NEW BULGARIAN UNIVERSITY

CENTRAL AND EAST EUROPEAN CENTER
FOR COGNITIVE SCIENCE

JUDGEMAP: A Model of Judgment, Based on the Cognitive Architecture DUAL

Georgi Petkov

A dissertation submitted in partial fulfillment of the requirements for the degree of Ph.D. in Cognitive Science

Supervisor:

Assoc. Professor Boicho Kokinov

SOFIA ● 2005

Summary

Judgment of a target stimulus can be systematically shifted in either direction depending on the context. The experimental findings are robust, but often controversial – seemingly similar experiments can lead to opposite effects depending on the stimuli or the procedure. There is no single theory that can explain and predict all these controversial data.

In addition, very few computational models of the judgment process are implemented in a way that simulation experiments can be run on them. The hope is that building such models could contribute to the understanding of these controversial data in a coherent way. Our assumption is that several mechanisms interplay in the process of judgment, each of them leading either toward assimilation or toward contrast, and the behavioral data reflect only the outcome of the competition between these mechanisms. That is why seemingly similar experiments could lead to opposite results.

Since the model of judgment should be dynamic and context-sensitive, it seems natural to use the dynamic and context-sensitive cognitive architecture DUAL as a basis for modeling. In this way, the model of judgment will be integrated with other cognitive processes rather than being considered in isolation.

Moreover, an attempt has been made to use the same basic DUAL mechanisms that have been used for analogy-making to model judgment as well. This exerts a strong restriction on the range of possible models of judgment and does not allow us to construct any possible mechanism that will lead to the desired data. On the contrary, if successful, the results would mean that the same mechanisms could produce analogy and judgment as well as the assimilation and contrast effects. This is a very challenging task and of course, it could not be fully accomplished within a dissertation, however, this is the driving force behind the studies.

An underlying assumption behind the JUDGEMAP model is that judgment is based on a process of mapping between a set of stimuli and the set of possible scale values. In this way, judgment is considered to be a close relative to analogy-making and based on the same mapping mechanisms. This mapping is trying to keep the relational structure of the two domains in correspondence, i.e. better stimuli to be mapped on higher ratings. Thus ordering relations play important role. Another important assumption is that the set of stimuli used as background when evaluating a single target stimulus (called comparison set) is dynamically constructed in WM by spreading activation mechanisms and based on similarity with the target stimulus. Then the target stimulus is included in this comparison set and all elements of this set are mapped on the scale competing for the specific scale values.

The result is that very few modifications of the existing AMBR mechanisms have been made and some new mechanisms have been developed, holding all principles of the DUAL architecture. A wide range of psychological data has been replicated with the model. In addition, the model has made some predictions that have later on been tested and confirmed. Some of these predictions have seemed quite strange at first but have turned out to be true. Confirmation of non-trivial predictions is one of the model's strengths. In addition, it has been demonstrated that the same mechanisms that are used for judgment can also produce choice and can even replicate some context effects on choice. At the same time, the model could not replicate all the existing experimental data and thus further development is needed, especially integrating it with perception and category learning.

Table of Contents

CHAPTER I
Introduction

1.1 Introduction to the Judgment Task

"Please, rate how long is the line that you see on the screen on an ascending scale from 1 to 7." This is a typical example of a judgment task. It has several parameters. Some of them are related to the nature of the stimuli – the property, which is judged, the category of the stimuli, etc. Others are related to the scale, on which the stimuli are judged, more precisely, which are the extreme points of the scale and what is its density.

Four of the parameters - the stimulus **S** to be judged, the property **P** that is judged, and the scale (from **a** to **b**), on which it is judged are usually explicitly presented. Sometimes the scale is presented with verbal labels, for example, "1 – (it is) not long at all, 2 – (it is) not long … 7 – (it is) very long"; or "1 – (it is) very short … 7 – (it is) very long".

However, people are able to judge even without such verbal labels. They understand unambiguously the instruction: "Please, rate the property **P** on a scale from **a** to **b**". It is *implicitly assumed* that smaller magnitudes of **P** should correspond to smaller ratings and vice versa, and that approximately equal differences between the magnitudes should correspond to approximately equal differences between their ratings.

The nature of the stimuli to be judged can be very broad – from lines, to human faces, to traits or emotions. Either a single dimension of the stimuli (for example, its length or redness) or a complex one (e.g., its attractiveness) may be judged.

The type of the scales can also vary. It could be ordinal and a limited one – for example, from **a** to **b**, but could also be a continuous one ("Judge this line length using a real number from 0 to 1). The scale boundaries could be predefined, or not, they can even be non-existing – for example, the instruction can be "Rate the stimuli with a real number – whatever you wish".

Some theories distinguish judgments of simple[1] stimuli, like lines, squares, etc. from judgments of complex ones – human performances, or emotions. However, it is not easy to find any reason to distinguish the *mechanisms* that underlie them. As Strack (Strack, 1992) said, all judgments should be based on calculations or on perceptions (or on both). This is true of both simple and complex stimuli.

The task of making a judgment is not manifested only in psychological laboratories with explicitly given instructions. Instead, people make judgments every day and everywhere. They estimate distances, colors, judge people – "I like this person, I do not like that one", etc. The judgments are strongly related with choice, decision-making, and other cognitive tasks. It seems reasonable to assume that the same basic mechanisms may underlie them.

[1] A judged stimulus is called *simple* if it is judged according to a single dimension that can be measured with physical instruments. A judged stimulus is called *more complex* than another judged stimulus if the former is judged according to more dimensions than the latter one.

1.2 Classical Approach – Judgment as Measuring

The early approaches to the mechanisms that underlie judgment treat the process of judgment as a precise measurement of the judged property. In other words, the assumption is that people have internal instruments which they use to gauge the stimuli. Then they monotonically transform the result according to the given scale. Psychophysics studies these measuring instruments, their capabilities and limitations.

This approach is intuitively simple and clear, but faces a problem when analyzing people's responses. People often rate the same stimulus with different ratings, even when it is presented twice within a short time. This dispersion in stimulus ratings is explained by psychophysics with some imprecision of our internal measuring instruments, i.e. they work with a certain noise.

Many experimental results, however, are in contradiction with this assumption. It has been shown that in certain conditions judgments shift *systematically*, not randomly. A certain change of stimuli distribution, or certain contextual elements could produce predictable changes of the mean and the dispersion of ratings. Hence, maybe it is wrong to treat the context just as a source of noise.

The number of such violations of the classical approach has grown during the last decades, and alternative theories have been proposed. A brief list of experiments that demonstrate systematic shifts in judgment, depending on different contextual manipulations, is presented in the next chapter.

1.3 Alternative Approaches to Judgment

There are three main approaches to describing the 'errors', i.e. the systematic shifts in responses due to certain contextual influences.

Sometimes researchers try to define the boundaries of these errors by studying systematically the input–output patterns. In other words, they perform series of experiments, each time shifting slightly some parameters, and recording the shifts in responses. Then the scientists interpolate mathematical formulas that represent the dependency of the responses on the input. Such theories are usually called *normative theories*.

Another way to study the phenomena is to presume that there are certain *heuristics* that people use, and then to verify this with psychological experiments. Heuristics are simple rules, whose application and use is easier than those of the more complicated and precise measurements. In most cases, these rules produce 'correct' answers, but under some special circumstances, they may produce systematic shifts. The theories that describe the processes with such heuristics are usually called *descriptive theories*.

Building *cognitive architectures* is a third approach. The researchers in this group try to define a small set of basic principles that underlie cognition in general. Different cognitive capabilities have to be formulated in terms of *models*, based on the architectures. The models based on the same architecture have to be integrated in a single whole. After that, the empirical data on a certain cognitive task have to be explained with the presumed basic principles and a minimum number of proposed novel mechanisms.

1.4 JUDGEMAP Model

JUDGEMAP is a computational cognitive model for judgment, based on the cognitive architecture DUAL (Kokinov, 1994b, c), which is implemented in a computer program. It is focused mainly on the judgment task, but one of its assumptions is that the same mechanisms underlie other cognitive abilities as well, choice in particular.

However, the capabilities of JUDGEMAP to model choice are only briefly demonstrated. The main work in this direction is subject to further research.

JUDGEMAP proposes several mechanisms that overlap in the course of time and influence one another.

The model assumes that a necessary condition to judge a certain stimulus is to have something to compare it with, i.e. a *context* is needed. The context may consist of some explicitly presented stimuli, of stimuli that have been used recently, and also of stimuli that are retrieved from memory because they are similar in some respect to the judged stimulus.

JUDGEMAP allows some *relations* between the available stimuli and their properties to be recognized. However, only *relevant* relations between *relevant* units are explicitly represented.

Based on the constructed relations, *mappings* between the constructed stimulus set and the available scale labels occur. The construction of such mappings is constrained by the requirements of the judgment task. Different *hypotheses* about the best answer emerge one after another. The inconsistent ones compete with each other; the consistent ones support each other. As a final result, one winner hypothesis is interpreted as the response of the system.

JUDGEMAP assumes that *choices* between alternatives are based on the same mechanisms. However, when choosing, the driving force is not to find the most proper rating for a certain stimulus, but to find the most proper stimulus that corresponds to the highest rating.

1.5 Verifications of the Model

A series of simulations has been run with JUDGEMAP, highlighting most of its capabilities and shortcomings.

JUDGEMAP proposes possible explanations for some of the known phenomena in terms of principles that the DUAL architecture assumes to be fundamental for cognition in general. However, since the model has not accounted for several behavioral data in judgment, additional improvements are needed. In particular, all stimuli representations are given to JUDGEMAP manually, i.e. they are predefined. If the model were integrated with abilities for categorization and high-level perception, its explanatory power would probably increase.

1.6 Overview of the Thesis

This dissertation presents a model of judgment. Chapter II presents briefly the experimental results in the field of relative judgment. Chapter III presents an overview of the main theories of judgment. In Chapter IV, the DUAL architecture and the JUDGEMAP model are briefly outlined. The detailed description of the mechanisms of JUDGEMAP is presented in Chapter V. In Chapter VI, the simulations performed by the model are presented. Some shortcomings of the model and open issues are listed in Chapter VII. Chapter VIII presents some conclusions that have been made.

CHAPTER II
The Role of Context in Judgment

Mussweiler (2003, p.472) pointed out that "when we evaluate a given target, we do not do so in a vacuum." A number of publications from the second half of 20^{th} century reported systematic shifts in judgment due to specific contextual manipulations. They violate the classical psychophysical theories that treat the context as a noise and give rise to alternative theories of judgment.

The contextual shifts in judgment are usually separated into two groups – assimilation effects and contrast ones. We speak about assimilation when by adding a contextual stimulus in the task the judgments systematically shift toward the magnitude of the context. The contrast effect is observed when the judgments shift away from the contextual stimulus.

Unfortunately, there is no single non-controversial scheme that can explain and predict exactly under which circumstances assimilation or contrast would be observed. Moreover, opposite effects have often been registered in almost identical experimental conditions. The hope is that better predictions will be possible by building models of judgment.

In this chapter, some of the observed data are presented. Since JUDGEMAP also suggests some ideas for modeling choice, a specific part of the contextual effects in choice is described at the end.

2.1 Assimilation Effects

Anderson (1966) gave participants three traits describing a person and asked them to judge the characteristics on a 1-20 numerical scale according to likeability. The result was that the rating of each trait was displaced toward the other two. People perceive the set of traits as a Gestalt when they describe a particular person. When they are asked to judge a single trait, they do not judge it in isolation but consider the whole Gestalt. Anderson called this effect 'positive context effect' and later it became known as the *assimilation effect*.

2.1.1 Assimilation toward a Contextual Stimulus

Wyer (1974) reported a more complicated replication of Anderson's experiment. He separated the adjectives used into three groups according to their likeability – low, medium, and high; and into two groups according to their ambiguity – low and high. He found that the assimilation effect increased when the target stimulus was ambiguous[1], independently from its degree of likeability.

Wedell, Parducci and Geiselman (1987) asked participants to judge attractiveness of female faces (presented in pictures). When pairs of pictures were

[1] A judged stimulus is called *ambiguous* either if it could be categorized in many different ways, or if the judgment task could be understood in many different ways, i.e., it is ambiguous to define which are the dimensions of the judged stimulus relevant to the judgment task.

given to be judged simultaneously, the two ratings within each pair shifted toward each other, in comparison to the case when the same faces were rated individually.

In all these experiments, complex (multi-dimensional) and ambiguous stimuli were used. In these cases, people first have to retrieve or construct complex criteria for judging (combining several dimensions of the stimuli) and then to use these criteria for rating the stimuli. This is in contrast with the simple cases where the given dimension is judged directly ("How long is the presented line"). Many other experiments are in favour of the hypothesis that when ambiguous stimuli have to be judged, the external contextual elements tend to produce an assimilation effect in judgments.

For example, Herr, Sherman & Fazio (1983) asked participants to judge the size of animals, even if they are unknown to them (authors also used non-words as labels of animals). Before making a judgment, participants were primed with different animals. The experimental results demonstrate that people tend to assimilate their ratings toward the size of the prime, but only in the condition where moderate exemplars were primes and ambiguous stimuli were judged. In some other conditions, the opposite effect appeared and later in the text (section 2.2.4) this experiment will be discussed again in more details.

Another demonstration of assimilation toward the prime elements when judging ambiguous stimuli was proposed by Strack, Schwarz & Gschneidinger (1985). They asked participants to list three positive events that had happened to them recently and after that to judge their overall satisfaction with their life. In a second group, participants were asked to list three negative events and to rate their overall satisfaction with life. Participants in the first group expressed higher satisfaction than those in the second. This result can be interpreted as an assimilation effect of the ratings toward the ratings of the recalled events.

In a series of experiments Stapel and Kooman (1996, 2000) used social stimuli (e.g. trait names or sentences that describe different behaviours) to demonstrate assimilation toward primed general frames (in contrast, when concrete exemplars are primed, the context beat out the judgments). Here again the judged dimensions were not well defined. The novelty is that the assimilation was toward general frames, not toward individual items.

Finally, Manis, Nelson & Shedler (1998) demonstrated an assimilation effect toward moderate stereotypes. They formed stereotypes about particular hospitals in their subjects by presenting them with a series of pathological cases. After that, they gave to the participants two types of tasks – to judge the pathology of specific cases, or to choose the more pathological case from a pair of cases. In both tasks[1] people assimilated their responses toward the induced stereotype, but only when this stereotype was relatively moderate. In the cases when the stereotype was extreme, a contrast effect was observed.

However, there are experiments that demonstrate assimilation toward a certain contextual stimulus even when the judged dimension is simple and unambiguous. Sherif, Taub & Hovland (1958) asked participants to judge weights on an ascending scale from 1 to 6 – "1 means lightest, etc., 6 means heaviest". In each trial, the participants hold the heaviest weight simultaneously with the target stimulus. The

[1] Assimilation in the case of choice means that people systematically reverse their choices in a specific direction (toward the stereotype) in comparison with a control group.

experimenters said to participants to think of this heaviest weight as a prototype for the rating 6 (this procedure usually is called *anchoring*). The result was that all target stimuli were overestimated, i.e. their ratings shifted toward the anchor in comparison to the ratings, given without an anchor. Parducci & Marshall (1962) replicated the experiment and received the same result. It is interesting that in both experiments the result had reversed in another experimental condition, when the anchor was more extreme (outside the range of the judged stimuli).

These experiments used quite an unusual modality for judgment. It seems that judging weights is an unambiguous and well defined task but maybe it is not. The task does not require any concepts to be retrieved or constructed. The only necessary criterion for judgment is an estimation of the changes in muscle tension. However, there are too many complex pressures that influence this tension – adaptation, lateral inhibition, tiredness, etc. and the information is brought through a number of sensors and has to be integrated. Thus, it might be the case that judgment of weights is actually quite a complex task (in contrast to judgment of length or redness).

2.1.2 Sequential Assimilation Effect

One typical judgment task is to ask participants to judge sequentially series of homogeneous stimuli that vary only in the value of the judged property. Almost all experiments of this type demonstrate a robust sequential assimilation effect – the rating of each stimulus tends to shift toward the rating of the previously judged one.

For example, Lockhead (1992) studied the integral nature of bi-dimensional stimuli – auditory tones with loudness and pitch. He demonstrated that judgments of each tone were assimilated toward the rating of the previous bi-dimensional stimulus. Moreover, assimilation was detected not only for the target dimension, but also for the combination of both dimensions.

Another study, reported by Petrov & Anderson (2000) showed similar results when participants judge lengths of line segments. Counter-intuitively, he found that assimilation is only toward the previous *rating*, but not toward the previous *stimulus*. More precisely, assimilation was measured with the standardized regression coefficients in a multiple linear regression with the following variables entering as predictors – the current stimulus S_t, the previous stimulus S_{t-1}, and the previous response R_{t-1}. The authors received standardized coefficient –0.25 (std. dev. 0.10) for the previous stimulus S_{t-1}, and +0.30 (std. dev. 0.10) for the previous response R_{t-1}, thus demonstrating evidence for stimulus-driven contrast and response-driven assimilation.

The sequential assimilation effect seems to be very robust and any theory of judgment should account for it. It is reported also in Ward (1973, 1979), Jesteadt, Luce & Green (1977). The fact that the assimilation is toward the ratings, not toward the stimuli, needs special attention. This observation is one of the crucial points in JUDGEMAP model.

Assimilation effects toward a contextual stimulus tend to be manifested when the target is ambiguous or complex and toward the previously used ratings. However, assimilation has been detected in other circumstances as well.

2.1.3 Other Manifestation of Assimilation Effects

2.1.3.1 Perceptually Driven Assimilation toward a Category

Goldstone (1995) asked participants to adjust the color of a given probe, in order to equalize it to the color of a simultaneously presented standard. The trick is that the probes are extracted from two categories – letters and numbers. The letters presented to the participants were predominantly more violet, whereas the numbers were predominantly more reddish. The result was that when the participants had to adjust the color of a given stimulus to match the color of a given standard they adjusted the letters to be more violet, and the numbers - more reddish. In other words, assimilation toward the *category* was demonstrated (in comparison with the adjustments of stimuli with the same color, but from different category). Furthermore, this assimilation seems to appear at a very low level of the cognitive processes – the task was just to adjust colors that the participants had seen.

In the same study Goldstone tried to find out which is the driving force behind this effect - whether the category 'draws up' its exemplar, or the other category 'beats it back'. The conclusion from his experiments is that both forces play a role, depending on the degree of similarity between the target and the contextual stimuli.

These results are again in favor of the hypothesis that when judging an unambiguous dimension, the ratings tend to be assimilated toward a certain category or sub-category. The novelty, however, is that Goldstone highlighted the role of perception in this process. He argued that the assimilation could not be explained without any perceptual mechanisms. To get ahead of the story, JUDGEMAP fails to explain some of the assimilation effects and I assume that exactly the lack of perceptual mechanisms in JUDGEMAP is the reason for these failures.

2.1.3.2 Assimilation toward a Number Value

Strack & Mussweiler (1997) asked participants to answer two questions. The first one was whether the Brandenburg Gate is shorter or taller than 150 meters. Soon after the answer, the same participants had to judge how tall the Brandenburg Gate (in meters) is. People answered the latter question with numbers that were closer to the anchor 150 meters, in comparison with a control condition. The same results have been demonstrated when people had to judge the mean temperature in Antarctica, when they were anchored with a certain value. However, this type of assimilation decreases (but does not disappear) if a different object or a different property of the same object had been judged. For example, the anchor about the height of Brandenburg Gate has less influence on the answers to the question of its width.

This result is in synchrony with the fact that the assimilation effect is toward the previous given rating (or primed numbers), not to the previous judged stimuli. Maybe the scale labels, considered as numbers, form an important part of the judgment process.

2.1.3.3 Assimilation toward the Density of the Stimulus Distribution

Usually the density of the stimulus distribution beats back the stimulus ratings (the famous frequency effect is presented later in the text). In some circumstances, however, assimilation can be observed.

Cooke et al (2004) asked participants to rate cheeses that depend on their price and quality. This is a case of a typical trade-off judgment, in which usually an inversed U-shape curve illustrates people's ratings. In other words, people rate the moderate stimuli higher than those, which have an extremely low price and extremely low quality, or vice versa. Cooke et al. demonstrated that if the overall distribution is skewed (for example, the cheep and low quality cheeses dominate), the peak of people's ratings shifts toward the skew, e.g., toward the point of concentration of more stimuli.

However, it is important to stress that this result is unique. The changes of the stimuli distribution usually produce the opposite effect (section 2.2.3).

2.1.4 Summary of the Assimilation Effects

The tendency to shift judgments toward a particular contextual element is robust and has been demonstrated in a number of experiments. Assimilation effects tend to be observed:

- toward the overall impression of a Gestalt, when the judged stimulus is part of this Gestalt (Anderson, 1974);
- toward the value of a primed stimulus, when this primed stimulus is relatively moderate, and the judged stimulus is ambiguous (Herr, Sherman, Fazio, 1983);
- toward primed general schema (Stapel, Kooman, 1996, 2000);
- toward moderate stereotypes (Manis, Nelson & Shedler, 1998);
- toward the prototypical value of the category (Goldstone, 1995);
- sometimes toward the rating of a stimulus, given simultaneously with the target one, in particular if the judged stimuli are too complex, like human faces (Wedell, Parducci & Geiselman 1987);
- toward the rating of the previous judged stimulus, when a series of homogeneous stimuli are judged (Lockhead 1992, Petrov, 2005);
- toward the primed numbers, when the task is for value estimation (Strack & Mussweiler, 1997).

Generally, assimilation tendencies could be observed in two main conditions. First, when people judge ambiguous dimensions of the stimuli, their ratings tend to be shifted toward the values of the stereotypes of available categories or sub-categories. Second, people tend to shift their ratings toward previously used ratings or other numbers.

2.2 Contrast Effects

Kenrick & Gutierres (1980) showed participants a TV-program with a strikingly attractive woman and after that asked them to rate female pictures. People underestimated all females in comparison with ratings, given in a second group, where people had watched another TV-program. The tendency to shift judgments away from the value of a certain contextual stimulus is called *contrast effect*. Some experimental conditions, in which contrast effects can be observed, are presented in this section.

2.2.1. Contrast from an Anchoring Stimulus

The same contrast effect was demonstrated with the anchoring procedure. For example, Sherif, Taub, Hovland (1958) and Parducci & Marshall (1962) asked participants to judge weights, instructing them that a certain stimulus – *anchor* - always has to be judged with the maximal value. People held both weights simultaneously. As was mentioned in the previous section, when the anchor had the same weight as the maximal one from the test set, assimilation occurred. When the anchor had been more extreme, however, the effect was reversed – a contrast effect was demonstrated.

Sarris & Parducci (1978) performed an analogous experiment, asking people to judge the size of squares on a numerical scale. The squares were presented sequentially, but people had to judge only the even ones. The odd squares served as anchors – the instruction was not to judge them. The anchors were equal or larger in size to the larger one from the test set and could vary from trial to trial. In all cases, a contrast effect was observed, but it was greater when the anchor was equal to the larger target square. The more extreme the anchors were, the smaller the contrast effect was.

In general, when people judge a certain simple property of a given stimulus, their judgments shift away from the value of the predefined prototypical examples from the same category. This observation is not quite surprising in the sense that people try to *differentiate* the target stimulus from the exemplary ones. As a result, they probably often pay more attention to the differences.

Unlike the contrast from concrete external exemplars described in this section, people often have to judge large series of homogeneous stimuli. The next section will review several experiments that demonstrate contrast effects due to the stimulus distribution.

2.2.2 Range Effect

The stimulus grade depends on the overall range of the set of judged stimuli. The experimental template for demonstrating the range effect is usually the following: A uniformly distributed set of homogeneous stimuli is given for judgment to the participants. After that, a subset is formed, removing all stimuli, whose physical value of the judged property exceeds a certain threshold. For stimuli that are more complicated, this physical value can be ambiguous or unknown. In such cases the mean ratings given by the participants can serve for this purpose. Once this restricted subset is formed, it can be given to participants (the same group after delay or another group) to be judged. Usually people tend to use all scale values in their judgments, i.e. the ratings of the stimuli in the restricted subset would capture the whole scale, not only the respective restricted part of it. This is called *range effect*.

The range effect is very robust. It has been demonstrated with a number of different stimuli under different conditions. Weddel, Parducci & Geiselman (1987) demonstrated it with female faces; Mellers & Cooke (1994) – with overall class performance, based on the percentage of correct scores on two independent exams; Parducci (1968), Parducci & Perret (1971) – with sizes of squares, etc.

It is important, however, to mention the observations by Wedell (1996), that when the stimulus distribution shifts, the ratings' shift is not immediate, but after some delay – several stimuli have to be presented in order for people to capture the new distribution.

To sum up, the Range effect is an excellent demonstration of the relativism in human judgments. People judge according to something – usually according to the category, in which the stimulus belongs. This is intuitively clear - for example, everybody understands what 'small' in 'a small elephant' means and what 'large' in 'a large butterfly' means – the elephants are compared to other elephants, the butterflies – to other butterflies. People do not judge the actual physical size of something without taking into account the category to which it belongs. Each theory of judgment should also reflect the effect of the category range.

The next section describes another aspect of the stimulus set that was demonstrated to affect judgments, namely, the *frequency effect*.

2.2.3 Frequency Effect

Frequency effect is observed when the stimulus set is not uniformly distributed, but the stimuli are concentrated closer to one of the extremes. In this case, people's ratings shift away from the direction of the density.

For example, Parducci & Perret (1971) asked participants to judge the size of squares. The overall set of squares consisted of nine square sizes, but was positively (e.g. the small squares were more than the large ones) or negatively (the large ones dominated) skewed. In the first case, people gave higher ratings than in the second one (for one and the same square size).

Wedell, Parducci & Geiselman (1987) performed an analogical experiment with similar results by using female faces for stimuli.

