobs and fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*, 72-107.

Palmer, C. F., Jones, R. K, Hennessy, B. L., Unze, M. G., and Pick, A. D. (1989). How is a trumpet known? The "basic object level" concept and perception of musical instruments. *American Journal of Psychology*, *102*, 17-37.

Palmer, S. E., Rosch, E., and Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. D. Baddeley (Eds), *Attention and performance, IX.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, Massachussett: MIT Press.

Perrott, E. (Ed.). (1971). *Diversity among living things.* London: John Murray.

Posey, D. A. (1979). *Ethnoentomology of the Gorotire Kayapó of Central Brazil.* Unpublished Ph. D. dissertation, University of Georgia, Athens.

Pothos, E. M. (1999). *Compression measure* [computer software], Bangor

Pothos, E. M. and Chater, N. (1998). Rational categories. In *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 848-853). New Jersey: Lawrence Erlbaum Associates, Publishers

Pothos, E. M. and Chater, N. (1999). *A simplicity principle in human classification.* Manuscript submitted for publication.

Pothos, E. M. (1999). Personal communication.

Quinn, P. C., Palmer, V., and Slater, A. M. (1999). Identification of gender in domestic-cat faces with and without training: Perceptual learning of a natural categorisation task. *Perception, 28*, 749-763.

Quine, W. V. O. (1960). *Word and object.* Cambridge, Massachussett: MIT Press.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes on scenes. *Psychological Science*, *8*, 368-373.

Rifkin, A. (1985). Evidence for a basic level in event taxonomies. *Memory & Cognition, 13*, 538-556.

Rosch, E. (1978). Principles of categorisation. In E. Rosch & B. B. Lloyd (Eds.), *Semantic factors in cognition* (pp. 137-168). Hillsdale, NJ: Erlbaum.

Rosch, E. and Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-352.

Schyns, P. G. (1997). New perspectives on category learning. *Trends in Cognitive Science, 1*, 183-189.

Schyns, P. G. (1998). Diagnostic recognition: Task constraints, objects information, and their interactions. *Cognition, 67*, 147-179.

Schyns, P. G. and Murphy, G. L. (1994). The ontology of part representation in object concepts. *The Psychology of Learning and Motivation*, *31*, 305-349.

Schyns, P. G. and Oliva, A. (1994). From blobs to boundary edges: evidence for time and scale dependant scene recognition. *Psychological Science*, *5*, 195-200.

Schyns, P. G. and Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*, 243-265.

Schyns, P. G. and Rodet, (1997). Categorisation creates functional features. *Journal of Experimental Psychology: Learning, Memory & Cognition, 23, 1-16.*

Schyns, P. G., Goldstone, R. L., and Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1-53.

Shaver, P., Schwarz, J. Kirson, D., and O'Connor, D. (1987). Emotion knowledge: further explorations of prototypeapproach. *Journal of Personality and Social Psychology*, *52*, 1061-1086.

Shepard, R. N. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Simons, D. J. and Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*, 261-267.

Smith, E. E. and Medin, D. L. (1981). *Categories and concepts.* Cambridge, Massachusetts: Harvard University Press.

Stross, B. (1973). Acquisition of botanical terminology by Tzeltal children. In M. S. Edmonson (Ed.) *Meaning in Mayan languages* (pp. 107-141). The Hague: Mouton.

Tanaka, J. W. and Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457-482.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., and Blanz, V. (1997). To what extent do unique patrs influence recognition across changes in viewpoint? *Psychological Science, 8*, 282-289.

Taylor, P. M. (1990). The folk biology of the Tobelo people: a study in folk classification. In *Smithsonian Contribution to Anthropology #34*. Washington, DC: Smithsonian Institute Press.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

Thorpe, S., Gegenfurtner, K. Fabre-Thorpe, M. J. and Bülthoff, H. H. (1999). Categorisation of complex natural images in extreme peripherical vision. *Perception, 28*, 61.

Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Tversky, B. and Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology, 15*, 121-149.

Tversky, B. and Hemenway, K. (1984). Objects, parts and categories. *Journal of Experimental Psychology: General, 113*, 169-191.

Tversky, B. and Hemenway, K. (1991). Parts and the basic level in natural categories and artificial stimuli: Comments on Murphy (1991). *Memory & Cognition, 19*, 439-442.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language, 35*, 434-453.

Wisniewski, E. J. and Gentner, D. (1991). On the combinatorial semantics of noun pairs: minor and major adjustements to meaning. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 241-284). Amsterdam: Elsevier Science.

Wisniewski, E. J. and Markman, A. B. (1993). The role of structural alignment in conceptual combination. In *Proceedings of the fifteenth cognitive science society* (pp. 1083-1086). Boulder, Colorado: Erlbaum.

Wisniewski, E. J. and Markman, A. B. (1996). Feature alignment in category learning. Manuscript submitted for publication.

Wisniewski, E. J. and Murphy, G. L. (1989). Superordinate and basic category names in discourse: a textual analysis. *Discourse Processing, 12*, 245-261.

Wolfe, J. M. (1999). How do you pay attention? *Science, 400*, 813-815.

Woodbury, M. A. (1949). On a probability distribution. *Annals of Mathematical Statistics, 20*, 311-313.

Woodman, G. F. and Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Science, 400*, 867-869.