Cooke & Mellers (1998) asked people to rate offers for apartment rents. The offers differed on three dimensions – their price, distance from the campus, and friend's opinion. In order to form skewed sets, the authors fixed two of the dimensions and varied the density around the lower or upper extremes of the range of the third dimension. In all conditions, the frequency effect was demonstrated.

Frequency effect is a manifestation of the tendency for people to use all available ratings an almost equal number of times (Parducci & Perret, 1971). This increases the precisions of judgments – it is easy but not informative to give to all stimuli one and the same rating.

Contrast effects in judgment can be observed depending on the anchors, the range and the skew of the stimulus distribution. There are other specific judgment situations, however, in which contrast effects appear. In the next section, some of them are presented.

2.2.4 Contrast from a Primed Stimulus

Herr, Sherman & Fazio (1983) asked participants to judge the size of animals. The animals were presented with the words that name them. Moreover, some of the names were non-words, but participants were instructed to judge the animals, even if they did not know them. Before making their judgments, participants were primed with different animals. As was described above, an assimilation effect appeared when the context animals were moderate exemplars and the judged stimuli - ambiguous. In the other cases, however, contrast effect was demonstrated. When the judged animals were unambiguous and well known, regardless of the size of the priming animals, the ratings shifted away from it. When the judged stimuli were ambiguous, the contrast effect appeared only when the priming animals were extreme in their size.

This experiment illustrates two different contrastive forces. On the one hand, people tend to shift their ratings away from a certain contextual element when the judged dimensions are clear and unambiguous. On the other hand, independently of the type of the judged stimulus, when the contextual elements are too extreme, again contrast is observed. The latter tendency may be viewed as an additional manifestation of the range effect. Not only the judged stimuli but also some available items of the same category could change the range.

For example, Manis, Nelson & Shedler (1988) induced in participants a stereotype about a patient's hospital – the patients were predominantly high pathological. A second hospital was presented, in which the patients were moderate pathological. People tended to judge one and the same case definition as more pathological, if it was said that the patient came from the second hospital than if he was from the first one. This contrast effect, however, appeared only when the stereotype about the first hospital was very extreme. Otherwise, an assimilation effect was observed.

Furthermore, the shifts in ratings when priming some items depend on whether concrete examples or whole categories serve for primes. For example, Stapel and Kooman (1996, 2000) asked people to judge different types of social stimuli – in different studies, they used trait names, actor's play, and sentences, which describe different behaviours. The authors demonstrated contrast from a primed concrete exemplar. As was mentioned in the previous section, the opposite assimilation effect arises when the priming is for general schemes.

In the previous section an experiment by Strack, Schwarz & Gschneidinger (1985) was described, in which the authors asked people to list three positive or negative recent events, and after that to judge their overall satisfaction with their life. In the same study authors asked participants in two additional groups to do the same, with the only exception that the listed events were not recent, but from the last ten years. In this case, the results reversed – people that recalled positive events, underestimated their life satisfaction, and vice versa.

This experimental result also could be interpreted as confirmation of the hypothesis that primed concrete exemplars produce contrast. Because the task of people was to rate their overall satisfaction, the examples from the last ten years are concrete exemplars – they could serve as a basis for comparison. In the opposite, the examples from the recent past are part of a relatively small sub-category of the whole life.

In conclusion, a certain item that serves for priming or is presented in the environment may cause the ratings of some stimuli to shift. This shift tends to be in a direction away from the contextual item when the judgment task is unambiguous, and when the contextual element could be recognized as an extreme exemplar of the category of the judged stimulus.

2.2.5 Summary of the Contrast Contextual Effects

The tendency to shift judgments away from a particular contextual element is called contrast effect and has been demonstrated in number of experiments. Contrast effect tends to be observed:

- when the contextual item is outside the range of the stimuli distribution or near its boundaries (Sherif, Taub, Hovland, 1958), (Manis, Nelson & Shedler, 1988);
- away from the priming stimuli, when the judged stimulus is not ambiguous (Herr, Sherman & Fazio, 1983);
- when the judged set is skewed (Parducci & Perret, 1971);
- away from the prototypical value for the category of the judged stimulus in the case when the task is to judge on a subjective scale (Marks, 1988).

Together with the assimilation and contrast effects, however, psychological studies report some other systematic shifts of ratings when people judge on a scale.

2.3 Other Contextual Effects in Judgment

2.3.1 Rating Preferences

Petrov & Anderson (2005) illustrated that the response distribution has a peak in the middle, even when the stimulus distribution is uniform. This can be numerically measured as the decrease in the variance of response distribution compared to the theoretical uniform one. This tendency - avoiding extreme ratings - is robust. It was also reported by Kokinov, Hristova & Petkov (2004).

It seems that this tendency is in contradiction with the frequency effect (the latter assumes that all ratings tend to be used an almost equal number of times). This contradiction is an additional challenge for the theories of judgment. The manifestations of the frequency effect, however, appear when a skewed set of stimuli is judged, whereas the non-uniform distribution of the ratings could be observed even when a uniform set of stimuli is judged.

2.3.2 The Role of Time Delays

Srull and Wyer (1980) investigated the influence of time between the priming and the target presentation. Their stimuli were trait concepts and paragraphs that describe a particular personal behaviour. The contextual stimulus, the target one, and the requested judgment of the target were positioned subsequently in time. The authors demonstrated that the assimilation effect decreases with the increase of the delay between the perception of the contextual stimulus and the target one. At the same time, the assimilation effect increases with the increase of the delay between the perception of the target stimulus and the moment of response[1].

Therefore, judgment (as well as cognition as a whole) seems to be a dynamic process. Hence, dynamic mechanisms should be proposed in order to describe the phenomena.

2.3.3 The Role of the Number of Scale Labels

Wedell, Parducci & Lane (1990) investigated the sizes of range and frequency effects when the scales vary. They asked people to judge the pathology of case definitions on 3-rate, 7-rate, and 100-rate scale. The authors concluded that the frequency effect decreases when more ratings are available on the scale. The range effect could decrease too, but only if each category (or at least the two extremes) was anchored with a verbal label or with an example.

This seems to be an important constraint for the mechanisms that may be considered responsible for the frequency effect. These mechanisms should capture the dependency between judgments and the number of the scale labels.

2.3.4 'Elasticity'

Hsee (1996) introduced a new interesting term – "*elasticity*". He noted that sometimes new, unexpected factors play an important part in the judgment task. The author gives the following example: Imagine a competition between two piano-

[1] The authors first presented the stimulus, and after a delay asked people to give a rating.

players. The first one plays a piece of music that is the more difficult one, but the second pianist makes less errors. The judge may never admit it but perhaps the crucial factor for his judgment is the fact that one of the players is from his country. According to the author, such new factors do not always arise but only when the main factors are "elastic", i.e. it is possible to interpret them in different ways. In the example, how does one compare the complexity versus the number of errors? In this example, the elasticity is in the combination of complicated piece of music and errors.

2.4 Contextual Effects in Choice

This section briefly reviews a few contextual effects demonstrated in choice situation.

JUDGEMAP model assumes that the tasks for judgment and choice are very close and that the same mechanisms underlie them. A similar view was expressed by Medin, Goldstone & Markman (1995). They argue that maybe there are common mechanisms underlying choice and similarity judgments. According to the authors, some of these mechanisms can be: generation of alternatives, establishment of correspondences between items, requirement of reference points, and justification of judgments. JUDGEMAP developed analogical mechanisms for judgment. That is why some contextual effects in choice are briefly presented here, and discussed later in the simulations.

Kahneman & Tversky (1979), Kahneman & Tversky (1984) demonstrated their famous examples of preference reversals because of pre-formulation of the task. When describing two social programs in terms of saved lives, one of them was preferred, whereas if the same two polices were described in terms of lives lost, the other one is preferred.

In addition, preference reversals could also result from presenting a third alternative to the two basic ones (Tversky, 1972; Shafir, Simmonson & Tversky, 1993). Participants were asked to choose between two gambles – the first one promised a higher profit, but with lower probability than the second one. The two gambles were adjusted so that each one was preferred by almost half of the people. The authors gave to another group of people to chose among three alternative gambles – the same two from the first condition, and a third one, which was less attractive than the first one both in its profit and probability to win. In comparison with the second gamble, however, the third one is better in one of the dimensions. Of course, nobody had chosen the third alternative (it was worse than the first one in the both dimensions), but now much more people chose the first gamble than the second one.

2.5 Summary of the Contextual Effects in Judgment

The following is a brief review of the main data from the psychological experiments on judgment:

First, judgments seem to be contextually dependent – they depend on the *whole set of given stimuli* and *their location in time* (their chronological order), on the attributes of the stimulus used in judgment (called also *stimulus levels*), on the *formulation of the task*, on the *previously solved tasks*, on *biases* primed from the experimenter, and on the *time of contextual manipulation*.

Second, the main observed effects of this dependency are *assimilation* and *contrast*. These terms are relevant only when both the target and the contextual stimuli can be judged on the same scale. Sometimes, the two opposite effects can be observed in experiments, performed under almost equal conditions, differing only in the type of stimuli, or in some details of the experimental procedure.

Usually, presentation of a *single contextual element* when judging unambiguous dimensions of the stimuli (no matter whether it is anchored with a rating or not) causes contrast effect (Saris & Parducci, 1978, Parducci & Marshal, 1962, Sherif, Taub, Hovland, 1958). In the opposite case, when the judged properties of the stimuli are too ambiguous, assimilation is observed (Wedell, Parducci & Geiselman 1987; Anderson, 1974; Herr, Sherman, Fazio, 1983). Sometimes, however, exceptions occur. For example, when the task was for judgment of weights and the anchor was exactly on the boundaries of the stimulus set (Wedell, Parducci & Geiselman 1987), again assimilation was observed. This exception is difficult to explain – maybe the tested modality is special in some way.

The influence of the whole set of the stimuli on judgment could be expressed with the tendency of the judgments to satisfy both the *range principle* and the *frequency principle* (Parducci, 1968).

The *sequential assimilation effect* (Lockhead, 1992) is also robust and has been observed in almost all experiments involving judgments of series of homogeneous stimuli.

The type of the scale also influences judgments. Parducci demonstrated that increasing the *number of the scale values* decreases frequency effect.

None of the existing theories can explain simultaneously all these pressures (JUDGEMAP does not make an exception). However, they highlight different aspects of the judgment task. The most famous theories and models of judgment, their contributions and shortcomings, are discussed in the next chapter.

CHAPTER III
Main Theories and Models of Judgment

3.1 Classification of the Theories of Judgment

The cognitive theories differ in the level of description at which they try to explain known phenomena.

Some explanations are mathematical in their form. They try to describe the final result as a function of the input – the given task and the surrounding context. These theories describe the phenomena on the level of input-output correspondences and are usually called *normative* theories.

When a certain bias (i.e. systematic error) is found, the researchers try to explain it. In this way *descriptive* theories are formed – they describe some heuristics that people use in solving tasks, and suggest an explanation how those heuristics are responsible for the systematic shifts in the responses.

Another type of theories (JUDGEMAP is included in this group) assumes that cognition is an emergent product of local interactions between a large number of units that share a small number of basic principles. From this point of view, each cognitive process should be treated not in isolation from the others. Instead, it should be based on a cognitive architecture, i.e., on a limited number of fixed basic principles and mechanisms. The contextual effects then should not be specially modeled, but they should emerge from these low-level principles.

Independently of their form, the theories and models of judgment could be separated in three main groups, according to their point of view about the nature of the judgment. The first point of view is that judgment is a process of *measuring the similarity/dissimilarity to a standard*. A second group of approaches looks at the judgment as a process of *classification*. Each rating forms a subcategory and the target stimulus has to be classified into one of those subcategories. The JUDGEMAP project assumes a third point of view. The target stimulus is included in the comparison set and then a *mapping* between the elements of the comparison set and the set of scale elements is established. This mapping should satisfy as much as possible the structural constraints specified in the instruction – higher stimulus magnitudes should receive higher ratings, and almost equal differences between magnitudes should correspond to almost equal differences between the corresponding ratings.

3.2 Judgment as Measuring the Distance to a Standard

The theories that treat judgment as a process of comparison fall in this group. According to them, the magnitude of the judged stimulus is compared to a certain standard. The theories, however, differ in the nature of this standard – where it comes from, how stable it is, whether the external context influences the standard, etc.

3.2.1 Psychological Scaling and Ideal Point Theory

Suppose that the experimenter presents some stimuli to a subject and asks him/her to judge one of their simple one-dimensional properties (like weight, sweetness or length). When a certain stimulus is given, in the observer's mind a psychological effect springs up, i.e. he/she perceives the stimulus. One fundamental hypothesis in classical psychophysics is that these psychological effects form a continuous and ordered space – the so-called 'psychological continuum'. This assumption underlies one of the main goals of psychophysics - to map the physical continuum with the correspondent psychological one. After this is done, a linear transformation transforms this psychological continuum to a given scale.

This methodology is called psychological scaling. The experimental results, however, show that even during a single judgment session people often give different ratings to the same stimulus, which contradicts the formula. According to the psychological scaling approach, this is due to the imprecision of our perceptions and thinking. This imprecision could be treated as a source of noise. Hence, one random value has to be added to the predicted by the model response.

In the case of judgment of subjective, ambiguous properties (e.g. female faces, sweetness of drinks, etc.), the Ideal Point Theory represents psychological effects again as points in continuous ordered abstract psychological space. One additional assumption, however, is that people keep in mind one ideal stimulus and the preferences are calculated as a function of the distance between the attribute values of the given stimulus and the ideal one.

Wedell and Petibone (1999) propose an improvement of the mathematical equations of Ideal Point Theory, explaining some non-symmetric effects of equivalent deviations above and below the ideal point.

Overall, both Psychological scaling methodology and Ideal Point Theory, in their classical form, do not consider any role of the context in the process. Thus, they fail to explain most of the experimental results that demonstrate systematic shifts in judgments.

3.2.2 Adaptation Level Theory (Helson)

Adaptation Level Theory (Helson, 1964) was first proposed in the field of perceptions. The fact of adaptation of perceptions has been well known for a long time. The sensory systems react only to change (the only exception is pain). When the irritating factor is constant, the psychological effect falls down to zero.

The assumptions of the theory, however, have been used both in the field of perception and in the task of judgment. According to them, the psychological effect of the entire context can be represented by the so-called adaptation level (**AL**) – just a single value, the "average" (for example, 4 in the 1-7 scale). After the **AL** is calculated, the judgment **J** is equalized to the ratio between the stimulus value being judged and the adaptation level **AL**. The **AL** value itself is given as a geometric mean of various factors: the value of the target stimulus, the value of the manipulated contextual stimuli, and whole context (the room in which the experiment is conducted, the people in it and so on). Each factor has its own relative power.

The Adaptation Level Theory always predicts contrast effect. In order to defend the theory and to use it even when the experimental data shows assimilation, Restle (1978) proposes its specification. He argues that it is wrong to assume that the weights of the factors are constant and independent from the context, studying some perceptual illusions with simple geometrical figures. He observed that when the contextual stimulus was very small, an assimilation effect appeared; when the contextual stimulus grew, this effect decreased to zero, turned into contrast, and the latter increased. Fitting the parameters of the equation with the experimental data, Restle proposes one quite complex but precise function.

A big advantage of the Adaptation Level Theory is its attempt to explain the contextual sensitivity of judgment. Not only the magnitude of the judged property, but also its relative position with regard to the whole set of stimuli, reflect judgments. The theory also tries to establish a bridge between perceptions and high-level cognitive processes (like judgment), assuming that the same principles underlie them.

The theory, however, does not provide a deeper explanation of the controversial assimilation and contrast effects. It uses too abstract and ambiguous terms (like 'the whole context'). This makes its experimental testing difficult.

3.2.3 Norm Theory (Kahneman & Miller)

A different point of view is expressed by Kahneman & Miller (1986). They suggest their Norm Theory - one constructivist idea based on the exemplar model of memory. According to it, the stimulus is compared with a norm. Instead of the classical Ideal–point theory, where the ideal point is constant and is stored in the memory in advance, the Kahneman's norm is constructive. The available exemplars, which the stimulus retrieves, create this norm. How does this mechanism work? Let only a single stimulus be given. It must create its own context. Because the number of its attributes is huge (maybe infinite), a small set of such attributes needs to be viewed like important. More probably, the attributes, which identify the stimulus, are the most important, i.e. these attributes, which are not central for the whole set of stimuli. "The general rule is that, other things being equal, the more mutable and less important of two attributes will have a disproportionately large effect in single–stimulus judgment" (Kahneman & Miller, 1986, p.142). On the contrary, when a set of stimuli is given, the identifying attributes are those, which are common for the whole set (they identify the set, no matter that a single stimulus is judged). Kahneman and Miller justify this idea by presenting data about the preferences of people to some bets.

In order to define the concept "norm", the authors stress that "normality" and "probability" are not the same thing. For example, compare the two sentences: (a) "The favorite lost the first set." and (b) "The favorite lost the first set, but won the match". Obviously (a) is more probable, but is also more surprising, hence – abnormal. According to the authors, the norm is exactly the opposite of the surprise.

The Norm Theory and DUAL cognitive architecture share the same idea about the constructive nature of the memory. The Norm Theory states that each stimulus is compared with a norm, which is yielded from a set of retrieved or constructed memories. However, JUDGEMAP proposes a different mechanism. Each stimulus is included in a constructed comparison set of stimuli that varies dynamically in time. The content of this set is: the target stimulus, the recent or primed stimuli, the typical stimuli (if they exist), and the closer to the already active stimuli (i.e., the stimuli, whose magnitude of the rated property is close to the ones of the already active stimuli). The rating of the target stimulus emerges from the mapping between this set and the required scale for judgment.

The next two subsections present respectively two models that assume comparison–making as a basic mechanism in judgment. According to these theories, however, instead of comparing the target stimulus with a single standard or norm, the response arises from many local comparisons between the target and other memory traces.

3.2.4 Comparison–Based Judgments (Mussweiler)

Mussweiler (2003) proposes a descriptive model of judgment, based on comparisons between the target stimuli and retrieved ones from the memory. According to the model, two different parallel processes are involved in judgment. One of them is responsible for assimilation effects, the other one – for contrast ones (fig. 3.1).

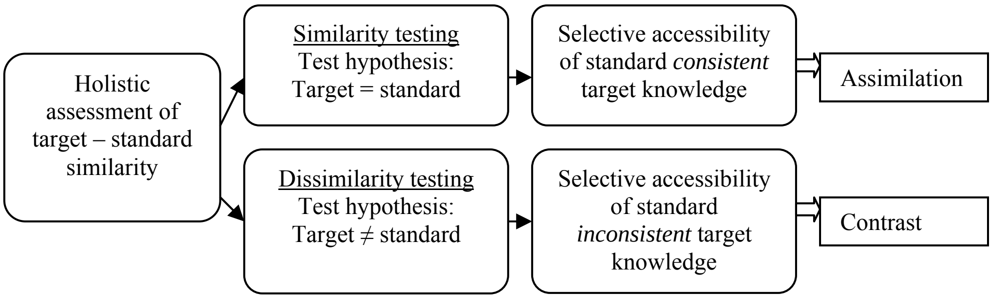

Fig.3.1. Comparison – Based Model of Mussweiler (taken from Mussweiler, 2003).

Mussweiler assumes that it is impossible to make judgments without context, i.e. using only the information about the judged magnitude, without anything to compare it with. It is necessary first to categorize it, to extract from memory other exemplars, and then to compare these memories to the target one. The author proposes two different kinds of such comparisons. On the one hand, the system checks the hypothesis that the target is similar to the standard. During this test, the common features become more salient and this creates a pressure to rate the target and standard with close ratings. On the other hand, the alternative hypothesis is checked in parallel. The features that differentiate target from the standard increase their attractiveness, and an opposite pressure arises. Thus, the final decision is biased from two conflicting forces and the kind of effect that arises depends on the strength of the arguments of similarity in comparison with the argument of dissimilarity.

Mussweiler's model and JUDGEMAP share the assumption that judgment is based on comparisons. The former, however, cannot determine in more detail the exact circumstances in which one or another mechanism wins. It fails to explain what the principal difference between the previously judged stimulus and the anchoring stimulus is. Why assimilation toward the former but contrast to the latter occurs?

3.2.5 Two-Path Model (Manis & Paskewitz)

The two–path model (Manis & Paskewitz, 1984, Manis, Nelson, Shedler, 1988) tries to explain a certain group of biases, observed in experiments. It is similar to

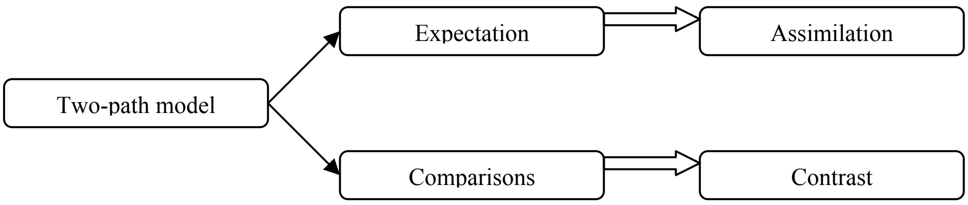

Fig.3.2. The two-path model of Manis & Paskewitz.

Mussweiler's model with proposal of two parallel conflicting pressures in judgment (fig.3.2).

According to the model, one of the driving forces in judgment is expectation. The cognitive system always has some expectations about the nature of the next stimulus. These expectations are based on stereotypes, recently met exemplars, etc. When the real stimulus becomes available, the system shifts it a little toward the expectation. This shift causes the assimilation effects in the final response.

In addition, the system makes comparisons between target and other exemplars, retrieved from memory. Contrary to Mussweiler's model, Manis and Paskewitz assume that comparisons always produce contrast effect.

The two-path model fits many empirical data excellently. Unfortunately, it fails to explain some other phenomena, e.g. why contrast appears when the context stimulus is outside or near the range boundaries, why sequential assimilation appears with respect to the previous ratings, but not to the previous stimuli.

3.3 Judgment as a Classification Task

The theories that treat judgment task as a classification task fall in this group. According to these theories, the category of the judged stimulus is subdivided into several subcategories. Each of these subcategories consists of exemplars that should be judged with a certain scale label. Thus, when people judge a certain stimulus, they try to find the most proper of these subcategories.

3.3.1 Range–Frequency Model (Parducci)

Allen Parducci accepts the relativistic idea that the judgment of a concrete stimulus depends on the entire set of presented stimuli and proposes his "Range-Frequency Model" (Parducci, 1965, 1968, 1973; Parducci & Perret, 1971, Wedell, Parducci, Geiselman, 1987). For several decades, he has improved and tested it with different stimuli – perceptual (Parducci, 1962), moral (Parducci, 1968), tasks of preference (Wedell, Parducci, Geiselman, 1987) and games (Parducci, 1973).

The Range–Frequency Theory is based on the competition and compromise between two principles:

The Range principle asserts that the judge uses the given categories to subdivide the psychological range. He/she fixes the two extreme values of the stimuli in the set and assigns the two extreme categories to them. After that, the set is subdivided into the necessary number of categories.

The Frequency principle asserts that the judge uses each category for a fixed proportion of his judgment. These proportions may vary for the different sets of categories but we can often assume equal proportions.

The two principles are in conflict when stimuli from the different parts of the range are presented with unequal frequencies. This conflict can be solved with a simple linear weighted sum:

$$\mathbf{J_i = wR_i + (1\text{-}w)F_i}$$

This equation expresses the Range–Frequency Model. Here $\mathbf{J_i}$ is the mean judgment of the **i**-th stimulus; $\mathbf{R_i}$ and $\mathbf{F_i}$ are respectively the range value and frequency value of the **i**-th stimulus and **w** is a weighting constant.

Frequency value is the mean of the grades that a given stimulus would elicit if each category was used with equal frequency and can be easily calculated. One needs to divide first the total number of stimuli presentations into the number of available categories. After that one needs to count by rank, to which category a particular stimulus should be assigned.

The Range value $\mathbf{R_i}$ is determined by the relationship between the presented stimulus and the two extreme stimuli defining the two extremes of the scale. It depends just on these three things. It cannot be calculated a priori without the actual data. Each $\mathbf{R_i}$ can be extracted by a substitution of $\mathbf{J_i}$ and $\mathbf{F_i}$ in the equation. This method may look strange, but the idea is that such a value can be used to predict the judgment of the same stimulus in another set (but with the same two extremes).

In the first version of the model ("Limen Model" - Parducci, 1965), Parducci presumed that according to the range principle the set of the stimuli was subdivided into the needed categories according to the principle of equal discrimination. However, after collecting huge experimental data, using squares with different sizes, he rejected this idea.

Parducci also studied the role of weight and found out that the relative power of the frequency principle decreases when the number of available categories increases.

In conclusion, the Range-Frequency Model is one of the first theories that underlie the important role of the context in judgment. In addition, the two principles, developed by Parducci are simple and clear, and at the same time, their validity is manifested in almost all experiments, independently of the type of stimuli that are judged.

However, the Range–Frequency Model does not predict an assimilation effect in any condition. In order to explain the assimilative results, when two target stimuli are presented simultaneously, Parducci assumes that a second assimilative force influences judgment.

Being a normative theory, the Range–Frequency Model does not propose any mechanisms that could be responsible for the systematic shifts of the ratings depending on the distribution of the stimulus set. In the next subsection, a computational model for judgment is presented.

3.3.2 The ANCHOR Model (Petrov & Anderson)

The ANCHOR model (Petrov & Anderson, 2000) is a computational model of judgment and absolute identification (this task is similar to judgment, but after each response, a feed-back signal is given about the right answer.), based on ACT-R architecture (Anderson & Lebiere, 1993). It views judgment as a two-stage process. At the first stage, the perceived magnitude of the target stimulus is compared to a set of anchors in memory. Each anchor corresponds to the prototype of the subcategory of stimuli and its corresponding rating label. At the second stage, an explicit correction strategy shifts a little the final response.

The magnitude of the judged property is transformed into an internal representation via exponential transformation and by adding some noise. Then this internal representation is mapped into one of the anchors in a stochastic manner. The probability to map onto a certain anchor is sensitive to similarity (between the target and anchor), anchor's base-level strength, and whether it was recently used. After an anchor has been chosen, a correction mechanism may shift the response with one or two grades in both directions. The correction reflects the difference between the target and anchor's magnitudes, and the appropriate knowledge about the approximate category width. Finally, a learning algorithm updates the locations and base-level activations of the anchors.

The ANCHOR model does not only describe the input–output correspondences, like normative theories. It does not describe in abstract terms some biases, and possible explanations for them, as the descriptive theories do. Instead, it is a complete computational model, whose starting points are assumed fundamental for all cognitive processes (namely, the principles of ACT-R architecture). The ANCHOR model demonstrates how many contextual effects – sequential assimilation, non-uniform distribution of the responses, systematic changes during the time of judgment, etc. - could emerge from these low-level mechanisms.

Unfortunately, the parameters in ANCHOR model should be fixed on a certain scale and certain type of stimuli, for example, in the reported simulations, the scale is a 9-point one and the stimuli are line segments. The model does not propose more general templates, which are independent from the scale and the concrete stimulus category. In addition, the model cannot explain the contextual effects, caused by properties of the stimuli, which are irrelevant to the task.

3.3.3 Descriptive Theories that Treat Judgment as Classification

In this subsection, three descriptive theories of judgment are briefly presented, together with their advantages and shortcomings.

The Integration theory proposed by Anderson (1971) assumes that some proportion of the overall impression, i.e. the stimulus together with its surrounding context, is induced onto the stimulus rating. For example, if a group of people were perceived, part of the overall evaluation of the group would be added to each individual member. Anderson calls this influence a 'halo effect'.

A similar approach was proposed by Wyer (1974). He assumes that if the perceived stimulus is ambiguous, then the category, in which it would be categorized, tends to be the most accessible and applicable one. A possible explanation of this phenomenon could be the neural net approach. Here the stimulus is represented by a set of associated semantic nodes and the contextual stimuli prime some of those nodes. Because according to this approach the meaning is coded through the overall pattern of activation across the nodes, Wyer calls this priming *"change of meaning"*.

Both the Integration Theory and the Change of Meaning approaches always predict assimilation toward the contextual elements. The next model in this group tries to explain both assimilation and contrast effects in a single frame.

According to the Inclusion – Exclusion Model (Schwarz & Bless, 1991), the default option in judgment is to include the accessible contextual information in the representation of the target. This gives rise to assimilation effects. Sometimes, however, it is possible to do exactly the opposite – to exclude the outside information that comes to mind from the representation of the target, and hence contrast would be observed. This may happen because this information comes for irrelevant to the task reason, because it does not belong to the category of the target, or because the subject was explicitly instructed not to use it.

The descriptive theories that treat judgment as a process of classification share with JUDGEMAP the idea that judgment cannot be separated from the overall context. However, the focus of the model differs from the focus of the presented in this subsection approaches.

3.3.4 The Role of Perception (Goldstone)

Goldstone (1998) highlights the role of perceptions in the process of judgment. He argues that there are many known phenomena in perception that could give rise to some of the contextual effects in high-level tasks like judgment.

He proposes four mechanisms that are merged in the theory of perceptual learning: By Attentional weighting, the relevant dimensions and features become more attractive, and the irrelevant ones become less attractive. By stimulus imprinting, precise detectors for the important stimuli and parts of stimuli are developed. By differentiation, the perceptual system becomes more accurate in distinguishing features that were undistinguishable before the respective task was presented. Finally, by unitization the system becomes able to detect whole constructions together with their parts.

One advantage of the theory is that it points to the important role of perception in judgment. Unfortunately, the theory lacks mechanisms that are more concrete and cannot demonstrate how these pressures can be modeled to work together in synchrony.

However, JUDGEMAP, since its focus is on different aspects of the judgment process, treats the mechanisms and principles, proposed by Goldstone (1998) as complementary ones. If a single model could merge them, this would increase the power of its explanations and predictions.

3.3.5 Exemplar–Based Random Walk Model (Nosofsky)

Nosofsky & Palmeri (1997) proposed the Exemplar–Based Random Walk Model (EBRW) for classification. As in the Anchor model, each scale label is associated with a prototype and all prototypes compete to match the target. This competition is resolved via prolonged random walk – at each step, one of the hypotheses increases its strength, and the other decreases it.

The memory in the model is exemplar-based. When a stimulus is given to be classified into one of two categories, the exemplars in memory are retrieved one after the other with races, determined by their similarities to the test item. A random walk counter with start value zero, increments a little in positive direction if evidence to categorize the target in the first category arise, or in negative direction, if the evidence is in favor of the second category. When the counter reaches one of the two predefined thresholds (criteria), the system gives a response.

EBWR shares with the JUDGEMAP model the idea that judgment could be treated as a result from a competition between alternative hypotheses. The two models differ, however, in that the mechanisms of EBWR are expressed only by mathematical equation, while JUDGEMAP, being designed on a cognitive architecture, integrates the judgment process with other cognitive abilities.

3.4 Summary of the Theories of Judgment

All theories and models make their own contribution and highlight the problems from different perspectives. It is difficult, however, to integrate all phenomena, known about human judgments in a single model. Unfortunately, none of the contemporary theories can explain satisfactorily the whole range of systematic shifts in judgments (JUDGEMAP is no exception).

Ideal Point Theory (in its classical form) does not give rise to any contextual effects at all. Adaptation Level Theory and the Range–Frequency model predict contrast effect in all cases, whereas Integration Theory and the Change-of-meaning approach deal only with assimilation effects.

Some theories propose the existence of two different pressures and treat judgment as a result of compromise between them. Comparison–Based Judgments and the Two– paths model are in this group.

Norm Theory and the Perceptual Learning approach propose more integrated mechanisms for judgment. Unfortunately, they deal with too abstract terms, do not propose any concrete mechanisms, and thus, it is difficult to verify them.

Finally, the ANCHOR model and the EBRW model are focused only on a narrow part of the problem – they put too many limitations on the type of the stimuli and of the scale.

CHAPTER IV
JUDGEMAP in Broad Strokes

4.1 Main Ideas

In contrast to the theories described in the previous chapter, I view the process of judgment as a process of mapping of a set of stimuli onto the set of scale elements. In other words, the judge always needs a whole set of elements to be rated – even if he/she has to evaluate one single stimulus he/she has to construct a set of elements to be evaluated that include this target element and then maps the whole set onto the scale (Figure 4.1.). This is in sharp contrast to the theories discussed so far that

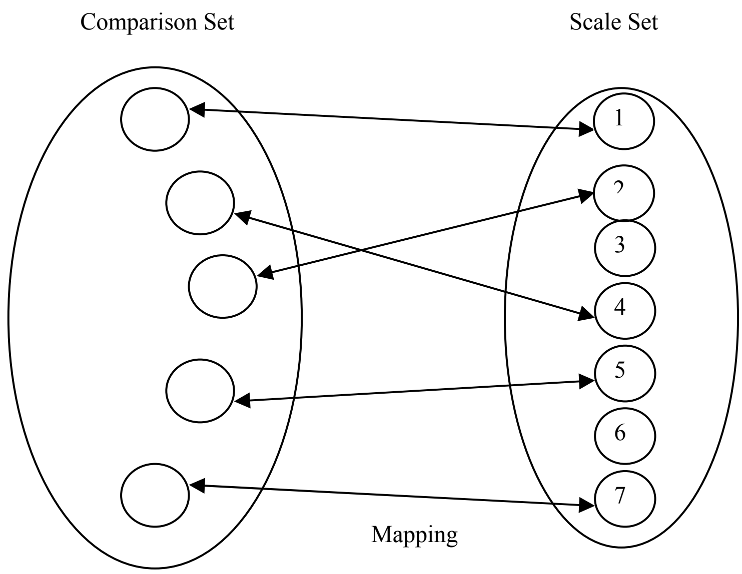

Figure 4.1 Judgment as mapping.

assume the presence of a standard or prototype or stereotype or anchors of the categories – in all these cases one can simply compare the target with this standard/prototype/anchor and as a result obtain a rating. JUDGEMAP approach makes no assumptions about the existence of such a centralized and static representation of a standard and thus requires dynamically constructing the comparison set. This peculiar characteristic of the current approach makes it unique in terms of its high context-sensitivity since the formation of the comparison set (which is crucial for the specific rating the target will obtain) is dynamic and can be influenced in many ways.

Thus the process of judgment can be described as consisting of two subprocesses:

- formation of the comparison set, and
- mapping the comparison set onto the scale elements set.

The comparison set is supposed to consist of those elements whose representations happen to become part of the Working Memory of the judge on the particular occasion. There are two main sources: comparison set elements may come from perception (if the judge encounters other elements in the environment) and from memory (if the target element reminds the judge of some previously encountered elements). The perceptual mechanisms are not modeled in the current version of JUDGEMAP-2, but the spreading activation mechanism of the DUAL cognitive architecture is a good candidate for explaining the process of “reminding”. Moreover, DUAL has already been successfully used for modeling memory processes (Kokinov, 1998, Kokinov & Petrov, 2001, Kokinov & Zareva-Toncheva, 2001).

The mapping process has to preserve the structure of ordering relations in the comparison set when finding their corresponding elements in the scale elements set (Figure 4.2.). Thus, this process seems very similar to the process of mapping in analogy-making (which always preserves the structure of relations). Therefore, the mapping mechanisms developed within the AMBR model of analogy-making (Kokinov, 1988, 1994a, 1998, Kokinov & Petrov, 2001) can be potentially useful for judgment as well.

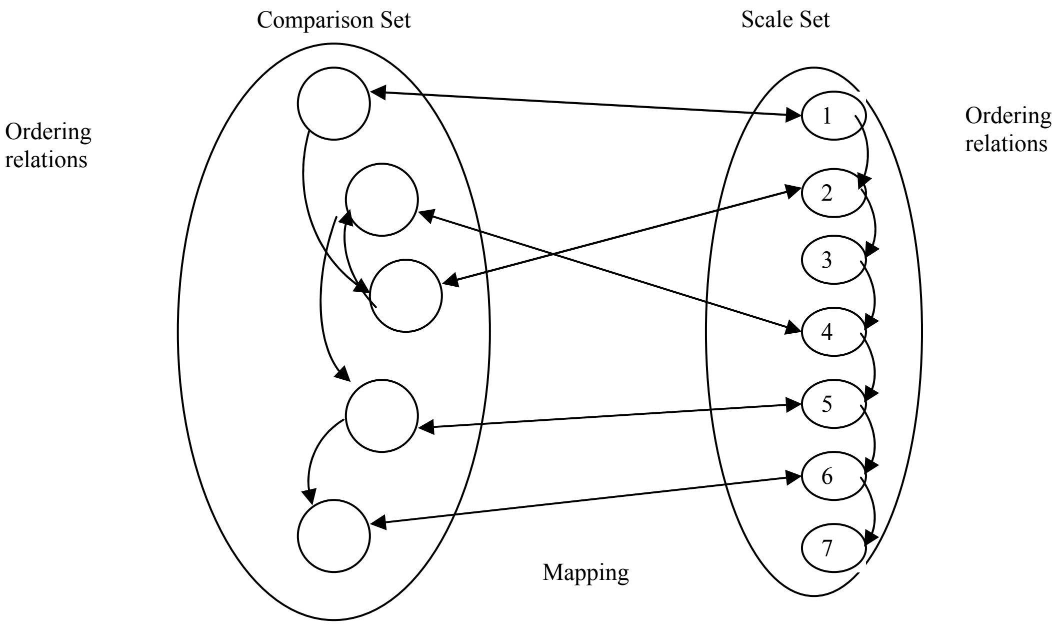

Figure 4.2: Mapping the relational structure of the ordering relations.

Thus, a natural idea would be to use the DUAL cognitive architecture (Kokinov, 1994b, 1994c) that has already been successfully used for modeling analogy and memory to build a model of judgment as well. Moreover, starting the modeling process from a general cognitive architecture has several important consequences:

- The principles of DUAL and its mechanisms and representations have been developed having in mind other cognitive processes and if these can be successfully applied to modeling judgment, this would be a good test for DUAL as a general cognitive architecture.
- The process of modeling will be severely constrained: one cannot build any possible mechanism and representation that would be simply useful for fitting

the experimental data. The model should keep the principles of DUAL and be based on the already developed mechanisms – new mechanisms can be build only if necessary and with great care. In this way, fitting the data is not a simple process – the data should rather naturally arise out of the embedded DUAL principles and mechanisms.

- The process of judgment will not be in isolation – it will be naturally integrated with other cognitive processes such as analogy, memory, and in the future with perception and learning. This will also allow for modeling and explaining the interactions between various cognitive processes.

Having reviewed the background information, it is now time to formulate the aims of the current dissertation.

4.2 Aims of the Current Dissertation

The main aim of the current dissertation is to build a computational model of human judgment, based on the cognitive architecture DUAL. The model should also use as much as possible the already developed mechanisms of mapping within the AMBR model of analogy-making and thus integrate judgment with other cognitive processes such as memory and analogy. The model should account for the contrast and assimilation effects described in chapter 2 and possibly make new predictions.

The model should also become a basis for future development of a model of decision-making based on DUAL.

What follows is a brief presentation of the cognitive architecture DUAL that serves as a basis for building the model and the related AMBR model of analogy-making.

4.3 DUAL Architecture

4.2.1 Basic Properties

DUAL is a cognitive architecture, launched by Kokinov (1994b,c). It consists of memory structures and processing mechanisms, organized around the following principles:

- *Hybridity* – DUAL combines the symbolic with connectionist approaches, by integrating them at the micro-level.
- *Emergent computations* – the global behavior of the architecture emerges from local interactions among a huge number of small entities, called DUAL-agents. There is no central executor that monitors the whole system.
- *Dynamics and context–sensitivity* – The behavior of DUAL changes continuously in response to the influence of the dynamic changes in context. There is no clear-cut boundary between the task and its context. Instead, the context is assumed to be the state of the system in any certain moment, i.e. the pattern of activation over the set of DUAL-agents. This pattern is assumed to reflect the relevance of the various pieces of knowledge in the current context. Some DUAL-agents might be relevant because the corresponding elements are currently perceived from the surrounding environment, others – because they reflect the

current goals of the system, and finally, some agents might be relevant because they were recently used and have some residual activation or are linked to recently used agents.

4.2.2 DUAL – agents

The basic structural and functional element in DUAL is the DUAL – agent. It is hybrid in two ways – it has both connectionists and symbolic aspects, and it serves both as a representational and a functional unit.

4.2.2.1 Connectionist's Aspect

From the connectionist's perspective, each agent is a node in a localist neural network. It continuously receives activation, updates its current activation level, and spreads it through associative links to other agents. An important feature of DUAL is that it distinguishes the semantic meaning of the agents from their relevance, considering them as independent. The *activation level* is a numeric value that codes the *relevance* of each agent. The pattern of activation does not represent any concept or scheme, but just the current context.

4.2.2.2 Symbolic Aspect

From the symbolic point of view, DUAL – agents are organized in a semantic network. Each agent 'stands' for something – an object, property, relation, etc. It has its own *micro-frame*. The micro-frames have *slots*, which in turn may have *facets*. There are two kinds of slots – general ones (*G-slots*), and frame-specific ones (*S-slots*).

```
Line
  :type :concept
  :subc (geometrical_figure 1.000)
  :slot1
    :type :aspect
    :c-coref (color-of-12 0.500)
```

Fig. 4.3. An example of a DUAL-agent

G-slots have labels, the meaning of which is invariant across the agents. Consider the example on fig. 4.3. The first line depicts the name of the agent – 'line'. The next two lines represent two different G-slots. The first one has a label *:type*. Its filler is a tag that denotes the type of the agent – :concept. The label *:subc* of the second G-slot points to the super-class of the agent. Its filler is a *reference*. The references are associations that consist of two parts. The first one is a name of another agent, whereas the second one concerns only the connectionist aspects. It is a real number that depicts the *weight* of the reference[1]. The whole activation that the agents send is distributed to all references, normalized by their weights.

The remaining part of the agent 'line' consists of one S-slot - *:slot1*. S-slots also have labels, but their labels are arbitrary, i.e., :slot1 in one agent may mean something very different from :slot1 in another agent. S-slots (and only S-slots) have facets, i.e., slots within slots. The same set of labels applies to both G-slots and facets.

[1] When the connectionist aspect is highlighted, the references are called *links*.

4.2.2.3 Symbolic Processing

DUAL – agents interact with each other. These interactions are relatively simple – they always involve two agents – one of them *sends* some information, and the other one *reads* it.

Each DUAL – agent has a *symbolic processor*. It can receive or construct symbolic structures, transform them, store them in its own *local memory*, and send them to neighboring agents. A typical symbolical transaction involves receiving a symbolic structure, comparing it with the other symbolic structures in the local memory, storing it, transforming it via specific to the agent's type routines, and sending its modification. Each one of these steps is discrete. DUAL – agents manipulate symbols sequentially, one after another, with a frequency that reflects the relevance of the respective agents.

4.2.2.4 Relationships between Connectionist and Symbolic Processing

All aspects of the agents are merged in a single whole and each one influences the others.

Only a small number of agents whose activation exceeds a certain threshold form the *Working Memory* (WM). The agents that are outside of the WM cannot perform any symbolic operations – they are assumed invisible. For the agents that are involved in the WM, the activation level determines the *speed* of the symbolic processing. Each elementary symbolic operation (namely *read*, *send*, *modify*, etc.) has a 'price' that is paid with activation. Whenever an agent wants to perform such an operation, it begins to accumulate activation in order to pay the required price. Only after it is ready, can it perform the operation. Therefore, the most active agents work rapidly, the less active ones – slowly, and the inactive ones do not work at all. In this manner, the relevance influences the symbolic part of the architecture.

There is also an opposite dependence. The symbolic operations cause new agents to be born, and new connections to be established. This changes the overall pattern of activation. Thus, the symbolic operations influence the pattern of relevance too.

4.2.3 The Coalitions of Agents

DUAL – agents are very simple and some of the more important properties of the architecture could be observed if looked at from a distant perspective.

DUAL – agents form *coalitions*, i.e., sets of agents, together with the pattern of interactions among them. Coalitions represent more complex entities like propositions or situations. However, the coalitions are not part of the strict computational description of the architecture. Instead, they enhance the conceptual understanding only.

Three important properties of the architecture become visible only at the level of coalitions. The coalitions are *decentralized*, *emergent*, and *dynamic*. None of these properties is presented in the individual agents. The coalitions vary in the intensity of the interactions among their members in comparison to the intensity of interactions with the outside agents. 'Tight' and 'loose' coalitions could be distinguished in respect to this ratio and there is a whole continuum between these two extremes.

Coalitions do not have clear-cut boundaries. Instead, the same agent can participate in two or more coalitions and to a different extent. In the course of time, the coalitions can become 'tighter' or 'looser', can break up, and new coalitions can emerge.

4.4 The AMBR Model

4.3.1 Main Ideas

Since JUDGEMAP treats judgment as a process of mapping (or roughly speaking, a kind of analogy - making), it is based on the Associative Memory Based Reasoning (AMBR) (Kokinov, 1994a, 1998, Kokinov & Petrov, 2001) model. AMBR is a model for analogy making based on DUAL. It treats analogy making as an emergent result of the common work of several overlapping sub-processes – perception, retrieval, mapping, transfer, evaluation, and learning. However, AMBR is a long-term project and unfortunately, at the current stage only the processes of retrieval, mapping, and transfer are modeled and integrated.

AMBR is capable of capturing some similarities between local structures of agents and mapping them, creating *hypotheses for correspondences*. It is a pressure for these mappings to grow, involving other agents. In this way, a Constraint Satisfaction Network is formed. Just as in the process of crystallization, the system strives to a stable equilibrium, changing quantitatively itself. Because the structure–based mappings emerge locally and grow, often some inconsistent hypotheses meet and compete with each other and sometimes blending between episodes occurs.

4.3.2 Mechanisms Used in AMBR

Several AMBR mechanisms will be briefly reviewed in this section. These mechanisms were explicitly designed for modeling the process of analogy – making.

4.3.2.1 Spreading Activation

The sources of activation are two special nodes – INPUT and GOAL. The node INPUT represents the perception of the system, whereas the node GOAL – the current tasks of the system. AMBR's work begins when some agents that represent the environment, are attached to INPUT, and some other agents, which represent the task – to GOAL. The activation then spreads through the Long-Term Memory (LTM) network.

The spreading activation mechanism defines the working memory, determines the speed of the symbolic processes performed by each individual agent, and underlies the relaxation of the constraint satisfaction network.

4.3.2.2 Marker Passing

AMBR marks the instance-agents[1], which enter in the WM; they in turn mark their respective concepts; then the markers spread to their neighbors that are up in the class hierarchy, and so on.

The main purpose of the mechanism is to justify some semantic similarities between two agents. Whenever two markers cross somewhere, AMBR creates a hypothesis about a correspondence between the two marker origins. The justification for this hypothesis is the fact that these two origins have a common super-class, i.e., they are similar in something. Note, that AMBR makes such inferences only if the whole paths of the markers involve only relevant agents.

[1] AMBR differentiates *instance-agents* (responsible for individual items) from *concept-agents* (responsible for classes of homogeneous entities).

4.3.2.3 Structural Correspondences.

Like marker passing, this mechanism creates hypotheses between agents. The difference is that the former is sensitive to semantic justifications, whereas the latter – to propositional ones. There are different kinds of structural correspondences – if two relations are mapped, their arguments should also be mapped; if two instances are mapped, their respective concepts should be mapped.

Because several mechanisms create hypotheses independently, it is possible for some of them to be duplicated, or some of them to be contradictory. A special mechanism, namely secretary's work, establishes inhibitory or excitatory links between them.

4.3.2.4 The Constraint Satisfaction Network

The Constraint Satisfaction Network (CSN) consists of hypotheses for correspondences and is interconnected with the main one. Each hypothesis receives activation from its arguments and from its justifications. It is also inhibited from its competitors. Thus, CSN simultaneously reflects the semantic, pragmatic, and structural pressures of the analogy–making task. Due to the CSN, the global behavior of the system emerges from the local interactions between agents.

4.3.2.5 Rating and Promotion

Because at some moment the system should finish its work, each agent rates its competing hypotheses at regular time intervals. If one of them holds for a long time as a leader, it is promoted to a winner.

Note, that there is not any central executor that monitors the CSN and decides whether the network is relaxed enough. Instead, some hypotheses become winners locally and in asynchrony. This allows blending between episodes to occur, or unique solutions to be found (of course, sometimes useless solutions are also proposed by the system).

4.3.2.6 Skolemization and Transfer

These mechanisms augment the descriptions of the episodes based on semantic or structural information. Their goal is to ensure a tolerance to lack of information, or to reformulations of the task. However, there is no need of these mechanisms in a judgment task.

4.5 The JUDGEMAP Model

Like AMBR, JUDGEMAP is based on the DUAL cognitive architecture. Moreover, the two models are integrated and share the same mechanisms.

4.5.1 The Agent's Types

JUDGEMAP consists of nothing but DUAL agents of various kinds. The individual agents carry information about the entities that they represent. Each agent's type keeps an agenda of possible operations that the agents of this type can perform. The type of each agent is marked with a tag in the :type G-slot of its micro-frame.

There are four major types of agents in JUDGEMAP: concepts, instances, hypotheses, and justifications. The first three are inherited from AMBR; the fourth one is a novel one. All types are systematized in a taxonomic manner and are illustrated in fig. 4.4.

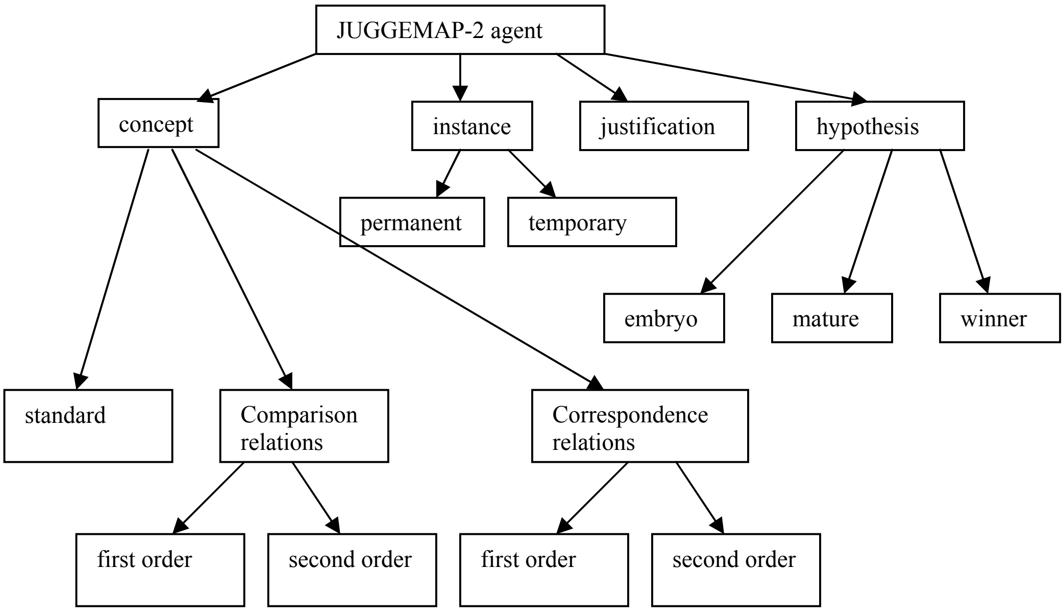

Fig.4.4. Main types of agents, used by JUDGEMAP

Concept–agents (for short *concepts*) represent classes of homogeneous entities and are organized in a semantic network. They are interconnected via vertical links, labeled :subc and :superc, for pointing respectively to their super-classes and some of their sub-classes. They are interconnected also horizontally, pointing to some associations and prototypical relations. Note that concepts can stand for objects as well as for relations and abstract terms.

JUDGEMAP proposes two special kinds of concepts:

Comparison-relations represent classes of specific relations that have two arguments and express a comparison between these arguments. Examples of comparison-relations are concepts like “longer_than”, “cheaper_than”, “better_than”, etc. They are equipped with the specific procedural knowledge that allows them to recognize manifestations of the relation for which they are responsible among relevant items.

Having comparisons between two objects, the model would theoretically be able to order the stimuli on an ordinal scale, i.e. it might be able to represent the fact that stimulus A is bigger than B, and B is bigger than C. However, in order to be able to map these stimuli on an interval scale it should also be able to differentiate between and compare two comparison relations, i.e. to be able to say that A is much bigger than B, than B is to C. People make such statements quite often. For example, most Bulgarian citizens would not only recognize that Everest is higher than the Rila mountain, and that the Rila mountain is higher than Vitosha, but they would definitely also recognize that the difference between Everest and Rila is (in fact much) bigger that between Rila and Vitosha. This would require a second-order relation between comparison relations. (And if we also need to be able to say that one difference is *much* bigger than another one, then we would also need third-order relations. However, we do not really need the third-order relations in order to be able to form interval scales and that is why we did not introduce such relations – in addition, they seem to be too complicated and it would not be justified to introduce them without checking whether people really use them.) Thus, JUDGEMAP involves first and second-order comparison relations.

Correspondence–relations represent the specific tasks of judgment. For example, if the task is to judge lengths of lines, one correspondence relation represents the knowledge "longer lines have to correspond to higher ratings". These correspondence-relations can be temporary agents that do not participate in Long-Term Memory. The correspondence relations trigger the mechanisms for the construction of hypotheses about correspondences.

Instance-agents (*instances*) represent individual entities. Each instance has a G-slot, labeled :inst-of that points to its respective category concept. Links, labeled :instance connect the concepts to some of their instances. Some instances are permanent – they are part of LTM and represent concrete memory traces. Other instances are temporary – they are constructed on the spot because of certain inferences (usually as a result of the work of comparison-relations). Some magnitudes can be represented with numbers that are filled in specific slots of the instances.

Hypothesis agents represent possible correspondences between entities (e.g., between stimuli and scale labels). They are temporary agents; they do not participate in LTM; and if they lose their relevance, they disappear. Hypotheses are organized in a constraint satisfaction network (consisting of excitatory and inhibitory links). From the relaxation of this network, the responses of the system emerge. Each hypothesis has its 'life cycle' – it can transform itself from *embryo* to *mature hypothesis*, and then to *winner*. These sub-types reflect the degree to which the hypotheses are novel and attractive.

Creation of a hypothesis requires a reason for this to be found. The reasons are inferred from the common effort of several agents and are expressed by justification agents (for short – *justification*). The justifications are also temporary. Their purpose is to combine many data into a single inference and then to create a hypothesis, which represents this inference.

4.5.2 JUDGEMAP Processing

One judgment session usually involves several overlapping processes.

The instance that represents the target stimulus to be judged is attached to the INPUT and GOAL nodes, assuming that on the one hand, people see the stimulus, and on the other, it is part of the task. The concept that represents the scale, together with some contextual elements (if any) is also attached to INPUT. The respective correspondence relation (e.g. bigger corresponds to higher rating) is connected to the GOAL node.

Due to the spreading activation mechanism, similar exemplars from LTM, relevant concepts, possibly together with their prototypes (if any), and various properties of the target stimulus can enter WM. Some recently judged stimuli might also stay active even if they are not similar to the target. The activation spreads through more abstract concepts and then again to their instances.

Other mechanisms do not wait until the activation stabilizes; they run in parallel. Each instance–agent that enters the WM emits markers. The concepts transmit them to their super-classes and via special mechanisms for exchanging *argument-related messages*, to the relations that may interest them. Some concepts send *argument-related requests* to other concepts, asking them for their markers. The receivers of such requests answer by sending back *argument-related answers* with the requested information. (This mechanism is to be discussed in more detail in the next chapter).

The comparison-relations collect and update the information that is relevant to their goals. They can compare two magnitudes or two local structures and if they satisfy the relation they dynamically create their own temporary instance.

Soon after the first such instances establish and send their markers, the correspondence relations begin their work. They check whether some established comparisons, together with some existing hypotheses could be viewed as reasons for other hypotheses. For example, the knowledge that "line-A is longer than line-B", together with the already established correspondence "line-B<-->rating-5" makes the hypotheses that "line-A could be judged with rating-6 or rating-7" reasonable.

The correspondence-relations build justification-agents on the basis of such inferences. The purpose of the justification-agents is to combine the effort of all agents that together explicit the respective inference. After their establishment, these justification-agents in their turn create new hypotheses for correspondences.

For various reasons and at various moments of time many new hypotheses emerge. Some of them duplicate each other and in this case, they combine their justifications; the contradictory ones create inhibitory links between themselves. In this way, a constraint satisfaction network is formed. When a certain hypothesis about the target wins over the competitors, it is considered as the response of the model. Then the system receives the next target stimulus without any initialization and continues with its judgment.

4.5.3 Back to the Phenomena

As mentioned above, the starting point for modeling JUDGEMAP were DUAL principles, which were assumed to be at the core of human cognition. Some of the mechanisms used by the model may seem too complicated, others unnecessary, etc. They allow, however, integrating JUDGEMAP with other models within the framework of a single architecture. From this perspective, judgment seems not to be a separate module but a natural part of human cognition.

However, the model would not contribute much, if it fails when verifying it against the experimental data. In this subsection, possible explanations of some of the observed effects are discussed in terms of JUDGEMAP mechanisms. The real verification, of course, could arise only from the simulation results presented in chapter 6.

In short, we assume that one of the main sources of assimilation effects is the association–based spreading activation, whereas the contrast effects arise from the mechanisms for creating comparisons and also from the competition between competing hypotheses.

We suppose that sequential assimilation occurs because of the residual activation of the previous *rating*. The overall pattern of activation over all scale values changes dynamically. There are always some ratings which are more active and some others which are less active. The hypotheses for correspondences, however, receive positive activation from two sources – from their justifications, and from their elements. The role of the justifications in the competition between hypotheses is intuitively clear. However, the role of the relevance of the *elements* of the hypotheses does not seem essential for the judgment task. It is a consequence of the basic mechanisms that underlie analogy–making (in particular - mappings).

JUDGEMAP points to an important differentiation: sequential assimilation is not towards the previous stimulus, but towards its rating – just as it was observed in the experimental data (Petrov & Anderson, 2000; Strack & Mussweiler, 1997).

The main sources of the contrast effects in JUDGEMAP are two. On the one hand, the comparison between the stimuli highlights their differences and hence creates a pressure for their ratings to be differentiated too. On the other hand, the soft version of the pressure for one-to-one mappings[1] causes a tendency for the scale labels to be used an almost equal number of times (in accordance with the frequency principle). This pressure is inherited from the AMBR model. The importance of one-to-one mapping is obvious in analogy making, but not in a judgment task. Thus, the assumption that the same mechanism produces the mapping in analogy and judgment explains the emergence of the frequency principle in judgment.

4.5.4 Comparison with Other Theories and Models

This section discusses the main ideas that JUDGEMAP shares with other theories.

For example, the creation of comparisons could be considered as a basic mechanism for the JUDGEMAP model. Comparisons between the target stimulus and the memory traces are also in the core of the Comparison-Based Judgments (Mussweiler, 2003) and the Two-Paths Model (Manis, Paskewitz, 1984). Like the latter JUDGEMAP assumes that comparisons always cause contrast, in contrast with the assumptions of the former model. The expectation path, proposed by the Two-paths model is analogical in many respects to the influence of the residual activation on DUAL's behavior.

However, JUDGEMAP differs from both models in several respects. First, being a computational model, it proposes more concrete mechanisms. Second, it creates comparisons dynamically, one after another, and in parallel with other working mechanisms. Third, several sub-mechanisms in JUDGEMAP reflect the relevance of the items: the speed of creation of comparisons; and the strength with which these comparisons would influence the other agents.

JUDGEMAP could be considered as being similar to the Norm Theory (Kahneman & Miller, 1986) in respect to the associative memory–based principles, responsible for the construction of the comparison set. One point in which they differ is that the Norm Theory assumes a construction of an explicit norm, whereas in JUDGEMAP such a norm could be viewed as an implicit emergent result. JUDGEMAP is also endowed with other additional mechanisms, which contribute to the judgment process, and is implemented in a computer program.

Several theories are complementary to JUDGEMAP, i.e. are focused on some complementary aspects of the problem. Perceptual Learning Theory (Goldstone, 1998) and Integration Theory (Anderson, 1971) fall in this group. One of the future

[1] Hypotheses that connect one and the same rating with different stimuli compete with each other

goals of the DUAL research group is to augment JUDGEMAP with mechanisms for perception and categorization.

JUDGEMAP shares with the EBWR model (Nosofsky, 1997) the idea that judgment is a result of a competition between alternative hypotheses, and that comparisons between the target and some memory traces justify these hypotheses.

However, there are several differences between the two models. EBWR consists of mathematical formulas and involves stochastic elements, whereas JUDGEMAP is a multi-agent deterministic model that is grounded on a larger cognitive architecture (and hence, it is integrated with other cognitive models). EBWR assumes strictly exemplar–based memory, whereas DUAL and the models based on it allow the combination of both exemplar–based and prototype–based approaches.

Both JUDGEMAP and the ANCHOR model (Petrov, 2005) are computational models, based on large cognitive architectures (ANCHOR is based on ACT-R). ANCHOR assumes a two-step process – first, simultaneous comparisons between the target stimulus and all prototypes of scale labels, followed by an explicit correction strategy. The ANCHOR model has an advantage with respect to its learning mechanism that accounts for the dynamical changes of the representation of the categories. Its behavior, however, (unlike JUDGEMAP's) is restricted by special predefined parameters that depend on the concrete type of stimuli and scale.

Since JUDGEMAP integrates analogy-making with judgment, and proposes concrete computational mechanisms, it makes a step ahead of many descriptive theories of judgment.

CHAPTER V
Detailed Description of the Model

5.1 Knowledge Encoding

All the knowledge of JUDGEMAP is represented by a network of a huge number of DUAL-agents. Each agent has a name. It is practical to use mnemonic names, like *line*, *color-of* and *longer*. However, these names are irrelevant to the program itself; it would work in the same way had the agents been named ag0001, ag0002, etc.

DUAL-agents also have specific types, which determine the set of operations that a given agent can perform. Again, words like *concept*, *instance*, and *hypothesis* are used for naming types, with full awareness that these names serve only the documentation of the model, not its actual work.

5.1.1 Representation of Concepts and Instances

The concept-agents represent classes of objects, whereas the instance-agents represent individual entities. This differentiation is popular in AI domain; it is known also as type vs. token, and class vs. instance distinction.

The concepts are arranged in a taxonomic semantic network via :subc and :superc links. The instance-agents are related to the concepts via links :inst-of and :instance (fig 5.1.).

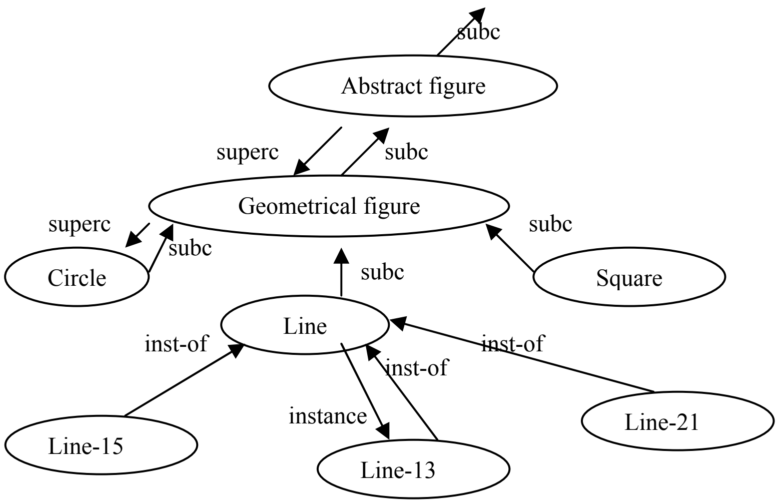

Fig. 5.1. The concepts and instances are represented by a hierarchical semantic network. The concepts are interconnected via :subc and :superc links, whereas the instances are connected to the concepts via :inst-of and :instance links. From each concept only few top-down links go out.

Each concept can be connected to zero, one, or more of its super-classes or sub-classes. It is an assumption, however, that the top-down links :superc and :instance represent only the most *salient* subclasses and instances, hence they cannot be too many (usually 1-4). The exact definition of the term *salient* is still under discussion within the DUAL research team. In JUDGEMAP two types of agents were defined to be more salient – the prototypes and the recently used instances. The top-down links from a concept to its prototypes are predefined and stable. The links to the recently used ones are temporal (named :t-instance); they are created at the moment when the concept receives a marker from them; and their weight slowly decreases in time.

Some instances are temporal and do not belong to long-term memory. They are constructed on the spot as a response to certain inferences, and if they fall out of WM, they disappear forever. In JUDGEMAP, such temporary instances represent the knowledge about comparisons between objects (for example*, line18 is longer than line10*). This knowledge is relevant in the context of judgment of lengths of lines, but after losing its relevance, the system forgets it. (Of course, the same knowledge can be constructed again, if needed.)

5.1.2 Representation of Relations and Propositions

Some multi-agent memory models separate objects from relations. JUDGEMAP does not follow this approach. Relations are represented in the same way as their arguments – via a network of concepts and instances. A piece of propositional knowledge (*line2-15 is longer than line1-10; and the latter is green*) is presented in fig. 5.2.

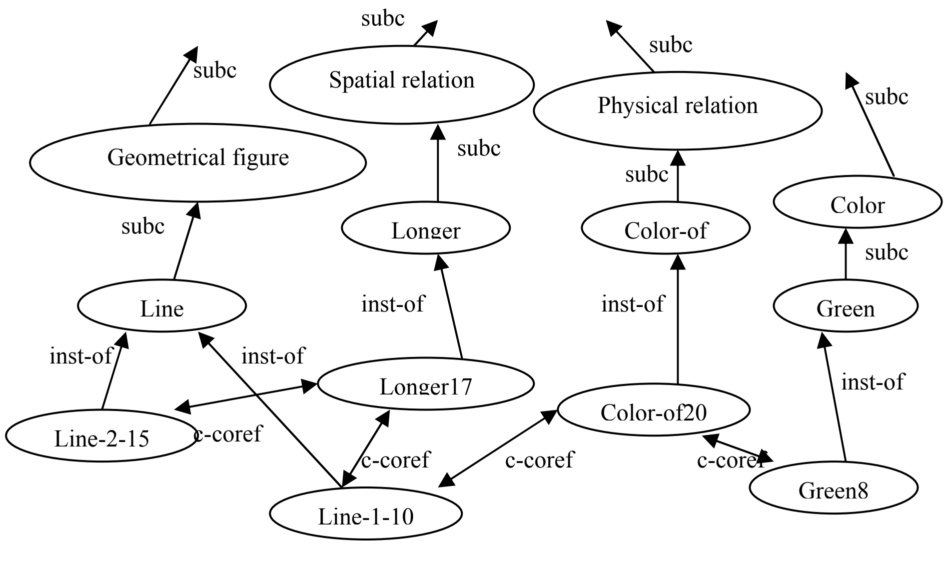

Fig. 5.2. The propositional knowledge is represented by coalitions of interconnected instances and concepts. (Fig. 5.3. demonstrates in more detail some of the agents and links.)

It is not difficult to distinguish the proposition *line2-15 is longer than line 1-10* from the inverse one – *line-1-10 is longer than line2-15*, because of the elaborated DUAL representational scheme (Kokinov, 1988, 1994a). DUAL-agents have G-slots, which represent some general knowledge about the agent, but also have S-slots, in which different aspects of the knowledge are organized. The more detailed description of the propositions *line2-15 is longer than line1-10* and *line-2-15 is green* is presented as a transcript in fig. 5.3.

```
Longer17
  :type (:instance temporary)
  :inst-of longer
  :slot1
    :inst-of (longer . :slot1)
    :c-coref line2-15
  :slot2
    :inst-of (longer . :slot2)
    :c-coref line1-10
```

```
Longer
  :type :concept
  :subc spatial_relation
  :slot1
    :subc (spatial_relation . :slot1)
    :c-coref object
  :slot2
    :subc (spatial_relation . :slot2)
    :c-coref object
```

```
Color-of-20
  :type :instance
  :inst-of color-of
  :slot1
    :inst-of (color-of . :slot1)
    :c-coref line1-10
  :slot2
    :inst-of (color-of . :slot2)
    :c-coref green8
```

```
Color-of
  :type :concept
  :subc physical_relation
  :slot1
    :subc (physical_relation . :slot1)
    :c-coref object
  :slot2
    :subc (physical_relation . :slot2)
    :c-coref color
```

```
Green8
  :type :instance
  :inst-of green
  :c-coref (color-of-20 . :slot2)
```

```
Line1-10
  :type :instance
  :inst-of line
  :c-coref ((longer17 . :slot2)
            (color-of-20 . :slot1))
```

```
Line2-15
  :type :instance
  :inst-of line
  :c-coref (longer17 . :slot1)
```

Fig.5.3. Coalition of seven agents, which represents two propositions – *line2-15 is longer than line1-10* and *line1-10 is green*. The coalition does not have any clear boundaries – it is interconnected with the main network. Fig. 5.2. presents the same knowledge in nodes – links diagram.

The conceptual co-references (c-coref) connect two complementary aspects of the same entity. In the example below, *line1-10* point to the same thing in the outside world as the second argument (:slot2) of *longer17*. That is why they are connected with the :c-coref link.

The S-slot labels, like :slot1, :slot2, etc. are absolutely arbitrary and their order does not have any prescribed meaning[1]. The arguments of a given relation are defined only by :inst-of or :subc facets, which are within each specific slot. It is possible for some aspects of an agent to be highlighted in a certain context, or some other aspects to disappear from WM. It is possible for an agent to have temporal slots – they can be constructed or removed on the spot. Hence, it is practical to keep the meaning of the arguments of the relations independent of any reordering of the S-slot labels. Only the pointers to their parents define their meaning. Such organization provides greater flexibility. It is possible for two different slots in a child-agent to inherit from one and the same slot in the parent-agent, or some parent's slots to be left unused in the child-agent, etc.

5.1.3 Representation of Perceived Properties

DUAL does not yet deal satisfactorily with perception. The only steps in this direction were made by constructing the PEAN model (Nestor, 2004), but unfortunately, this model is not fully integrated with the architecture. Hence, it is not

[1] The only exceptions are the most abstract relations. The order of their S-slots (if any) *does* have a prescribed meaning.

yet defined how JUDGEMAP could construct new perceived instances on its own and how it can interconnect them with the others. Instead, ready-made instances are constructed manually and presented to the model.

In order to judge a certain stimulus, several conditions appear necessary. First, one needs to recognize and classify the stimulus, i.e. to fill its :inst-of slot. Second, one needs to recognize its relevant aspects, i.e. its properties. This means a coalition of DUAL-agents to be created and interconnected via :c-coref links. Third, one needs to represent somehow the quantitative knowledge about these aspects. This is done by filling in a special slot, named :amount, with a real number (e.g. the magnitude of the stimulus). Leaving some questions for future discussions (for example, is it necessary to recognize the stimulus in order to judge its properties), the work of JUDGEMAP begins after manually creating coalitions like the one shown in fig. 5.4.

Line15	**Length-18**
:type :instance	:type :instance
:inst-of line	:inst-of length
:c-coref length-18	:amount 125
	:c-coref line15

Fig. 5.4. Representation of the information that *line15* has *length 125*.

The number 125 in the slot :amount of *length-18* does not mean meters, inches, or something like that. It is just an internal representation of the physical magnitude. For simple stimuli, like lines' lengths, the exact function that transforms the physical magnitude into its internal representation can be experimentally obtained using Weber's low as a template and adjusting the coefficients in it.

JUDGEMAP, however, uses numbers as fillers of the :amount slots also for much more ambiguous and uncertain properties. In the context of judging food, a coalition, representing the aspects of a certain cheese, is shown on fig. 5.5. Here, it is

Price-2	**Cheese-15**	**Quality-12**
:type :instance	:type :instance	:type :instance
:inst-of price	:inst-of cheese	:inst-of quality
:c-coref cheese-15	:c-coref (price-2 quality-12)	:c-coref cheese-15
:amount 18		:amount 150

Fig.5.5 Representation of a certain instance of cheese, together with its price and quality (see text for details).

not clear what the concrete physical magnitudes, responsible for the quality (taste) of the cheese are but the model assumes that the internal psychological representation of the taste may be *coded* with a number for the model's purposes[1].

Representation of internal magnitudes with real numbers is used in other models for judgment as well, but still looks too artificial. Note, however, that JUDGEMAP does not use these numbers in any complex calculations. The only purpose of the :amount slot is to be used in construction of comparisons between entities (the mechanism is described in detail further in the text). For example, if the system 'knows' that *line10* is long 10, *line12* – 12, *line20* – 20, then it is able to make

[1] However, JUDGEMAP does not deal yet with the judgment of very abstract concepts like the beauty of an idea.

inferences that *line20 is longer than line12*, *line12 is longer than line10*, and that the difference between lengths of *line20* and *line12* is bigger than the difference between lengths of *line12* and *line10*. (fig . 5.6.)

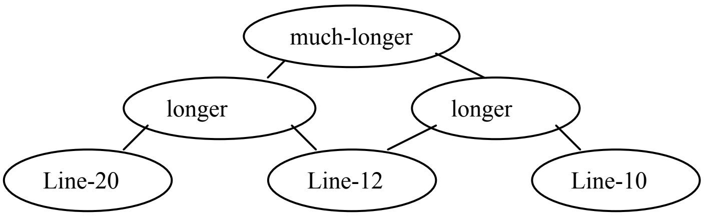

Fig. 5.6. The system is able to make comparisons between instances and, inductively, comparisons between comparisons.

It is possible, however, that human perceptual mechanisms may build these comparisons more directly, without using any internal magnitudes. Nevertheless, since JUDGEMAP focuses on the mechanisms of the judgment process, it was assumed that the system is able to make the discussed comparisons between items.

5.1.4 Representation of Scales

Like each other piece of knowledge, scales are represented in DUAL as coalitions of micro-agents (fig. 5.7.).

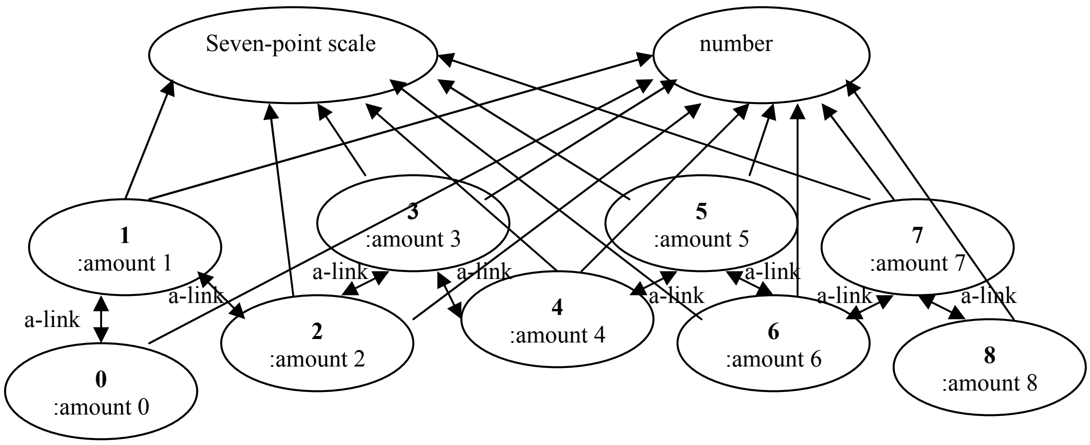

Fig. 5.7. Representation of the scale from 1 to 7. Scale labels are just numbers and can participate in other scales or relations. Neighboring ratings are connected with associative links.

Each individual rating is represented by a single instance-agent, which has :amount slot, filled with the concrete number. It also has :inst-of link to the concept number. The ratings, participating in a certain scale (and only they), also have in their :inst-of slot links to the 'head' of the scale (which is a concept-agent). Of course, one number can participate in many different scales, and it is not possible to fill its :inst-of slot with too many concepts. Only the prototypical scales have to be defined. Because this question is not essential for JUDGEMAP purposes, it is left open, waiting to be solved together with the more general problems of recognition and generalization in the DUAL architecture. For the simulations, JUDGEMAP does not use more than one scale simultaneously, and does not use too many scales at all (actually in the simulations a 2-point, a 3-point, a 7-point, and 100-point scales are used).

As shown in fig. 5.7. :a-links connect the neighboring rating labels. The *associative links* (a-links) are used in DUAL-based models to represent associations between entities that cannot be represented with other types of links (the classical example is cow - milk). A-links do not have any semantic meaning; they serve only the connectionist aspect of the system. Thus, the order of the scale is not explicitly represented. However, it can be inferred by comparisons between the fillers of the :amount slots.

It is not predetermined that a-links necessary connect neighboring ratings and only them. For example, in a 100-point scale a-links connect also the rating labels 10 with 20, 20 with 30, etc. Moreover, the a-links can be asymmetrical. They are usually weighted higher in the direction from smaller to larger rating, than vice versa (it is more difficult to count from 10 to 1 than from 1 to10). A special tool is implemented in DUAL that generates randomly variations of the knowledge base, varying the weights of the :instance and :a-links. In this way various "individuals" are simulated who perform the same judgment task.

5.1.5 Representation of the Task for Judgment

The task "Please, judge the property **P** of the stimulus **S** on scale **A**" consists of three explicit elements – **P**, **S**, and **A**, and some implicit ones. For example, in judgment of lengths of lines, it is assumed that longer lines have to receive higher ratings. Moreover, it is implicitly assumed that to almost equal differences between two lengths have to correspond almost equal differences between their ratings.

Knowledge of this sort is explicitly represented in JUDGEMAP by a special kind of concepts, called *correspondence–relations*. An example of such an agent is shown in fig. 5.8.

```
Longer <-> higher rating
  :type :concept
  :subc correspondence–relation-rang1
  :a-link much-longer<-> much-higher-rating
  :slot1
    :subc (correspondence–relation-rang1 . :slot1)
    :argument longer
  :slot2
    :subc (correspondence–relation-rang1 . :slot2)
    :argument seven-point-scale
```

Fig. 5.8. An example of correspondence–relation of rank1. It represents the knowledge that longer lines have to receive higher ratings.

Correspondence–relations are usually temporal agents. They exist only during the period the respective task for judgment is relevant. The facets of type :argument in their S-slots serve both the connectionist and the symbolic aspects of the model.

Symbolic messages and their responses can be exchanged through those links. Using such messages, the correspondence–relations become 'aware' about number of relevant properties and are able to make inferences about them (detailed description of the process is presented in 5.3.4).

As shown in fig. 5.8. the implicit knowledge about judgment is represented by an a-link to another correspondence–relation –much-longer<-> much-higher-rating (fig. 5.9).

```
Much-longer <-> Much-higher rating
  :type :concept
  :subc correspondence–relation-rang2
  :slot1
    :subc (correspondence–relation-rang2 . :slot1)
    :argument more-longer
  :slot2
    :subc (correspondence–relation-rang2 . :slot2)
    :argument seven-point-scale
```

Fig. 5.9. An example of correspondence–relation of rank2. It represents the knowledge that if the difference of two lengths is higher than the difference of two other, this has to reflect analogically onto their ratings too.

It represents the more complicated information that in judgment one has to choose ratings not only according to the relative position of the stimulus, but also according to its absolute range. One example is shown in fig. 5.10.

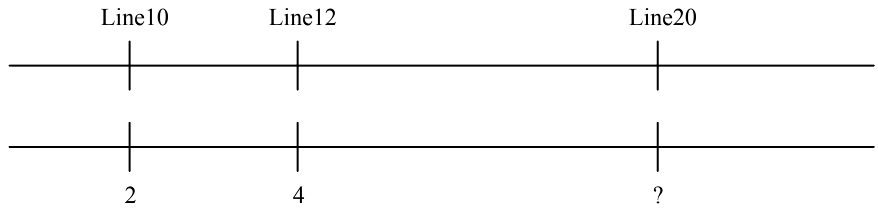

Fig. 5.10. The purpose of correspondence–relations of higher order is to ensure pressure to judge also according to the absolute position of the stimulus.

Line20 is longer than both line10 and line12, and obviously, it is a good idea to judge it with a rating higher than 4. However, because the middle line (line12) is closer to line10 than to line20, there is a pressure also to judge line20 with even greater ratings – 6 or 7.

Inductively, correspondence–relations of rang 3, 4, etc. can be constructed. However, this requires much more resources – working memory, calculations, etc. The correspondence–relations of higher order receive activation only from those of lower order. Hence the higher their order is, the less active they are, until they become inactive at all. Without putting any strong limitations, JUDGEMAP uses in the simulations only first and second order correspondence–relations.

5.2 Spreading Activation

The activation level in DUAL-based models represents the relevance of the respective agent to the context. The spreading activation mechanism defines the working memory of the system as the set of all agents, whose relevance exceeds a certain threshold. Activation also determines the speed of each symbolic processor.

5.2.1 Activation Functions

From a connectionist point of view, the DUAL-based models work like huge connectionist networks. Each agent receives activation from many other agents and sums it via the input function **n=n (t)**, where **t** is the time. All links are excitatory (with the exception of hypothesis – agents, described in section 5.4.6), and this guarantees that **n (t)>=0**.

From a conceptual point of view, it is assumed that time is continuous. The computational realization, however, uses quanta of time, the length of which is negligible small in comparison to the macroscopic time scale.

The connectionist processor of each agent calculates the new activation level $\mathbf{a_{new}}$ at each elementary cycle on the basis of the previous level $\mathbf{a_{old}}$ and the net input **n**, using the same activation function, like the one used by the AMBR model (Petrov, Kokinov).

$$
\begin{cases}
\mathbf{a_{new} = 0}, & \text{if } \mathbf{a_{old} + E.n.(M-a_{old}) - d.a_{old} < \Theta} \\
\mathbf{a_{new} = M}, & \text{if } \mathbf{a_{old} + E.n.(M-a_{old}) - d.a_{old} > M} \\
\mathbf{a_{new} = a_{old} + E.n.(M-a_{old}) - d.a_{old}}, & \text{otherwise,}
\end{cases}
$$

where **M** = const is the maximal activation value, **E** and **d** are parameters that control the rate of excitation and decay respectively, $\mathbf{\Theta}$ is a threshold. As was mentioned above, $\mathbf{a_{new}}$ is the new activation level, $\mathbf{a_{old}}$ – the old one, **n** is the net input.

This function has the following properties:

1) It is a non-decreasing function with respect to the input **n**.
2) The activation level of each agent is bound between **0** and **M**.
3) All activation levels, which are smaller than the threshold $\mathbf{\Theta}$ clip to zero.
4) It is a spontaneous decay and each agent loses its activation in the absence of external support.

At each connectionist cycle each agent also sends activation, the amount of which is calculated through the output function **o (t) = a (t)**, i.e. the total output is equal to the activation level. This is done by normalizing the weights of all outgoing links so that their sum is always equal to 1. The normalization of the weights is one of the mechanisms, responsible for preventing working memory from unlimited growth.

5.2.2 INPUT and GOAL Nodes

The INPUT and GOAL nodes are the sources of activation. Their activation level is always equal to the maximal value **M**. By attaching different agents to them, respectively perceptions and tasks are simulated. Each simulation of judgment of a set of stimuli looks like the following: The respective correspondence–relation, together with the concept for the scale is attached to a GOAL node (for example – the relation *longer<->higher-rating* and the concept *seven-point-scale*). Then the stimuli are attached to both GOAL and INPUT sequentially, one after another, and are removed from there immediately after a response is given. Various context or priming objects can be attached or removed from INPUT at any moment in time. It is possible for several stimuli to be given simultaneously for judgment; it is also possible for predefined correspondences between stimuli and ratings (anchors) to be constructed and attached to INPUT or GOAL.

Each agent, attached to GOAL, receives a tag :t-driver in it's :type slot., which stays there while the agent stays on GOAL. The agents, connected to :t-driver agent, also receive the same tag (to remember, c-coref links connect two aspects of one and the same entity). Various mechanisms need the information about whether a certain agent is :t-driver or not.

5.3 Marker-Passing and Exchanging Symbols

Generally speaking, the marker-passing mechanism serves to find out whether a path between two agents is present or not. All symbolic interactions between agents, i.e. exchanging messages, are local and involve only two neighbors. The marker passing mechanism is not an exception, but it allows information about agents to be carried through longer distance as an emergent result.

5.3.1 Marker Emission and Marker Passing

Each instance-agent emits a marker when entering the WM – it sends it to its parent concepts through the :inst-of links. The marker itself is a symbolic structure that contains the name of the origin and information about whether the origin is :t-driver or not.

When it receives a marker, each concept agent analyzes it and performs with it the operations, which are specified in its symbolic routine (if any) – for example, answers to an argument-related request (this mechanism is described in the next section). After that, the concept stores a copy of the marker in its buffer (temporary local memory), and sends it to its neighbors, which are up in the class hierarchy through :subc links. In addition, if the origin is a direct instance of the concept, the latter creates a temporary top-down link of type :t-instance to the former. The weight of this link slowly decreases in time.

To summarize, the presence of a marker in a certain concept codes the information that the concept 'knows':

1) The marker origin is in WM;
2) The marker origin is an instance of the respective concept – direct or by inheritance;
3) Whether marker origin is :t-driver or not (i.e. is it attached to GOAL node or not).

Note that the number of markers, stored in a particular concept, stay relatively small, because of the natural limitations of their spreading, emerging from the basic principles of DUAL architecture. Marker emission and marker passing are symbolic operations. Thus, only the active agents can emit, receive or transmit markers. Moreover, these operations require time, and this time depends on the activation level of the agents. Consequently, spreading markers is limited by the boundaries of WM. The speed of this spreading is extremely slow in its periphery. Because all processes in JUDGEMAP run in parallel, it often happens that the decision has been made and a response has been given before the very slow markers reach the respective concepts. Thus, only the most relevant agents participate in the decision-making process.

In addition, only instance-agents emit markers, whereas the concepts only pass the existing ones. When an instance – agent leaves WM, all its markers are removed. As a final result, JUDGEMAP simulations show that even the most active concepts store no more than 5-7 markers at each particular moment of time. The concepts analyze markers in a temporal order, reflecting their potential usefulness for the particular task and in the particular environment.

5.3.2 Argument–Related Messages

Sometimes concepts may need for their work to know the instances of another concept. This 'communication' is possible due to the links of type :argument and the messages that are exchanged through them. An example of a concept with :argument links is shown on fig. 5. 11.

```
Longer-line
  :type :concept
  :subc longer
  :argument (line more-length)
  :slot1
    :subc (longer . :slot1)
    :c-coref line
  :slot2
    :subc (longer . :slot2)
    :c-coref line
```

Fig. 5.11. An example of comparison-relation and the use of :argument links (see text for details).

Links :argument between concepts point to the respective class of instances of interest. When a concept–agent enters WM, it sends an *argument-related-request* through it's :argument links. This message means simply: "Send me the markers that you receive". When a concept receives such a request, it memorizes it in its local memory, and each time it receives a marker, it sends an *argument-related-answer*. These answers are just like the markers, but they do not spread further. Only the symbolic processor of its receiver uses them.

5.4 Comparison-Relations

The role of the comparison-relations is to *recognize* relations, for which they are responsible. For example, the relation *longer-line* (see fig 5.11.) receives information about the active lines via exchanging of markers and argument-related messages. If reasonable, it creates instance-agents of itself and interconnects them with the respective neighbors (fig. 5.12.).

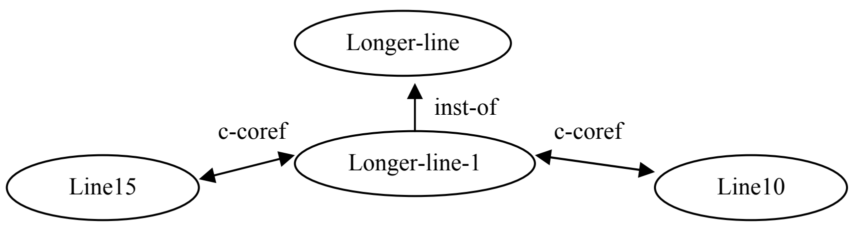

Fig. 5.12. The role of the comparison relations (longer-line) is to recognize the relations, for which they are responsible, and to create instances of themselves (longer-line-1).

There are two types of comparison-relations, depending on whether their arguments have :amount slot or not. For example, the arguments of *more-length* are lengths, which have :amount slot, whereas the arguments of *longer-line* are lines, which do not.

5.4.1 Comparisons between Amounts

The comparison-relations of the first type simply compare the two amounts and depending on this comparison decide how to order the pointers in the S-slots of the new instance, e.g. define which amount is higher and which one - smaller. If, for example, the agent *length-of-line-12* has :amount slot with filler 200, the agent *length-of-line-19* has :amount slot with filler 400, then the comparison-relation *more-length*, after receiving the necessary messages, would create a new comparison-instance-agent, which codes *that length-of-line-19 is more than length-of-line-12*. The new instance also has :amount slot, which is filled with the difference of the two amounts, e.g. 200.

Note that even simple objects like lines may have many different aspects and the information about them is distributed over many different agents. Hence, *line-19* is something different from *length-of-line-19*, *color-of-line-19*, *position-of-line-19*, etc. In the same way, the comparison-relation *more-length* is something different from the comparison-relation *longer-line*, no matter that they represent the same information.

5.4.2 Comparisons on the Base of Structures

The comparison-relations of the second type work in a slightly more complicated manner. *Longer-line* (see fig. 5.12,) has :argument links both to the concepts *line* and *more-length* (the same *more-length* that was discussed above). Its behavior does not depend on any amounts, but only on the structural relationships. In short, if it receives argument-related-answers about: *line19*, *line12* and *length-of-line-19 is more than length-of-line-12*, it starts the construction of the relation *line-19 is longer than line-12*.

Note that because of the specific way of symbolic computations in DUAL, there is no risk of “combinatorial explosion”. The new relations would not capture all pairs of instances. They emerge slowly, one after another, in an order that reflects their relevance. Just after the creation of the most relevant ones, the other processes in JUDGEMAP begin to work in parallel, and often the system gives a response long before the less active instances are compared. In addition, as many relations around a certain agent emerge, the so-called “fan-out” effect[1] decreases the activation levels and creates a pressure for the next relations to be created with more difficulty.

5.4.3 Comparisons of Higher-Order Relations

All new relations are instance-agents in their nature, and hence have all the rights as the other instances, i.e. they emit markers, they can participate in argument-related messages, etc. Therefore, they can also be arguments of another comparison–instance. In this way, some comparison-relations may compare comparisons. The relations that compare comparisons are called second-order comparisons (fig 5.13).

[1] ‘fan-out’ effect – because of the normalization of the weights of the links, the more output neighbors an agent has, the less activation it would send to them.

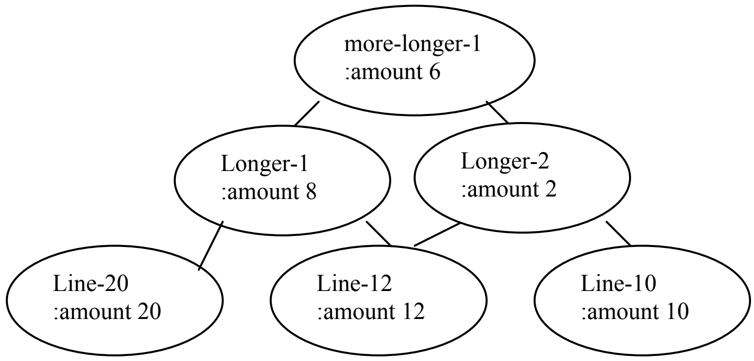

Fig. 5.13. Some comparison relations are responsible for comparisons of higher order. The information that the difference between line-20 and line-12 is higher than the difference between line-12 and line-10 is represented.

These comparison-relations (having comparisons for arguments) do not differ from the other comparison-relations in any way. Their activation level, however, is lower, because they receive activation only from their arguments. The conceptual idea is that such agents extract from the environment and encode more complicated information, involving more objects. This has a price, of course. They have less activation and work more slowly.

Inductively, it is possible to define comparisons of third, fourth, and higher order. JUDGEMAP, however, uses only those from first and second order[1], assuming that the more complicated ones would work so slowly that their work would not influence the processes.

The comparison-relations from higher order are an important innovation. Their purpose is to equip the model with a capability to use the properties of the interval scales, not only of the ordinal ones.

5.4.4 Fault Comparison-Relations

It is possible for two instances to be equal in their amount. In this case the respective comparison-relation that compares them, again creates a new relation between them, but marks it as a *fault comparison instance*. This mark is used by the mechanisms for creating hypotheses.

5.5 Correspondence–Relations and Justification Agents

The *correspondence–relations* represent the information that a greater amount of the judged property has to correspond to higher rating (the information given in the instruction for judgment). These relations typically stay attached to GOAL node. The role of the correspondence–relations is to trigger the creation of new hypotheses between stimuli and ratings.

[1] By convention, the name of the relations of second order replicates the names of those of first order, adding *more* at beginning, for example – *longer* and *more-longer*.

One of the assumptions behind all DUAL–based models is that each inference must have its justifications. The justifications themselves are usually semantic or structural, and hence are represented by DUAL agents. JUDGEMAP uses a separate class of justification agents, whose role is to combine different justifications in a single 'head' agent.

The correspondence–relations are very similar to the comparison-relations, according to their micro-frame. Usually they have :argument links and via exchange of argument-related messages, they collect information about the available comparisons and grade labels (fig. 5. 14).

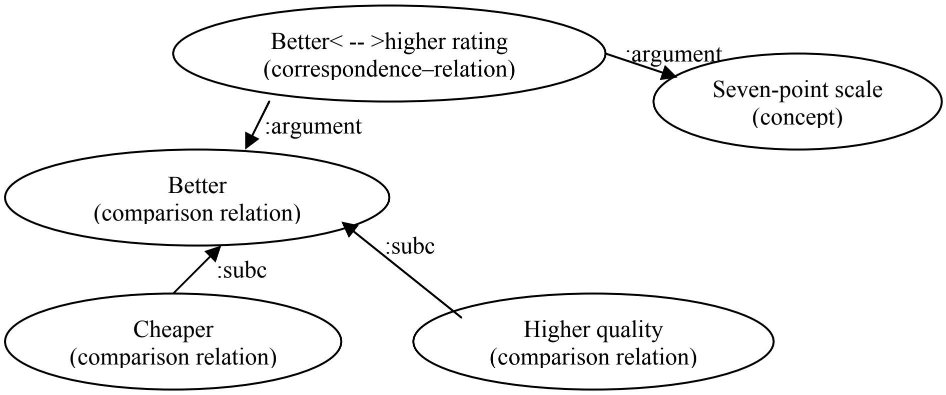

Fig. 5. 14. An example of a correspondence–relation (better < -- > higher rating), and its :argument links to comparison relations and scale concepts.

The correspondence-relations, however, differ from the comparison ones in respect to the operations that they perform, when they receive argument-related answers. When a new comparison arrives, the correspondence–relations extract the hypotheses that one of the arguments of the comparison also have, and trigger creation of new hypotheses, which involve the other argument of the comparison and the appropriate scale labels[1]. For example, let us say for variety's sake, the task is to rate cheeses on the basis of their price and quality. Suppose that there also exist a hypothesis to rate 'chese1' with rating 4. Let this cheese have price 3 and quality 10. Suppose that now the target stimulus is 'cheese2' with price 2 and quality 5. The system would detect that 'cheese2' is cheaper than 'cheese1' and would create a new comparison instance-agent about this fact. The marker that this instance emits, would come at a certain moment into the concept 'cheaper', then would continue to the concept 'better', and through an argument-related answer would receive in the correspondence–relation 'better < -- > higher rating'. Now the correspondence–relation would detect that a hypotheses to rate 'cheese1' with grade 4 exist, and hence, would trigger creation of hypotheses that involve the target 'cheese2' and the available ratings, which are higher than 4. Of course, in parallel, the relation 'cheese1 has better quality than cheese2' would emerge, and this would be a justification to rate the target 'cheese2' with lower ratings. The final decision here would depend on other comparisons with other cheeses and on the relative *relevance* of the price and the quality.

Note, that the meaning of 'better' is not strictly predefined by its prototypes. In a certain context, a different concept can be recognized as its subclass. For example, if

[1] New hypotheses can be created only for target stimuli, not for retrieved ones.

the necessary agents are active enough, the system may infer that ‘more yellowish’ is better, because yellow cheeses suit the tablecloth more.

Analogically to the comparison-relations, the correspondence ones also have orders. The ones described above are of first order. The correspondence–relations of second order create hypotheses based on comparisons of second order, etc. For

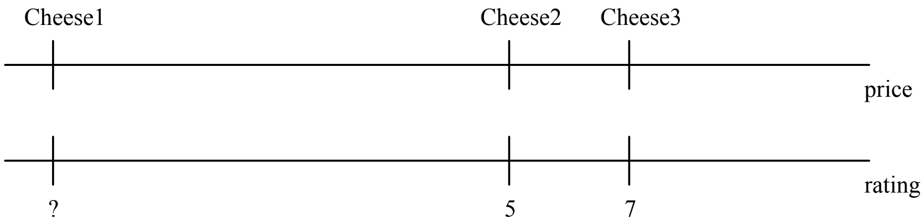

Fig. 5. 15. The correspondence–relations of second order make inferences on the basis of comparisons of second order, and the available hypotheses and ratings. (See text for details)

example, in the situation shown on fig. 5. 15, the information that the difference in price between ‘cheese1’ and ‘cheese2’ is higher than the difference between the prices of ‘cheese2’ and ‘cheese3’, together with the information that ‘cheese2’ was rated with 5, ‘cheese3’ – with 7, forms a justification to rate the target ‘cheese1’ with ratings that are smaller than 3.

The exact algorithm for creating hypotheses is the following one. When a certain correspondence–relation is ready to infer a new hypothesis, it gathers all justifications for this inference – usually one comparison and one or more hypotheses. After that, it creates a new agent – justification agent, whose only work is to create the new hypothesis. The correspondence–relation does not do this job directly, because the speed of this creation has to depend on its justifications. The speed of the processes is a very important element of DUAL specification, and it must be carefully considered. The justification agents ensure that the order of hypotheses creation would reflect their relevance to the context.

5.6 Constraint Satisfaction

5.6.1 Main Ideas

The Constraint Satisfaction Network (CSN) consists of hypotheses and justification agents, which are interconnected with the main semantic long-term network. The justifications and the hypotheses are created locally and asynchronously. Thus, the CSN emerges dynamically in parallel with the other mechanisms of the model.

The construction of hypotheses is based on reasons. In order to be created, each hypothesis has to have justifications – groups of other agents, which together describe the semantic or structural reason for the hypothesis.

No time is spent waiting for the CSN to settle in order to read out the ‘solution’ from the activation pattern. Instead, the target stimulus along gives a response when it is ready without informing all agents of this. This allows cognition to be viewed as a continuous process, without breaks between the given tasks.

Each hypothesis agent has its ‘life cycle’, making several ‘metamorphoses’. CSN, however, involves hypotheses of a different kind.

5.6.2 Hypothesis Agents

From a declarative point of view, hypothesis agents keep three main pieces of information – the two agents being mapped (called hypothesis *arguments* or just *elements*), and a list of justifications for this correspondence[1]. Each one piece of information is written in a separate slot (Fig. 5. 16).

```
Line15<-->grade3
  :type (:mature :hypothesis :temporary)
  :t-link ((Line15<-->grade5  – 0.3)
          (Line10<-->grade3  – 0.3))
  :slot1
    :c-coref line15
  :slot2
    :c-coref grade3
  :slot3
    :c-coref (justif->line15>line6==grade2
             justif-> grade5==line20>line15)
```

Fig. 5. 16. An example of a hypothesis agent. It represents the hypothesis to rate line15 with grade3; two justifications for this, and two alternative hypotheses (competitors).

As is shown in figure 5.16., there is one more G-slot labeled :t-link. These links are temporary and connect hypotheses to hypotheses. They serve only the connectionist aspect of the architecture.

The AMBR model postulates two types of :t-links – excitatory (links between coherent hypotheses), and inhibitory (links between conflicting hypotheses). The latter embody the *one-to-one mapping constraint*. This means that, for example, the hypotheses X<-->Y and X<-->Z are viewed as contradictory, because X should not be mapped to two or more elements, e.g. Y and Z. This constrain is well known in analogy making, but is more complicated in a judgment task. It is obviously that one and the same stimulus should not be mapped onto two different ratings, but the opposite is possible – different stimuli can be mapped onto the same rating.

Our belief was that the one-to-one mapping constraint is a more fundamental principle and can also be detected outside the boundaries of the task for analogy-making. It plays a very important role in DUAL–based models. The architecture combines the mechanisms of associative memory and relevance-based mappings. Without a pressure for one-to-one mapping the most active elements would be mapped onto each other again and again. Thus, the system would not be able to run out of its fixation, and would not be able to search for other paths. Thus, the pressure for one-to-one mapping balances the associative and mapping mechanisms.

This, however, does not mean that it is not possible for different stimuli to be judged with the same rating. The tendency for one-to-one mapping is only a pressure – one of many constraints. Only the agents, marked as :t-driver have to report *only one* of their winner hypotheses. For the judgment task those :t-driver agents are the judged stimuli (one or more), whereas in the choice task the maximal rating is marked

[1] In AMBR model the hypotheses carry out in addition the information about the respective situation. JUDGEMAP, however, does not need his information, and the respective slot stays empty.

as :t-driver and has to report only one stimulus that corresponds to it, i.e. the final choice.

JUDGEMAP uses only inhibitory :t-links. The reason is the following: In AMBR, justification agents are not postulated, because in analogy-making the justifications are always single agents, and it is not necessary to create their replications. Thus, the coherent hypotheses are connected directly with excitatory links. In JUDGEMAP, however, the justification agents serve as mediators between the coherent hypotheses (fig. 5.17), and thus the direct links between the hypotheses are only inhibitory.

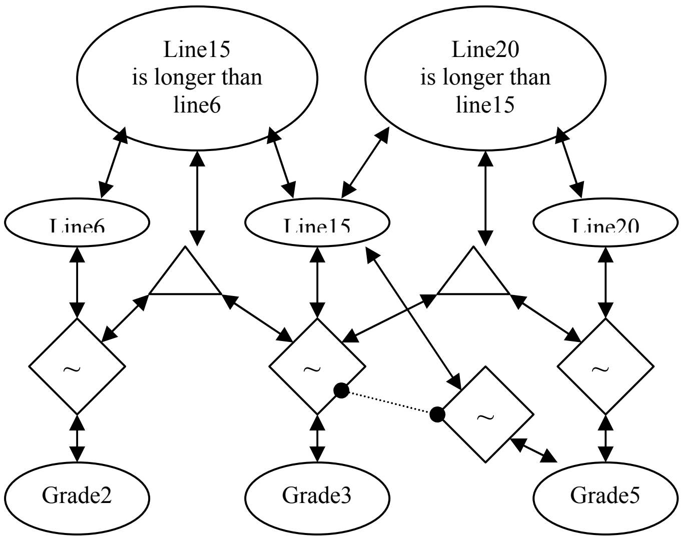

Fig. 5. 17. An example of the connections between entities (small ovals), comparison instances (large ovals), justification agents (triangles), and hypothesis agents (diamonds). The conflicting hypotheses line15<-->grade3 and line15<-->grade5 are connected with an inhibitory link. (Compare with the transcript in fig. 5.16.)

5.6.3 Secretary Messages

Maintaining inhibitory links poses one problem. There is not any one central executor in JUDGEMAP that 'sees' all hypotheses, identifying the conflicting ones. Rather, the hypotheses emerge one by one, and their creators have local information only. Then, how does the hypothesis X<-->Y 'know' that there is a rival hypothesis (X<-->Z) in order to create an inhibitory link to it?

The solution is for the hypothesis to 'ask' its elements. In short, the first job that a hypothesis agent does after its creation is to send *hypothesis-registration-requests* to its two arguments. These requests are symbolic structures that carry information about the name of the hypothesis-sender, and also it's other argument. The job of the instance-agent that receives such a request is to make sequentially two things: First, if the hypothesis is novel (it does not duplicate an already existing one), it registers it in a special buffer of its own, called secretary. This registration has several purposes – it is used in analyzing the next registration requests; supports the rating mechanism, responsible for finding out the final winners (later in the text), etc. Second, the respective instance-agent sends back to the hypothesis *a hypothesis-registration-answer*.

The hypothesis-registration-requests are symbolic structures that, roughly speaking, state one of the two orders – *resign* or *establish*.

The answer 'resign' means either that it comes too late – there is already a winner, or that the new hypothesis is a duplicate of an existing one. To be a duplicate means that it connects the same two arguments, probably for a different reason. For example, suppose that line-A is longer than line-B, and line-B was judged with 6. This would serve as a justification to rate line-A with, let's say, 4. It is possible at a later moment for the information that line-A is shorter than line-C, and line-C was judged with 3 to spring up, and hence the system would construct another justification-agent that would create a second hypothesis about the same – judge line-A with 4.

The answer 'resign' also carries information about the hypothesis in favor of whom to resign. When a hypothesis agent receives such an answer, it hands over all its justifications to the favorite, and dies. In such a way the hypotheses agents usually collect many justifications, which support them with activation.

The answer 'establish' means that the correspondence is a novel one. This answer also carries a list of the conflicting hypotheses, to which inhibitory links have to be added.

5.6.4 Life Cycle of the Hypothesis Agents

There are three main types of hypotheses – *embryo*, *mature*, and *winner*; the type is marked in their :type slot.

Each hypothesis begins its life cycle as an embryo. During this period, it sends hypothesis-registration requests to its respective arguments and waits for answers. When both answers arrive, the hypothesis analyzes them and makes a decision. Three variants are possible – both answers to be 'resign', both to be 'establish', or to contradict each other (the last case is possible because of the asynchronous and parallel nature of the model). If the answers are contradictory, the hypothesis carefully checks what the reason is, and then decides what to do next. The possible reasons for contradictory answers are three. First, maybe a duplicate exists, but one of the arguments was not yet informed about it. Second, the answer 'resign' has come because one of the arguments already had a winner hypothesis, i.e. there is no more need of the hypothesis – the system has finished the tasks concerning the respective entity. Third, maybe there was a duplicate, but it has died (again one of the arguments was not yet informed about this). The new hypothesis should resign, if the first or the second event has happened, or be established in the third case.

If the embryo hypothesis decides to resign, it hands over all its justifications to its duplicate (if such mentioned in the hypothesis request answers), and fizzles out. This is the mechanism that allows for multiple justifications for one and the same hypothesis, instead of having redundancies of correspondences.

Otherwise, if the embryo decides to establish, it transforms itself to a mature hypothesis. In this case, it connects itself to its competitors (mentioned in the hypothesis request answers) with inhibitory links. For fair play, it informs those competitors to do the same.

The third phase of the hypothesis's life cycle begins if they receive a special message from one of the arguments that promotes it into a winner. The mechanisms of rating and promotion are described in the next sub-section.

5.6.5 Rating and Promotion

These two mechanisms were proposed by the AMBR model (Petrov, 1988). They may look too complicated for the purposes of judgment and choice, where the task is to report in the end for only one correspondence – of the judged stimulus (or several, if several stimuli are given to judge simultaneously, but always not too many). However, the mechanisms have been designed to work in analogy–making between huge domains, where many agents have to find corresponding elements. The architecture' requirements were this to be done locally, without any central executors.

The main ideas are the following: Some of the instance-agents (the :t-driver ones) are authorized to use the rating mechanism. The purpose of this mechanism is to monitor all hypotheses that involve the agent and to send *promotion incentives* to those that emerge as stable and unambiguous leaders.

Each authorized instance keeps a data structure, called *rating table*. The *individual ratings* for each registered mature hypothesis are stored in this table. Individual ratings are just a numbers that characterize the relative success of the respective hypothesis – how long, how recently, and how strongly has it been a leader, according to its rivals. The instance-agents periodically (in a fixed time interval and whenever a hypothesis registration request come) adjust the individual ratings. The rating of the current leader increases, whereas those of the other hypotheses decrease. The amount of the change is proportional to the difference between the activation levels of the leader and its closest competitors.

When the individual rating of some hypothesis exceeds a certain threshold, the respective instance-agent sends to it a symbolic structure, called *promotion incentive*. In addition, it eliminates all loser hypotheses. When a hypothesis agent receives such message, it transforms itself to winner hypothesis. If the task is for judgment, this winner is the response of the system about the respective stimulus.

5.6.6 Hypothesis Activation Functions

Hypothesis agents differ from the other agents in that they receive both positive and negative activation. That is why they have two separate input values – **enet** and **inet**. The former is equal to the total sum of all input activation that comes from excitatory links, whereas the latter – to the total sum of the inhibitory activation.

The calculation of the current activation level involves three steps:

First, the old activation level is adjusted with a linear transformation that ensures that the adjusted value would be non-negative:

$$\mathbf{a^{|} = a + Z},$$

where $\mathbf{a^{|}}$ is the adjusted value, **a** is the original one, and **Z** is hypothesis zero level – the level, to which activation strives if the input is zero.

Second, the new activation level is calculated, according to the formula:

$$| \, a^{|}_{new} = m, \quad \text{if } -d^* a^{|} + E^*enet^*(M - a^{|}) + I^*inet^*(a^{|} - m) < m,$$

$$| \, a^{|}_{new} = M, \quad \text{if } -d^* a^{|} + E^*enet^*(M - a^{|}) + I^*inet^*(a^{|} - m) > M,$$

$$| \, a^{|}_{new} = -d^* a^{|} + E^*enet^*(M - a^{|}) + I^*inet^*(a^{|} - m), \text{ otherwise.}$$

Here $\mathbf{a}'_{\mathbf{new}}$ is the new adjusted activation level; **d** and **E** are respectively decay and excitatory constants; **M** and **m** are respectively the maximal and minimal value the activation can reach; and **I** is a parameter that reflects the power of inhibition between hypotheses. **a** is the original (non-adjusted) value, **enet** and **inet** are respectively the total excitatory and inhibitory inputs.

Third, the activation level is again re-adjusted with inversed linear transformation:

$\mathbf{a} = \mathbf{a}' - \mathbf{Z}$,

where **a** is the original value, $\mathbf{a}'$ is the adjusted one, and **Z** is hypothesis zero level.

These formulas ensure that the adjusted activation level has the same properties as the activations of the other agents in DUAL.

The hypothesis output function also has some specifics. The novelty concerns the embryo hypotheses. Their output is always equal to zero, i.e. from a connectionist point of view they do not influence the other agents.

5.6.7 The Constraint Satisfaction Network

In the course of time many hypothesis agents, interconnected each other and with the other agents, emerge. They, together with their justifications, form the Constraint Satisfaction Network. Its purpose is to solve the constraint satisfaction problem by the cooperative work of all hypotheses, justifications, and entities. The network involves several kinds of links:

The links from LTM to CSN represents the semantic constraints – the more relevant a certain entity is, the more relevant are its correspondences. The links with opposite direction – from CSN to LTM - allow the CSN to influence the overall pattern of activation – the better a certain inference looks, the more attractive its elements are.

The links from LTM to the justification agents, and those from justification agents to the hypotheses represent the structural constraints – they reflect the instructions for the task as well as some implicit assumptions behind these instructions.

The opposite links – from hypotheses to justifications and then to LTM also play an important role. They could be viewed as feedback from the hypotheses to the entities about how relevant the constructed hypotheses are.

Finally, the inhibitory links between competing hypotheses represent the pressure for one-to-one mappings – one very important for analogy–making constraint.

5.7 Putting Everything Together

5.7.1 Presentation of the Task and Judgment of the First Stimulus

This subsection illustrates how all mechanisms of the model work together. Let the task be to judge sequentially lengths of lines on a seven-point scale.

The session begins by attaching the correspondence–relation 'longer<-->higher rating' to GOAL node, and the concept 'seven-point scale' to INPUT node. Their micro-frames are predefined and are shown in fig.5.18. Initially, the prototypes of the

```
longer<-->higher rating
  :type :concept
  :subc correspondence
  :argument (longer seven-point-scale)
  :a-link (higher-length--higher-rating 0.200)

seven-point-scale
  :type :concept
  :subc scale
  :t-instance (grade-1 grade-7))
```

Fig. 5.18. At the beginning of the simulation, the correspondence relation ‘longer<-->higher rating’ is attached to the GOAL node, the concept ‘seven-point-scale’ – to the INPUT.

scales are assumed to be their end-points. The links to them are temporary and their weight decreases with the course of time. The initial choice of prototypical ratings, however, matters only for the first several judgments. It does not influence the statistical results over judgment of large sets of stimuli.

Now, it is time to choose also one line and to attach it to both GOAL and INPUT nodes. Let it be ‘line-200’ with length 200 (fig.5.19)

```
line-200
  :type :instance
  :inst-of line
  :c-coref length-200)
```

Fig.5.19. The agent ‘line-200’ represents the first stimulus to be judged and is attached both to GOAL and INPUT nodes.

The activation spreads in the network and brings relevant concepts and instances into WM. Fig.5.20 illustrates at which moment the agents pass the threshold of the WM (the agents, attached to GOAL or INPUT are not reported). The markers from the instance-agents (scale labels, 'line-200', 'length-200, etc) spread upward to their super-classes, and then to other concepts as argument-related answers. The concepts

```
;;T=0.000, #$seven-point-scale enters in WM
;;T=0.000, #$longer--higher-rating enters in WM
;;T=0.000, #$line-200 enters in WM
;;T=0.100,<ARG-REL-REQ-#$longer--higher-rating> receives in #$seven-point-scale
;;T=0.200, #$grade-1 enters in WM
;;T=0.200, #$grade-7 enters in WM
;;T=0.200, #$length-200 enters in WM
;;T=0.300, #$grade-0 enters in WM
;;T=0.300, #$grade-2 enters in WM
;;T=0.300, #$grade-6 enters in WM
;;T=0.300, #$grade-8 enters in WM
;;T=0.300, #$relation enters in WM
;;T=0.300, #$object enters in WM
;;T=0.300, <MARKER-#$line-200> receives in #$line
;;T=0.300,<ARG-REL-REQ-#$longer--higher-rating> receives in #$longer
;;T=0.400, #$number enters in WM
;;T=0.400, #$property enters in WM
;;T=0.400, <MARKER-#$grade-7> receives in #$seven-point-scale
;;T=0.400, <MARKER-#$grade-1> receives in #$seven-point-scale
;;T=0.500, #$grade-3 enters in WM
;;T=0.500, #$grade-5 enters in WM
;;T=0.500,<ARG-REL-REQ-#$longer> receives in #$line
;;T=0.600, #$grade-4 enters in WM
```

Fig.5.20. Part of the transcript of the simulation run is showed. The first column (T=0.20) shows the time moment, the next point the event. ARG-REL-REQ-*agent* means argument-related request, sent by *agent*, MRK-*agent* means marker, emitted by *agent*.

that receive markers check whether the origins are their direct instances, and if yes – create :t-instance links to them (see in fig.1.21. the continuation of the transcript).

```
;;T=0.600, <MARKER-#$length-200> receives in #$length
;;T=0.900, <ARG-REL-ANS-#$grade-7> receives in #$longer--higher-rating
;;T=0.900, <MARKER-#$grade-6> receives in #$seven-point-scale
;;T=0.900, <MARKER-#$grade-2> receives in #$seven-point-scale
;;T=1.000, #$more-higher-length enters in WM
;;T=1.200, #$more-longer enters in WM
;;T=1.200,<ARG-REL-REQ-#$longer> receives in #$higher-length
;;T=1.200,<ARG-REL-REQ-#$higher-length--higher-rating> receives in #$seven-point-scale
;;T=1.200,<ARG-REL-REQ-#$higher-length> receives in #$length
;;T=1.300, <ARG-REL-ANS-#$line-200> receives in #$longer
;;T=1.400, <MARKER-#$grade-5> receives in #$seven-point-scale
;;T=1.400, <MARKER-#$grade-3> receives in #$seven-point-scale
;;T=1.400, <ARG-REL-ANS-#$grade-1> receives in #$longer--higher-rating
;;T=1.600, <MARKER-#$grade-4> receives in #$seven-point-scale
;;T=1.900, <ARG-REL-ANS-#$grade-6> receives in #$longer--higher-rating
;;T=2.200, <MARKER-#$length-200> receives in #$property
;;T=2.200,<ARG-REL-REQ-#$more-higher-length> receives in #$higher-length
;;T=2.400, <ARG-REL-ANS-#$length-200> receives in #$higher-length
;;T=2.600, <ARG-REL-ANS-#$grade-2> receives in #$longer--higher-rating
;;T=2.700, <ARG-REL-ANS-#$grade-6> receives in #$higher-length--higher-rating
;;T=2.800, <ARG-REL-ANS-#$grade-2> receives in #$higher-length--higher-rating
;;T=2.800, <ARG-REL-ANS-#$grade-7> receives in #$higher-length--higher-rating
;;T=2.900, <ARG-REL-ANS-#$grade-1> receives in #$higher-length--higher-rating
;;T=3.100,<ARG-REL-REQ-#$more-longer> receives in #$more-higher-length
;;T=3.500,<ARG-REL-REQ-#$higher-length--higher-rating> receives in #$more-longer
;;T=3.600, <ARG-REL-ANS-#$grade-5> receives in #$longer--higher-rating
;;T=3.600, <ARG-REL-ANS-#$grade-5> receives in #$higher-length--higher-rating
;;T=4.900, <ARG-REL-ANS-#$grade-3> receives in #$longer--higher-rating
;;T=4.900, <ARG-REL-ANS-#$grade-3> receives in #$higher-length--higher-rating
;;T=5.000,<ARG-REL-REQ-#$more-longer> receives in #$longer
;;T=5.800, <ARG-REL-ANS-#$grade-4> receives in #$longer--higher-rating
;;T=5.800, <ARG-REL-ANS-#$grade-4> receives in #$higher-length--higher-rating
```

Fig. 5. 21. Continuation of the transcript form fig.5. 20. Again only part of the events is presented.

However, no other lines and lengths would enter in WM, and no comparisons would be formed. This initial moment – the judgment of the first stimulus needs a special comment. The system must judge in a ‘vacuum’ – without anything to compare the stimulus with. My belief is that this never happens to people. They have extremely huge knowledge bases and in all cases, they are able to retrieve (or to construct) something similar to the target. One possibility to deal with this problem in JUDGEMAP is to use some predefined prototypes of the concept ‘line’ – may be the two extremes (assuming that people construct them from the boundaries of the screen), or the middle one. The additional problem then would be that these prototypes would bias judgments in an artificial and unwanted manner. For this reason, a different solution was chosen in the model. If and only if such a case appears – the target stimulus has not found any rating to map it to for a long time - the system creates one hypothesis without any reason (justification). It just takes the most active rating and maps the target onto it. (fig. 5. 22).

```
A winner by scrap: Stimulus #$line-200 <-> Grade #$grade-4
In time 249,90 the stimulus #$line-200was judged with rating #$grade-4 WM - 24 ag.,act 29.543 Justs: 0
```

Fig. 5.22. In some pathological cases, a hypothesis can be created without reasons.

5.7.2 Judgment of the Next Stimuli

Once the first stimulus has been judged, it is removed from the GOAL list but not from WM, i.e., it still receives activation from its concept via :t-instance link. However, removing the respective stimulus from the GOAL list causes erasing the tag :t-driver from its :type slot.

Let the second stimulus to be judged be ‘line-500’ with length 500. The initialization procedure is the same – the stimulus is attached to the GOAL, and :t-driver tag is added in its :type slot. The new stimulus emits marker, activates its neighbors, etc.

```
;;T=249.900, #$line-500 enters in WM
;;T=250.000, <MARKER-#$line-500> receives in #$line
;;T=250.100, #$length-500 enters in WM
;;T=250.400, <MARKER-#$length-500> receives in #$length
;;T=250.700, <ARG-REL-ANS-#$line-500> receives in #$longer
;;T=250.900, <MARKER-#$line-500> receives in #$object
;;T=251.400, <ARG-REL-ANS-#$length-500> receives in #$higher-length
;;T=251.600, <MARKER-#$length-500> receives in #$property
;;T=252.900, #${length-500}>{length-200} enters in WM
```

Fig.5.23. The beginning of the judgment of the second stimulus – ‘line-500’.

At time T=250.7 (fig. 5. 23.) the comparison-relation ‘higher-length’ receives an argument-related message about ‘length-500’. Now the comparison-concept is ready to apply its symbolic routine – at time T=252.9 it creates a new instance that represents the information that ‘length-500’ has a higher amount than ‘length-200’.

```
;;T=256.300, <MARKER-#${length-500}>{length-200}> receives in #$higher-length
;;T=257.600, <ARG-REL-ANS-#${length-500}>{length-200}> receives in #$longer
;;T=257.900, <ARG-REL-ANS-#${length-500}>{length-200}> receives in #$more- higher -length
;;T=258.400, <MARKER-#${length-500}>{length-200}> receives in #$relation
;;T=258.700, #${line-500}>{line-200} enters in WM
;;T=261.800, <MARKER-#${line-500}>{line-200}> receives in #$longer
;;T=263.000, <ARG-REL-ANS-#${line-500}>{line-200}> receives in #$longer—higher-rating
;;T=263.100, #$justif-\#${line-500}>{line-200}- enters in WM
;;T=263.200, <ARG-REL-ANS-#${line-500}>{line-200}> receives in #$more-longer
;;T=263.400, <MARKER-#${line-500}>{line-200}> receives in #$relation
;;T=281.800, #$line-500<==>grade-5 enters in WM
;;T=286.200, #$line-500<==>grade-6 enters in WM
;;T=292.700, #$line-500<==>grade-7 enters in WM
At time 292.900 the stimulus #$line-500 was judged with rating #$grade-5 WM - 32 ag.,act 33.641 Justs: 1
```

Fig.5.24. The rest of the judgment of the second stimulus – ‘line-500’.

This new instance (like each one instance-agent) also emits a marker, which arrives in the comparison-relation ‘longer’ (the arguments of ‘longer’ are lines) at time 257.6 (fig.5. 24.). ‘Longer’ detects for with two objects the comparison between lengths is responsible – respectively ‘line-500’ and ‘line-200’ – and creates the instance-agent ‘line-500 is longer than line-200’[1].

[1] It may seem that having two separate comparisons about one and the same information is unnecessary redundancy. The advantages of such separation, however, are more salient when judging complex stimuli, according to several relevant dimensions. For example, it is important to separate the relation ‘better cheese’ from the relations about cheese’s properties – ‘better price’ or ‘better quality’.

At time 263.0 the correspondence–relation longer<-->higher-rating ‘hears’ about this new instance. It creates a justification agent and interconnects it with the comparison ‘line-500 is longer than line-200’ and to the correspondence line200<==>rating4. The correspondence–relation also sends a message to the new justification agent with instructions what hypotheses to create, namely – hypotheses that ‘line-500’ may correspond to ‘grade-5’, ‘grade-6’ or ‘grade-7’. The justification agent creates them prolonged in time, one after another, in order that reflects the activation levels of the hypothesis’s arguments.

At time 281.8 the first hypothesis is born – line-500<==>rating-5. At this moment the rating mechanism begins to work – at each step it increases a little bit the individual rating of the hypothesis.

At time 286.2 a competitor arrives - line-500<==>rating-6. However, it is not strong enough to stop the run of the first one. At time 292.9 the promotion mechanism proclaims line-500<==>rating-5 a winner (the emergence of the third hypothesis in fact does not matter – it comes too late). The response is reported; all loser hypotheses are fizzled out, and ‘line-500’ is removed from GOAL list.

Let the third stimulus be ‘line-1400’. Let us skip some details and look at a later moment which hypotheses compete with each other and what their support is.

The first hypothesis proposes to judge the stimulus with rating-7. It receives support from three justifications – line-1400 is longer than line-200, judged with 4; it is also longer than line-500, judged with 5, and ‘line-1400 is longer than line-200’ is more than ‘line-1400 is longer than line-500’. It also receives activation from their elements, respectively ‘line-1400’ and ‘rating-7’. In addition, the hypothesis is inhibited from its competitors.

The second hypothesis proposes to judge the stimulus with rating-6. It receives activation only from two justifications – the first and the second justification of the previous hypothesis. The competition between these two hypotheses, however, was long, because the second hypothesis receives more activation from its arguments than the first one. This happens, because the agent ‘grade-6’, being a winner in the previous trial, is very relevant.

One more hypothesis – about rating-5, also exists, but it is too weak to do anything more. In the end, however, the first hypothesis wins the competition and line-144 is judged with 7.

Let the fourth stimulus to be judged be ‘line-300’ with length 300. One may think that now the system is close to a ‘combinatory explosion’ – too many comparisons and correspondences could be created. However, this is not the case, because of the special role of the working memory. In fact, at the moment when the hypothesis ‘line-300<==>rating-4’ was promoted, it was supported only by 2 justifications, namely ‘line-1400 is longer than line-300’, and ‘line-500 is longer than line-300’. There could be found more possible reasons to judge line-300 with 4, but most of them do not have enough time to be established. The rating and promotion mechanisms do not wait until all moderately relevant hypotheses emerge. Instead, they report immediately if the current state is satisfactory enough. The competition during judgment of line-300 was so short mainly because the competing hypotheses were too weak. During the previous judgments, the system was ‘focused’ on the larger ratings (i.e., that they were quite active). Because each hypothesis receives activation not only from its justifications, but also from its elements, the small ratings lose the competition. In addition, the contradictory information – that line-200 was judged with 4, has lost much of its activation due to the decay.

Later in the judgment process the old instances fizzle out one after another. Note that the order of their fizzling out is not strictly determined by the order in which they have been judged. Instead, fizzling out reflects inversely their relevance. If a certain line justifies many winner correspondences, it also receives feedback activation from them and hence stays in WM longer. If it is inconsistent with the recent winners, it loses its support and dies.

5.7.3 Variants of the Judgment Task

JUDGEMAP has the potential for dealing with variations of the judgment tasks.

It does not put any limitations on the complexity of the stimuli or on the scales. If the stimulus has to be judged according to two or more dimensions, each of them is involved in creation of justifications, hypotheses, etc. on its own, but all the acts run in parallel and thus influence each other. The final result depends on the collective effort of all dimensions.

It is possible to judge several stimuli simultaneously – in this case all of them are attached to the INPUT list and marked as :t-drivers. It is also possible to model the so called anchoring procedure (one stimulus is presented together with the suggested rating while a second one has to be rated) – in this case the first item is attached to the INPUT list without marking it as :t-driver, while the second one is attached to the INPUT list as :t-driver.

A special case of judgment task would be if a set of stimuli is given and one has to judge which of them would best correspond to a certain scale label. (Which one would be the best 3, or the best 5). JUDGEMAP is able to simulate this task by attaching the :t-driver tag to the rating label rather than to the stimuli. In the very special case when we use a two-point scale, the task could be interpreted as a choice task: "Please, find the stimulus that best fits the higher rating, i.e. the one that is preferred".

From this point of view, JUDGEMAP can also be treated as a model of choice.

5.7.4 The Role of these Mechanisms

The behaviour of the JUDGEMAP model results from the combined work of many mechanisms. Some of them may seem too complicated or even unnecessary for the judgment task. However, JUDGEMAP is part of the DUAL architecture and thus the used mechanisms cannot be considered just as mechanisms of judgment. Instead, JUDGEMAP can be treated as a system of basic mechanisms that account for judgment, analogy-making, choice, and decision-making. Some of the proposed mechanisms may look inessential for one or another of the mentioned cognitive activities but these mechanisms are mutually integrated and form a united whole.

However, it is important to analyse the exact role of each one of the mechanisms in the process of judgment.

The spreading activation mechanism defines the working memory, thus it is responsible for the construction of the comparison set. Activation represents the relevance of each agent and determines the speed at which it performs symbolic operations.

The spreading activation mechanism is undistinguishable from the associative organisation of memory. The close associations of the most relevant items become also very relevant and thus their influence on the process becomes greater. This is the main source of the assimilation effects, observed by JUDGEMAP. More precisely, the decision of the system to judge a certain stimulus with a certain rating increases the relevance of the respective rating together with its neighbours and thus sequential assimilation is produced.

The marker-passing mechanism serves to find a path between two items. It is an excellent mechanism for making an analogy between two distant domains. In JUDGEMAP, however, a very small amount of its power is actually used. The markers and the other messages that the agents exchange during the judgment process usually spread through not more than four or five agents. Thus, it is possible JUDGEMAP to work using a simplified version of the mechanism. However, for the purpose of integration between the DUAL-based models, the marker passing mechanism is enlarged with the sub-mechanisms of exchanging argument-related messages.

The comparison relations serve for a kind of recognition. Their work is essential for the construction of a structure among the comparison set, based on the ordering relation. This structure is then mapped to the other ordering-relation based structure, namely the structure among the available scale labels. As the constructed comparison relations highlight the difference between the items (they create pressure even too close for different stimuli to be judged with different ratings), the mechanism for recognition of relations is one of the main sources of the contrast effects, demonstrated by JUDGEMAP (the other main source is the pressure for one-to-one mapping).

The comparison relations of second order allow the properties of the interval scales to be combined with the principle for local computations only. Without the relations of higher order, the model would be not able to differentiate, for example, the set of magnitudes 1, 2, and 10 from the set of magnitudes 1, 5, and 10. However, people judge differently these two sets. This differentiation is one of the manifestations of the Range effect (Parducci, 1968; Parducci & Perret, 1971; Weddel, Parducci & Geiselman, 1987; Mellers & Cooke, 1994).

In fact the correspondence relations create hypotheses for correspondences and thus are the basis for the construction of the constraint satisfaction network. The work of the correspondence relations is closely related to the work of the justification agents. In order for a new hypothesis to be created, the common effort of several agents is necessary – more often one correspondence relation, one comparison instance and one hypothesis. However, which of them is to decide how fast the new hypothesis should be created? As was already mentioned above, one of the basic constraints assumed by the DUAL architecture is that the relevance of each item determines the speed of its symbolic operations. The justification agents receive their activation from all justifications for the potential new hypothesis, thus their activation level represents the overall relevance of this potential new hypothesis and determines the speed at which to create it.

The hypotheses for correspondences are the basic components of the constraint satisfaction network. They have highly complex processing equipment, namely an elaborated life cycle and sub-mechanisms for rating and promotion inherited by the AMBR model. This complexity is important for analogy-making between two large situations based on emergency from local interactions only. For the judgment process, however, only a small part of their power is used because the mapping is between a relatively small number of items and relations. However, JUDGEMAP is designed to allow the same basic mechanisms to be used for modelling different cognitive processes.

CHAPTER VI
Simulations

6.1 Overview of Simulations

In order to test the behavior of the model, several simulations organized in five groups were performed. In the simulations of the first group the distribution of the stimulus set was varied. The role of the scale was explored in the second group. The simulations in the third group tested the influence of a single contextual stimulus on judgments. In the fourth group, the judged stimuli had two relevant dimensions. The fifth group consisted of only one simulation – the capability of the model to make choice was tested.

The stimuli from the first three groups were labeled 'lines' and their judged property was labeled 'length'. As mentioned above, the names of the DUAL-agents are arbitrary, i.e., they can be named ag0001, ag0002, etc. without changing the behavior of the system. However, it is much easier to document the model using names that are more mnemonic. From this point of view, the capabilities of JUDGEMAP to judge stimuli that have a single relevant dimension were illustrated with judgments of lengths of lines. Analogically, the simulations in the fourth group demonstrated how the model can judge stimuli with respect to two relevant dimensions, independently of the names used for the stimuli. In the simulation from the fifth group, JUDGEMAP made a choice between alternatives, called 'gambles", according to their two relevant properties.

In all simulations the stimuli were represented by coalitions of DUAL-agents. One agent was responsible for the stimulus itself and others, for its properties (relevant to the task or not). Each agent pointed to its super-class. There were :c-coref links between the head agent and all the properties. The agents that represent properties could have an :amount slot that was filled in with the magnitude of the respective property.

Each scale was represented by one concept-agent responsible for the scale itself and a set of agents representing each scale label. All scale labels pointed with :inst-of links to the head concept. Each scale label also had an :amount slot filled in with the respective number. The neighboring ratings were interconnected with associative links.

The comparison-relations that were used by the model stayed in LTM. They would be activated by the respective correspondence relations. For example, if the agent 'longer<-->higher-rating' was active enough, it would also activate the agents 'longer' and 'higher-rating'.

For each simulation, correspondence relations from first and second order were designed, in order for the instruction to be represented.

All simulations were performed using the same pattern of the values for all parameters of the model.

Overview of the simulations.

The first group of simulations tests the role of the overall set of stimuli in judgment. It consists of four simulations.

In the first one a uniformly distributed set is judged. It tests whether the overall correlation between the magnitudes and the ratings is satisfactory; whether the distribution of the ratings is uniform or not; whether the sequential assimilation effect can be observed; and whether the size of the Working Memory stays stable in the course of time.

In the second simulation the model judges a subset of a restricted range and the results from this simulation are compared with those from the previous one. Thus, it is tested whether the model is capable of reproducing the Range Principle.

The third simulation tests the Frequency principle. The model judges a skewed subset and the results from this simulation are compared with those from the first one.

The stimuli judged in the fourth simulation are uniformly distributed but a dimension irrelevant to the task separates them into two subsets. The first of these subsets is positively skewed according to the relevant to the task dimension and the second one – negatively skewed. Thus, this simulation tests a novel prediction, that the irrelevant to the task dimensions take part in judgment.

The second group of simulations tests the behavior of the model when the scale varies. A uniformly distributed set and a skewed subset are judged on a three-point scale and on a 100-point scale. Thus, the relative weight of the frequency pressure is calculated on a 3-point and a 100-point scale. This weight is calculated also on a 7-point scale using the results of the simulations from the first group. After that all weights are compared in order to test whether the role of the skew decreases when the number of scale labels increase as the experimental data (Wedell, Parducci & Lane, 1990) shows.

The third group consists of four simulations and two psychological experiments. In the first three simulations a uniformly distributed set is judged but together with a single contextual element (*anchor*). In the first simulation the anchor is outside the stimulus range, in the second one, in the upper limit and in the third one, in the middle of the stimulus set. The results are compared with those received in a psychological experiment, which tested the hypothesis that the presence of an anchor (in the limit or in the middle of the stimulus set) 'improves' judgment, i.e., causes the small stimuli to be judged with lower ratings and the large stimuli to be judged with higher ratings.

In the fourth simulation from this group JUDGEMAP judges pairs of stimuli simultaneously. The received contrast effect was then confirmed by a second psychological experiment.

In the fourth group of simulations two-dimensional stimuli are judged. The capabilities of the model to reproduce the sequential assimilation effect, the range and frequency effects when judging complex stimuli are tested. In the first two simulations two sets of stimuli in which the two dimensions correlate are judged, the first one being uniform and the second, skewed. In the third simulation from this group the two dimensions of the stimuli correlate negatively, i.e. a natural trade-off judgment situation is simulated.

Finally, the fifth group consists of a single simulation testing whether the mechanisms of JUDGEMAP can be used to model choice-making.

6.2 The Role of the Distribution of the Stimulus Set

6.2.1 Uniform Distribution

In the first simulation JUDGEMAP judged 112 uniformly distributed line lengths. The simulation had two purposes. First, it served to test the overall behavior of the model – whether JUDGEMAP was capable to give higher ratings to higher magnitudes and vice versa; whether the distribution of ratings would be non-uniform, as had been demonstrated by empirical studies (Petrov & Anderson, 2005); whether the model reproduced the sequential assimilation; whether a 'combinatorial explosion' was avoided. Second, the results from this simulation served for comparison with the results from various other simulations – when judging restricted or skewed sets, etc.

Stimuli and procedure.

The lengths of a set of 112 lines were judged on a 7-point scale. The lines were separated into 14 length groups, with 8 lines in each group. The first 8 lines had length 100, the last 8 – 1400. The increment of lengths was 100. All 112 lines were randomly ordered and were presented to the model sequentially, one after another. The procedure was repeated 15 times (varying randomly the ordering of the set), thus receiving totally 1680 responses from the model.

Results and discussions.

Linearity of the scale. To estimate whether the model judges linearly the stimuli, a function of the form $\mathbf{R = a + k.S^n}$ was fitted to the data. Here **R** was the response of the model, **S** was the length of the line, **a**, **k**, and **n** were estimated parameters. The 95% confidence interval for the exponent **n** was 0.96 – 1.26. This allowed assuming that JUDGEMAP judges linearly the length.

Correct Judgment. The overall correlation between length and rating was **0.762**, $p<0.01$. This correlation is in agreement with those received in various experiments, e.g. Petrov & Anderson (2003) received a correlation 0.71, Kokinov, Hristova & Petkov (2004) – 0.88 when asking people to judge line length. The two experiments varied in their procedure and in their stimuli range. Thus, the correlation received by the model could be thought as satisfactory, i.e., *the model is capable to judge*.

This result is a consequence of the fact that JUDGEMAP makes its judgments on the basis of justifications for them. However, it is important to stress that it judges the first stimulus in each session without anything to compare it with, a fact that never occurs when people judge. Thus, the correlation received by the model was a bit lower than the one received by the empirical data.

Non-uniform distribution of the ratings. The overall distribution of the ratings is shown on fig.6.1. The standard deviation of the ratings could serve as a numerical value for estimation of non-uniformity (Petrov & Anderson 2005). The std. dev received from the simulation was **1.21** (min 1.09, max 1.28, mean 1.21, std. dev. 0.05 across the 15 runs), whereas for assumed uniform distribution it should be 2.03. People prefer the middle ratings – this observation is robust and was reported by many studies. For example, Kokinov, Hristova & Petkov (2004) received for the standard deviation of the ratings mean 1.59, min 1.15, max 1.98, std. dev 0.19.

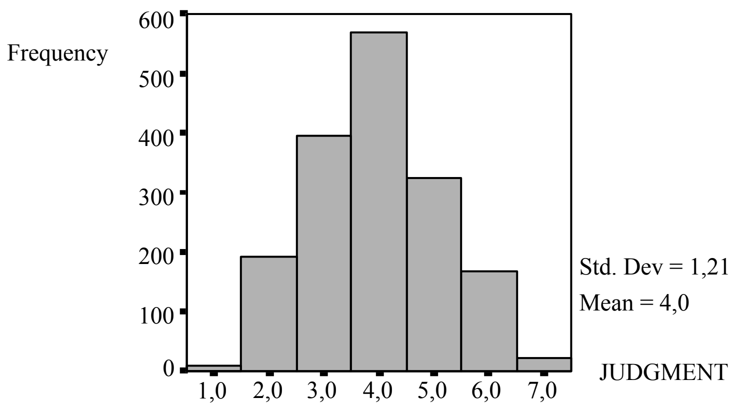

Fig.6.1. Ratings distribution when uniform set of stimuli was judged.

JUDGEMAP rarely gives extreme grades because usually hypotheses for them are created at later moments. For example, suppose that a certain line was rated 4. If the next line had to be longer, then the first created hypothesis would more probably be for 5, the second one would be for 6, etc. Being created first, these hypotheses had an advantage and only a justification with too high an activation level should be found in order to give the extreme rating.

Sequential assimilation. In order to estimate the sequential assimilation effect, a multiple regression was performed with the following variables entering as predictors: the current stimulus $\mathbf{S_t}$, the previous stimulus $\mathbf{S_{t-1}}$, and the previous rating $\mathbf{R_{i-1}}$. The estimation of the overall accuracy of the model was $R^2=0.67$, $p<0.001$. For the current stimulus $\mathbf{S_t}$, the standardized coefficient $\boldsymbol{\beta_s}$ was 0.76 (partial correlation 0.798), for the previous stimulus $\mathbf{S_{t-1}}$, $\boldsymbol{\beta_{s-1}}$ was –0.42 (partial correlation -0.432), and for the previous rating $\mathbf{R_{i-1}}$, $\boldsymbol{\beta_{r-1}}$ was +0.39 (partial correlation 0.406). The signs of the regression coefficients of the time-lagged variables were of special interest. Thus, assimilation towards the previous rating and contrast with the previous magnitude were observed. This result is in accordance with the results obtained in psychological experiments (Petrov & Anderson, 2000, Kokinov, Hristova & Petkov, 2004).

JUDGEMAP was capable to illustrate sequential assimilation because the arguments of each hypothesis were one of its sources of activation. The pattern of activation, in turn, varied dynamically. When the system gave a certain response, a temporary :t-instance link was created from the scale concept to the respective rating. As a result the activation of this respective rating grew up together with the activations of its neighbors. In this way the previous ratings assimilated the current responses.

This was not the case of the previous magnitudes, i.e., the magnitudes of the previous stimuli. On the contrary, these previous magnitudes participated in comparisons with the current one, and therefore, contrast with respect to them was observed.

Working Memory size. The next analysis checked whether JUDGEMAP avoids the so-called ‘combinatorial explosion’. It was tested how the size of the WM varies in the course of judgment. This dependency is illustrated in fig.6.2.

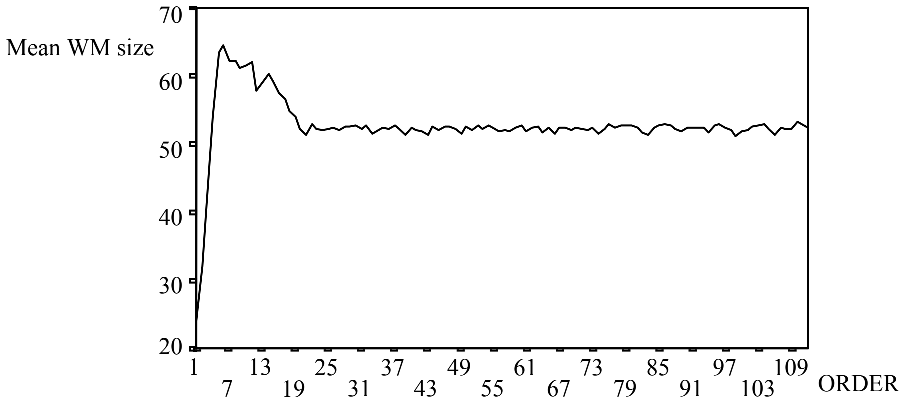

Fig.6.2. Stabilization of the size of the WM. The number of active agents (ordinate) is calculated at the moment when the system gives its response.

Fig.6.2 illustrates that the number of agents that participated in the WM first grew, around the 12th stimulus started to decrease, and stabilized around the 19th one. In other words, the system accumulated enough 'experience' with the stimuli between the 12th and 19th presented stimuli. Thus, the size of the WM stabilizes and allows the system to judge sequentially an unspecified number of stimuli.

The simulation 6.2.1 ('Uniform distribution') demonstrated several of the capabilities of the model. The correlation between length and rating was in the range of the respective correlations obtained in the psychological experiments. The ratings were non-uniformly distributed despite the fact that the stimulus set was uniform just as in the psychological experiments. The model reproduced the sequential assimilation effect as well.

However, the effects received by the model were stronger in comparison with the empirical data, e.g., when testing the sequential assimilation effect $\boldsymbol{\beta}_{s-1}$ was –0.42 in the simulation, whereas $\boldsymbol{\beta}_{s-1}$ was –0.25 in the experimental data (Petrov & Anderson, 2000); $\boldsymbol{\beta}_{r-1}$ was +0.39 in the simulation, whereas $\boldsymbol{\beta}_{r-1}$ was –0.30 in the experimental data (Petrov & Anderson, 2000). A similar tendency was observed in the other results – the non-uniformity of the responses was overestimated, whereas the overall correlation between length and rating was around the lower limit of the empirical data. However, I interpret this as a relatively small deviation from human behavior. After all, when people judge they have previous experience that helps them and also various external contextual elements (e.g. the size of the screen) influence their judgments and blur the various systematic effects. Thus, the main goal of JUDGEMAP is not to fit quantitatively the empirical data but test whether the proposed mechanisms could qualitatively produce the same effects.

In the next simulation, JUDGEMAP judged a subset with restricted range in order to test whether the model can simulate the robust Range Effect (Parducci, 1968), i.e., whether the grades depend on the minimal and maximal magnitudes in the stimulus set.

6.2.2 Restricted Range

Usually people tend to use all scale values in their judgments, i.e. the ratings of the stimuli in the restricted subset would capture the whole scale, not only the respective restricted part of it. (Parducci, 1968; Weddel, Parducci & Geiselman, 1987;

Parducci & Perret, 1971; Mellers & Cooke 1994). This is called a *range effect*. The purpose of this simulation is to test whether JUDGEMAP is able to reproduce it.

Stimuli and procedure.

Only a part of the stimuli used in the first simulation was used. From the overall set of 112 lines, those lines were extracted whose length exceeded 1000. Thus, only 10 groups of line lengths were presented, with 8 stimuli in each group. The first 8 lines had a length 100, the last 8 – 1000. The increment of length was 100. The whole set of lines consisted of 80 lines. The procedure remained the same – the overall set was judged 15 times, each time the order of the stimuli was randomly presented.

Results and discussions.

All effects observed in the first simulation were obtained in the second too. The correlation between magnitudes and ratings was **0.794**[1] (correct judgment), the standard deviation of the ratings was **1.21** (non-uniform distribution). The standardized regression coefficient for the previous ratings was **0.43** (sequential assimilation), for the previous lengths was **-0.41** (contrast from the previous magnitude).

The dependency of the ratings from the magnitudes is represented on fig.6.3, together with the respective dependency from the first simulation ('Uniform distribution') for comparison.

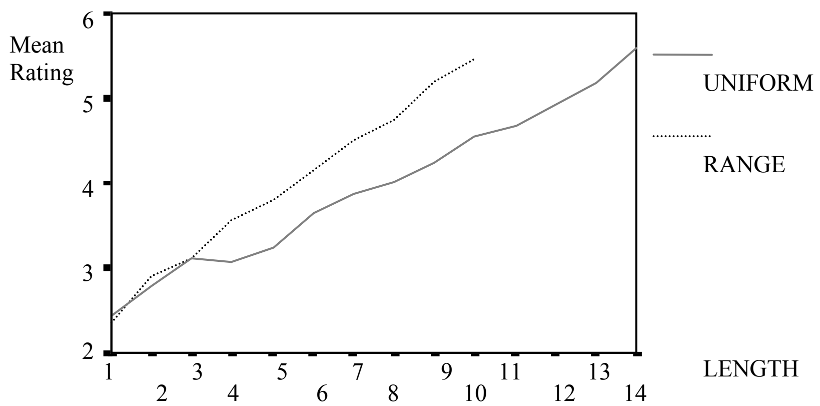

Fig.6.3. Main ratings, given by the model in two conditions – when a set of lengths varied between 100 and 1400 (solid line), and when it varied between 100 and 1000 (dashed line).

The mean ratings were aggregated by length and simulation. Then for each length (between 1 and 10) the mean ratings from the first simulation were extracted from the mean ratings from the second one. The received difference was tested with a t-test, and was found to be significantly different from zero ($t(9)=4.260$, mean difference was **0.485**, $p<0.003$). Thus, it is shown that the model adapted the given rating with the range of the stimuli's set.

The received result was an expected consequence from JUDGEMAP's work. The model creates hypotheses only on the basis of comparisons between magnitudes, and thus the ratings reflect the range of the set of stimuli. In addition, the model is able to create comparisons of higher order, i.e., comparisons between comparisons. They play an important role in the process of judgment because their work is the only

[1] All reported correlations and regression coefficients to end of the thesis were with $p<0.01$.

mechanism in the model that takes into account the fact that judged magnitudes form an interval scale, not a range one.

The result from the second simulation demonstrated that the model is capable of producing the *Range Effect* (Parducci, 1965). The next step was to test whether the model can simulate the *Frequency Effect* (Parducci, 1965).

6.2.3 Skewed Set

The frequency effect is observed when the stimulus set is not uniformly distributed but the stimuli are concentrated closer to one of the extremes. In this case, people's ratings shift away from the direction of the density (Parducci & Perret, 1971; Wedell, Parducci & Geiselman, 1987); Cooke & Mellers, 1998).

Stimuli and procedure.

Again, the third simulation used only part of the stimuli used in the first one. The overall set consisted of 56 lines that separated into 14 length groups. The lengths in the first group had magnitude 100, in the last one – 1400. The increment of lengths was 100. In the first and in the second group there were 7 lines in each, in the third and the fourth ones - 6 lines in each, etc., in the last two groups there was one line in each (see table 6.5). Thus, the overall set of 56 lines was positively skewed. The procedure remained the same – the overall set was judged 15 times, each time the order of the stimuli presentation was random.

Results and discussions.

All of the effects observed in the first two simulations were obtained in the third one. The correlation between magnitudes and ratings was **0.759** (correct judgment); the standard deviation of the ratings was **1.23** (non-uniform distribution). The standardized regression coefficient for the previous ratings was **0.36** (sequential assimilation), for the previous lengths was **-0.37** (contrast from the previous magnitude).

The dependency of the ratings from the magnitudes is represented on fig.6.4, together with the respective dependency obtained in the first (uniform) simulation for

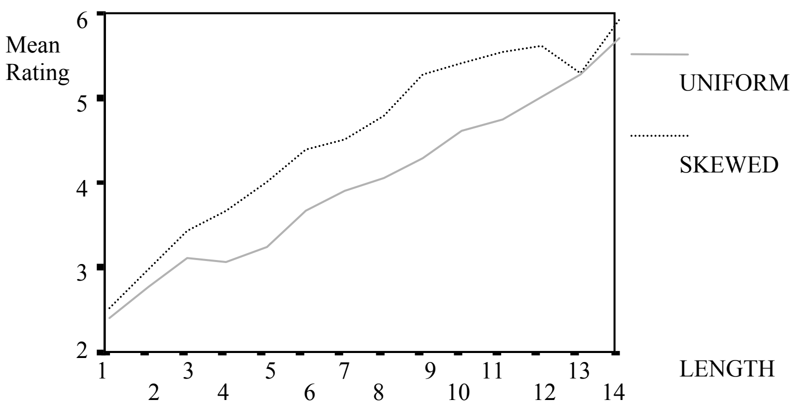

Fig.6.4. Dependency of the main judgments on the length of lines in two conditions – when a uniformly distributed set of lengths was presented (solid line), and when a positively skewed set of lengths was judged (dashed line).

comparison. The procedure for the comparison with the first simulation was the same as in the second simulation ('Restricted Range). The mean ratings are aggregated for each length and for each simulation. Then the results from the first simulation ('Uniform distribution') were subtracted from those of the current one. The received difference was tested with t-test, and was found to be significantly different from zero ($t(13)=6.690$, mean difference was **0.511**, $p<0.001$).

The received result is a consequence of the pressure for one-to-one mapping, inherited by JUDGEMAP from the AMBR model of analogy-making (Kokinov, 1994a, 1998, Kokinov & Petrov, 2001). This pressure was not included in the model for the sake of obtaining this contrast effect in the judgment task. However, it is assumed to be important for analogy–making. Thus, since JUDGEMAP is integrated with AMBR, the pressure for one-to-one mapping influences also the judgment process.

This pressure prevented JUDGEMAP from using the same rating many times, and therefore it tends to use all of them an almost equal number of times. The result is that when a positively skewed set was judged, all stimuli were overestimated. In conclusion, the result from this simulation demonstrated that the model is capable of simulating the *Frequency Effect* (Parducci, 1965) as a side effect of the assumption that judgment is based on mapping.

The next simulation was based on the ability of the model to simulate the frequency effect. It tested the behavior of the model when a uniform set of lines was judged but the set is distributed into two skewed subsets according to an irrelevant to the task characteristic.

6.2.4 The Role of the Irrelevant Dimensions

On the basis of the mechanisms of the JUDGEMAP model one novel prediction can be formulated. Because of the spreading activation the pattern of activation in the comparison set depended not only on the relevant to the task properties but also on some irrelevant ones. Suppose that the task was to judge line lengths but the lines were colored. Thus, the comparison set would consist of both red and green lines but if the target line was green, the green lines in it would be a little bit more active because of the additional activation that would spread through the concept 'green' and then back to the green instances. Consequently, the hypotheses that were justified by comparisons with green lines would have more chance to win. If the overall set was constructed in a manner that more often the green lines were shorter, those stronger hypotheses would predominantly be about the high ratings. Conversely, if more often the red lines were longer, then the red lines would be underestimated. Thus, lines with the same length would be judged with higher rating, if they were green than if they were red.

Stimuli and procedure.

Again, lengths of a set of 112 lines were judged on a 7-point scale. All lengths were distributed in the same way as in the simulation 'uniform distribution', thus forming uniformly distributed set. The lines were separated into 14 length groups, with 8 lines in each group. The first 8 lines had length 100, the last 8 – 1400. The increment of lengths was 100.

However, in addition to the length, one more line property was represented with an additional DUAL–agent – the lines were colored, either in red or in green. The 56 green lines formed positively skewed subset according to their length. The same lines used in the simulation 'skewed set' were green. The rest 56 lines were red. They formed a negatively skewed subset according to their length (see table.6.5.).

lengths	*number of the green lines*	*number of the red lines*
1 & 2	7	1
3 & 4	6	2
5 & 6	5	3
7 & 8	4	4
9 & 10	3	5
11 & 12	2	6
13 & 14	1	7

Table. 6.5. Distribution of the stimuli used in the fourth simulation.

The line color was irrelevant to the judgment task. The respective correspondence relation attached to the GOAL node required only that the longer lines receive higher ratings.

All 112 lines were randomly ordered and were presented to the model sequentially, one by one. The procedure was repeated 57 times (varying randomly the ordering of the set), thus receiving totally 6384 responses from the model.

Results and discussions.

All effects observed in the first simulation remained the same. The correlation between magnitudes and ratings was **0.765** (correct judgment); the standard deviation of the ratings was **1.90** (non-uniform distribution). The standardized regression coefficient for the previous ratings was **0.12** (sequential assimilation), for the previous lengths was **-0.18** (contrast from the previous magnitude).

However, a difference in the ratings of the lines was found depending on their color (see fig.6.6). The mean ratings were averaged by length and color. These

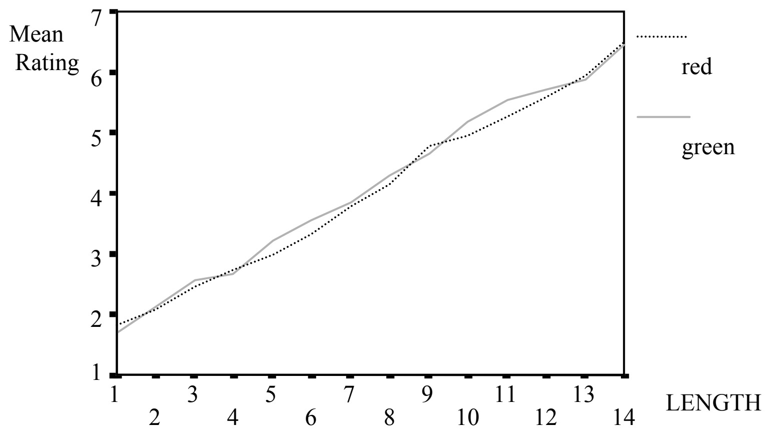

Fig.6.6. Results from the simulation. The solid lines represent the ratings of the positively skewed green lines, the dashed one – the ratings of the negatively skewed red lines.

aggregated ratings were tested with Repeated Measurement Analysis and the result was that the color significantly influenced judgments ($F(1,13)=3.990$, $p<0.05$). In other words, a green line with the same magnitude as a red line obtained a higher rating.

The difference, however, was very small. The mean rating of the green lines was **4.101**, whereas for the red ones - **4.025**. The difference was **0.076**. Moreover, the difference was not significant when only 15 runs of the program were performed. Therefore, this simulation differs from the others by the number of the runs of the program – 57 instead of 15 (60 runs were planed but 3 of them failed).

This prediction of the model was then tested with a psychological experiment performed by Kokinov, Hristova & Petkov (2004). They formed a set of lines whose lengths were uniformly distributed. The whole set, however, was subdivided into two subsets according to an irrelevant for the judgment task dimension – the color of the lines. Half of the lines were green and they were positively skewed according to their length, i.e. the short green lines dominate the long green ones. Conversely, the other halves of the lines were red and negatively skewed. The participants had to judge the lengths of the whole set randomly presented. The result was that the same length was judged with higher rating when it was in green than when it was in red. The difference in the ratings was 0.046 and turned out to be significant $p=0.026$. Thus, the psychological experiment confirmed the prediction of the model that the irrelevant dimension (color of the line) also takes part in judgment of length. In addition, the psychological experiment also confirmed the small size of the effect.

Other experiments also studied the role of the irrelevant characteristics in judgment and demonstrated similar results.

For example, Marks (1988) reported "differential context effects". He formed a set of sounds to be judged according to their loudness. Half of the sounds were 500Hz tones of a relatively low sound pressure level; the other halves were 2500Hz tones of relatively high sound pressure. The 70dB tones from the first half were judged to be as loud as the 73dB tones of the second half. The experiment was counterbalanced - in another group the 500Hz tones were of relatively high sound pressure, the 2500Hz tones, of low sound pressure. Now the 70dB tones from the first group were judged to be as loud as the 57dB tones from the second one.

In similarly designed studies using horizontally and vertically oriented lines, Arieh & Marks (2002), Armstrong & Marks (1997) tried to find out at what level of the cognitive processes this effect appeared and argued that this is was a low-level retinal effect. The JUDGEMAP model is not currently capable of explaining perceptual-driven effects. However, the mechanisms that underlie it propose a natural explanation[1] of the role of the irrelevant dimension in higher level cognitive processing.

Thus, in a series of four simulations the influence of the stimulus distribution was tested. The model is sensitive to the magnitude of the stimuli, to the previously given ratings and to the range and frequency of the stimulus set. The model uses all available properties of the stimuli - both relevant and irrelevant ones to produce the judgment.

The next step is to explore the behavior of JUDGEMAP when the scale varies.

[1] Namely, the role of spreading activation in formation of the comparison set and as a result producing a contrast effect with the dominating set of green lines (when a green line is judged).

6.3 The Role of the Scale

Wedell, Parducci & Geiselman (1987) demonstrated experimentally that the role of the skew of the stimulus set (frequency pressure) decreases when the number of the available ratings increases. In order to test whether JUDGEMAP can replicate this result, a uniform set and a skewed one were judged on a 3-point and on a 100-point scale. The used sets were the same as the two used in the previous simulations ('uniform distribution' and 'skewed set').

6.3.1 Judgment on a Three-Point Scale

Stimuli and procedure.

In this subsection two additional simulations are reported. The same set of stimuli was used – 112 lines separated into 8 length groups with 8 lines in each group but now the scale was a three-point one.

First, the whole set of 112 lines was judged on a three-point scale 15 times, presented randomly. Second, a positively skewed subset of 56 lines was judged 15 times on a three-point scale and again following the same procedure.

Results and discussions.

The results from both simulations are shown on fig.6.7.

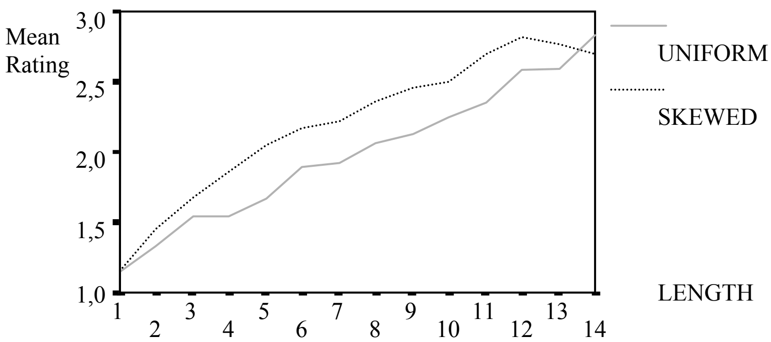

Fig.6.7. Mean ratings of the length of line lines in two conditions – uniformly distributed set (solid line), and positively skewed set (dashed line). The judgments were on a three-point scale.

The robust frequency effect was observed again – the distribution of the differences between the mean ratings for each length was significantly different from zero when analyzed with a t-test ($t(13)=5.630$, the mean difference was **0.216**, $p<0.001$).

In order to obtain more information, the influence of the skew when the judgment was on a three-point scale and when it was on a 7-point scale (simulation 6.2.3 - 'skewed set') were compared. The methodology proposed by Wedell, Parducci & Geiselman (1987) was used to perform this comparison. For each simulation the data were averaged by length. Then the following formula was used:

$\mathbf{w} = (\mathbf{C}_{i+} - \mathbf{C}_{i-}) / (\mathbf{F}_{i+} - \mathbf{F}_{i-})$, where

w is the weight of the frequency value; $\mathbf{C}_{i+}$ and $\mathbf{C}_{i-}$ are the mean ratings given by the model to each length when positively and negatively sets respectively were judged[1]; $\mathbf{F_{i+}}$ and $\mathbf{F_{i-}}$ are the frequency values[2] of the positively and negatively sets respectively.

Thus, the received weight of the frequency value, averaged by length was **0.79** (std. dev. 0.37) when judging on a three-point scale, and **0.53** (std. dev. 0.24) when judging on a 7-point scale. The difference was significant, analyzed with a paired t-test ($t(12)=2.18$, the mean difference was 0.26, $p=0.05$).

Why JUDGEMAP was capable to reproduce this result would be discussed later, after testing the behavior of the model when judging on a 100-point scale.

6.3.2 Judgment on a Hundred-Point Scale

Stimuli and procedure.

The same two simulations were also performed using a 100-point scale. First, the whole set of 112 lines, randomly presented, was judged on a 100-point scale 15 times. Second, a positively skewed subset of 56 lines was judged 15 times on a 100-point scale following the same procedure.

Results and discussions.

The averaged weight of the frequency value was **0.17** (std. dev. 0.58), calculated using the same methodology (Wedell. Parducci & Geiselman, 1987). As mentioned above for a three-point scale the respective weight was **0.79**, for a 7-point scale – **0.53**. All differences between these three weights were significant, analyzed with a paired t-test – respectively ($t(12)=2.92$, mean difference was 0.62, $p<0.02$) for a 3-point and a 100-point scales, and ($t(12)=2.91$, mean difference was 0.36, $p<0.02$) for a 7-point and a 100-point scales.

Thus, the conclusion is that the influence of the skew on JUDGEMAP's ratings increases when the scale range decreases, as demonstrated experimentally by Wedell, Parducci & Lane (1990).

The mechanism responsible for this result again is the pressure for one-to-one mapping. However, when judging on a three-point scale the common inhibitory input for each hypothesis is higher in comparison with the situation when judging on a 100-point scale. The reason for this is simple – when judging a certain stimulus on a three-point scale, the system can create only three alternative hypotheses. However, usually all three available ratings have already been used many times by other instances that are in the WM. Consequently, when judging on a three-point scale a greater number of inhibitory links to each of the hypotheses would be created in comparison with the judgment on a 100-point scale. Thus, the relative weight of the frequency principle decreases when the number of the available categories increases.

The next group of simulations also uses a 7-point scale. Their purpose is to test the behavior of the model when the so-called 'anchors' are presented. The anchor is a stimulus that is presented together with its rating. It serves as a reference point during judgment. The influences of extreme anchors, of anchors in the limits, and in the middle of the stimulus set were explored.

---

[1] It would be redundant to simulate separately positively and negatively skewed sets. The values for the negatively skewed set were calculated by subtracting the difference between the skewed and the uniform sets from the data from the uniform set.

[2] Frequency values depend only on the stimulus distribution, not on the responses. Frequency values were calculated with the counting algorithm, used by Parducci & Perrett (1971).

6.4 The Role of the Single Contextual Stimuli

Suppose that a certain stimulus connected with a certain rating via a winner hypothesis stays attached to the INPUT during the judgment of all set of stimuli. On the one hand, the activation of the respective rating label would stay permanently higher (because it receives activation from INPUT) and hence would produce assimilative pressure. Because of this assimilative pressure, the first several stimuli would be judged with ratings close to the anchor. On the other hand, however, this concentration around the respective anchor would increase the contrastive tendency because of the pressure for one-to-one mapping. Thus, from a long-term perspective, it is possible for the contrast effect to gain superiority or at least to weaken the assimilation.

Several simulations explored the behavior of the model when anchors were attached to the INPUT node.

6.4.1 Anchor outside of the Stimuli's Range

Stimuli and procedure.

The stimuli, and the procedure repeated those of the first simulation – 'uniform distribution' (112 lines, distributed in 14 groups of line's lengths, with 8 lines in each group, judged on 7-point scale) with only one exception – one additional line, together with its rating was attached at all times to the INPUT node. Thus, this additional stimulus was interpreted as anchor.

In this simulation, the anchor line had length 1600, i.e. it was outside the range of the judged stimuli (this range was between 100 and 1400). A winner hypothesis between this anchor and the rating 7 was created manually. The stimulus, together with its rating stayed attached to INPUT all the time. The 112 lines were judged 15 times, each time in a different, randomly selected, order.

Results and Discussions.

The dependency of the ratings from the magnitudes is represented on fig.6.8, together with the respective dependency from the simulation 'Uniform distribution' (the same stimuli without anchor) for comparison.

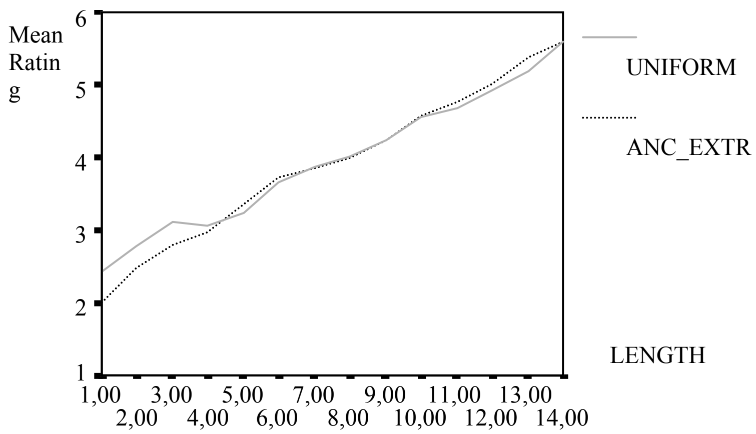

Fig.6.8. Main ratings on line length in two conditions – without anchors (solid line), and with anchor, whose magnitude was larger than the upper boundary of the range of the stimuli's set (dashed line).

There was no significant difference between the mean ratings in the two conditions – with and without anchor. This may look contradictory to the range effect, which was illustrated in the simulation 'restricted range'. However, in the current simulation the range restriction was too small and hence the size of the range effect disappeared.

However, some tendencies should be noted. It is shown in fig. 6.8. that the short lines were underestimated, i.e. a contrast effect, according to the anchor, was observed. The difference between the mean ratings of the first four length sizes was significant ($t(3)=-4.004$, the mean of the differences was **–0.283**, $p<0.029$). This contrast effect, decreased when the length increased.

This was due to the competition between the two opposite forces. On the one hand, the permanently higher activation level of the anchor 'rating 7' produced assimilation, i.e. overestimation of the ratings (this tendency was stronger for the long lines). On the other hand, from a long-term perspective, the same assimilative tendency increased the strength of the pressure for one-to-one mapping and hence produced contrast effect.

In the next simulation, JUDGEMAP judged line lengths on a 7-point scale when an anchor on the boundary of the stimulus set was attached to the INPUT node.

6.4.2 Anchor at the Boundary of the Stimuli's Range

The experimental data obtained in psychological studies turned out to be controversial as well, when the anchor was exactly on the boundary of the stimulus set. Sherif, Taub, Hovland (1958) and Parducci & Marshall (1962) received assimilation when they asked people to judge weights. Sarris & Parducci (1978) received contrast when they asked people to judge the size of squares.

Stimuli and procedure.

Everything remained the same with the exception that the anchor was exactly on the *boundary* of the stimuli's range. The 113^{th} line had length 1400, and was connected to the rating 7 with a winner hypothesis. This 113^{th} line stayed attached to the INPUT node all the time.

Results and Discussions.

The dependency of the ratings from the magnitudes is represented in fig.6.9,

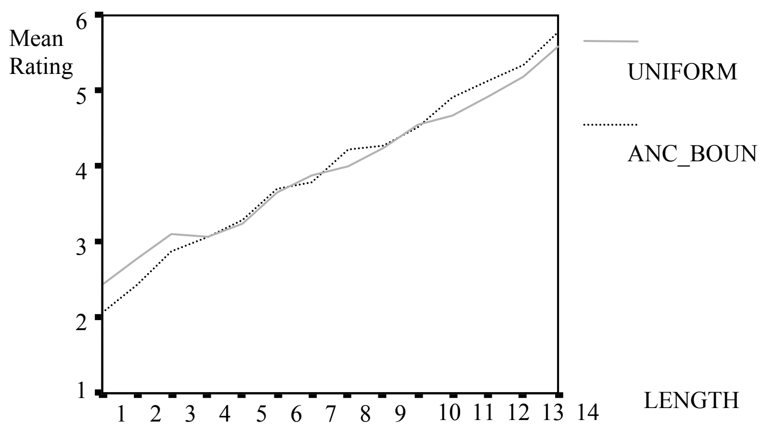

Fig.6.9. Plot of the main ratings in two conditions – without anchors (solid line), and with anchor on the upper boundary of the range of the stimulus set (dashed line).

together with the respective dependency from the simulation 'uniform distribution' (the same stimuli without anchor) for comparison.

Again, there was no significant difference between the mean ratings in both conditions. However, a contrast effect appeared again when the short lines were judged, but now the long lines were overestimated, i.e., the result for these long lines could be interpreted as assimilation to the anchor. The contrast when judging the first four lengths was almost significant ($t(3)=-2.760$, the mean difference was **–0.235**, $p<0.07$), the assimilation when judging the last four lengths was significant ($t(3)=11.399$, the mean difference was **0.197**, $p<0.01$).

This result illustrated the combined contradictory influences of the spreading activation mechanism and the comparison-based construction of hypotheses. Looking at overall length's range, neither contrast, nor assimilation could be concluded. Maybe this is the reason why the psychological experiments demonstrated ambiguous results – small shifts in both directions was observed, depending on the type of stimuli and the procedure.

However, the results from the simulation 'boundary anchor' could be interpreted in a different way. It could be speculated that the judgments were *improved*, i.e., the ratings of short lines became lower when an anchor was presented and vice versa – the ratings of long lines become higher when accompanied by an anchor. This was tested in the following way: For each of the 15 runs of the program, the standardized regression coefficient of the rating depending on the length was calculated performing a linear regression with a single variable entering as predictor – the line length. The same procedure was repeated with the data from the simulation 'uniform distribution'. Finally, the 15 regression coefficients from the simulation 'anchor at the boundary' were compared with the 15 regression coefficients from the simulation 'uniform distribution'. The mean of the standardized regression coefficients for the 'anchor at the boundary' condition was **0.82**, std.dev. 0.03; for the 'uniform distribution condition – respectively **0.77**, std.dev. 0.06. The difference was significant ($F(1,28)=10.06$, $p<0.01$).

The interpretation that anchors 'improve' judgment makes sense – it shouldn't come as a surprise that the presentation of anchors helps judgments. In terms of JUDGEMAP, the higher the number of 'correct' hypotheses that stay in the WM, the higher the probability of the system finding out the current 'correct' answer. However, in order to test this new hypothesis, namely that the anchors improve judgments, one simulation in which the anchor was in the middle of the stimulus range was made. In addition, a psychological experiment was performed, which tested whether the anchors improve judgments.

6.4.3 Anchor in the Middle of the Stimuli's Range

Stimuli and procedure.

Everything remained the same with the exception that the anchor was exactly in the *middle* of the stimuli's range. The 113^{th} line had length 800, and was connected to the rating 4 with a winner hypothesis.

Results and Discussions.

The dependency of the ratings from the magnitudes is represented on fig.6.10, together with the respective dependency from the simulation 'uniform distribution' (the same stimuli without anchor) for comparison.

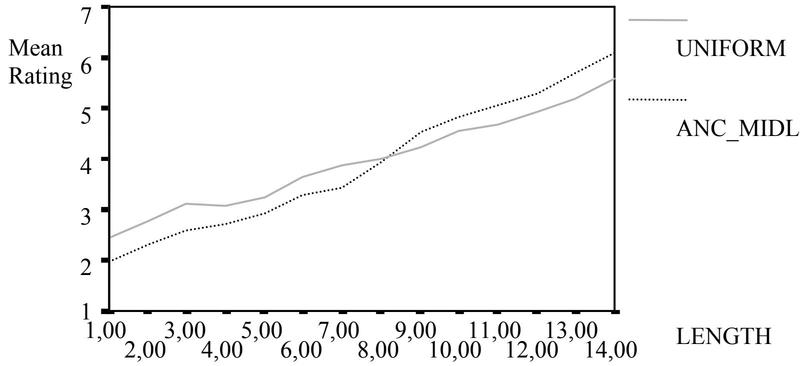

Fig.6.10. Plot of the mean ratings in two conditions – without anchors (solid line), and with anchor in the middle of the range of the stimulus set (dashed line).

A clear contrast effect was observed. All short lines were underestimated, whereas all long lines were overestimated. The difference between the mean ratings with and without anchor was positive for the lengths between 8 and 14, and negative for lengths between 1 and 7. Thus, the total shift away from the anchor was significant ($t(13)=8.94$, the mean difference was **0.37**, $p<0.01$).

The standardized regression coefficients were compared with the respective ones from the simulation 'uniform distribution', following the procedure from the previous simulation. The mean standardized regression coefficients for the 'anchor in the middle' condition were **0.90**, std.dev. 0.02; and for the 'uniform distribution condition – respectively **0.77**, std.dev. 0.06. The difference was significant ($F(1,28)=69.05$, $p<0.01$).

This result, together with the results from the previous simulation, is in favor of the hypothesis that presentation of anchors improves judgments. This hypothesis is intuitively clear but, nevertheless, it is necessary to test it with a psychological experiment.

6.4.4 Psychological Experiment 1 – Anchoring

Stimulus material

A set of 56 lines segment was used. They were distributed into 14 length groups, with 4 lines in each group. The shortest lines were 7 pixels, the longest ones – 410 pixels, with increment 31 pixels. Thus, the longest lines were almost as long as half of the screen's width. The lines were projected in red color on gray background on 17" PC monitor.

Procedure

The experiment consisted of three parts, the order of which varied randomly.

The first part (named control part) consisted of 56 trials – participants had to judge all 56 lines on a 7-point scale sequentially, in a random order.

The second part (named middle-anchor part) again consisted of 56 trials but at each trial two lines were projected. The line on the right side of the screen was from length group 8, i.e. the middle line and was marked with the number 4. Participants were instructed to judge the line on the left side on a 7-point scale, having in mind that the right line is a prototype for the rating 4. All 56 lines were presented on the left side in a random order.

The third part (named boundary-anchor part) was exactly like the second one, with the only difference that the line on the right side was the longest one (length group 14) and was marked with the number 7.

For each subject, the order of the three parts varied randomly, in order to distribute the effect of learning.

Participants

6 voluntary students from New Bulgarian University participated in the experiment.

Results and Discussions

The overall distribution of the ratings is illustrated on fig.6.11.

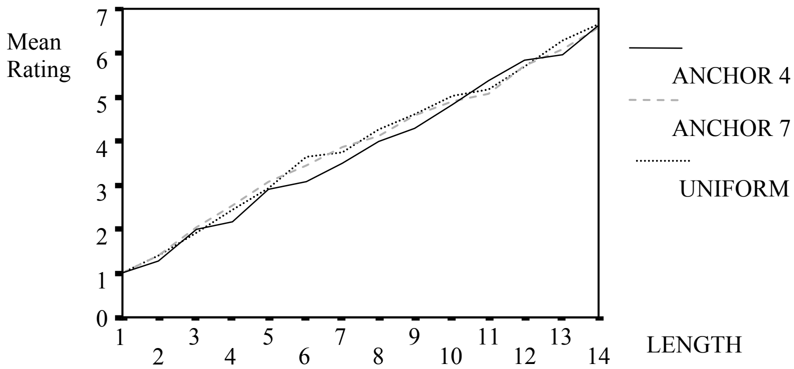

Fig. 6.11. The results from Psychological Experiment I – Anchoring. The mean ratings for the control condition (without anchors) are presented with a dashed black curve, the mean ratings from the condition 'anchor 4' (see text for details) are presented with a solid black curve, and the mean ratings for the condition 'anchor 7' – with a dashed gray curve.

The slopes of the three curves were of special interest. Thus, the procedure from the previous two simulations was repeated. The standardized regression coefficients for the dependency of judgments on lengths were calculated for each subject and each part of the experiment. There were three such regression coefficients for each subject – one for each part of the experiment, named respectively control, middle-anchor, and boundary anchor coefficients. The mean regression coefficients were respectively **0.95**, std.dev 0.014 for the control condition, **0.98**, std.dev 0.009 for the middle-anchor condition, and **0.97**, std.dev 0.007 for the boundary-anchor condition. Two pairs of coefficients were compared with Repeated Measurement Analysis. On the one hand, the coefficients from the middle-anchor condition were significantly different from those of the control condition ($F(1,5)=8.61$, $p<0.033$). On the other hand, the coefficients from the boundary-anchor condition were also significantly different from those of the control condition ($F(1,5)=10.44$, $p<0.024$).

The results allow the conclusion that the regression coefficients increase when an anchor is presented compared to judgments without anchors. This means that presentation of anchors improves judgments and thus the predictions of simulations 'anchor on the boundary' and 'anchor in the middle' were confirmed.

However, the results from the experiment did not replicate quantitatively the results from the simulations. All empirical regression coefficients were higher than the corresponding ones from the simulations. Thus, people were more precise in their judgments than the simulation. One possible reason could be the fact that people have much more previous experience. Even without being able to predict the exact quantities in the empirical data, JUDGEMAP predicted the qualitative tendencies, e.g. the tendency to improve judgments when anchors are presented.

In the next simulation, again two stimuli were given simultaneously to the model. However, neither of them was anchored. Instead, both of the stimuli were marked as target ones and thus the task of the model was to judge them both.

6.4.5 Simultaneously Judgment of Two Stimuli

Stimuli and procedure.

The same set of 112 lines was judged on 7-point scale. The difference was that the stimuli were judged in pairs. The whole set was randomly separated into 56 pairs, and the lines in each pair were given simultaneously as target to the system. The model gave its responses at unspecified times – when it was ready with one of the lines, it gave its rating, and continued with the other one. The whole procedure was repeated 15 times, each times the pairs and their order was randomly selected.

Results and Discussions.

The correlation between the lengths and the ratings was **0.878** ($p<0.001$), e.g. the model was able to deal with such a task: to judge simultaneously more than one stimulus.

In order to obtain more information, however, the influence of one of the stimuli in each pair on the judgment of the other stimulus was tested. The mean ratings for each length were calculated, and then from each single judgment the correspondent mean rating was subtracted, thus the deviation from the mean rating was calculated. The dependency of the received deviation from the second stimulus's length in the pair is presented in fig.6.12.

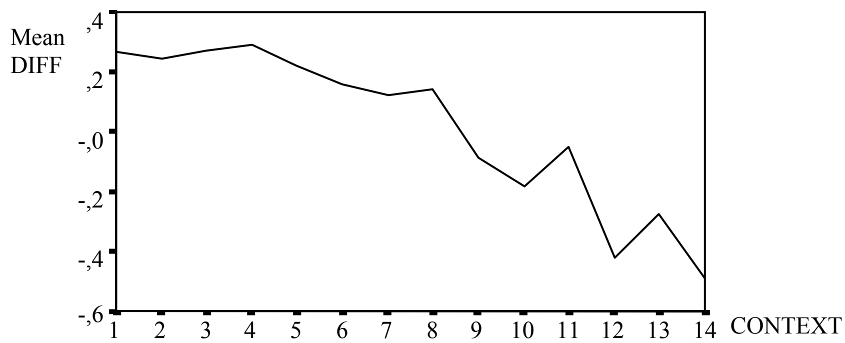

Fig.6.12. Mean shift of the ratings, depending on the magnitude of judged simultaneously stimuli.

It can be seen from fig.6.12. that when the contextual stimulus was short, the ratings were overestimated, and vice versa – when the second line in the pair was long, the ratings were underestimated. This tendency was calculated numerically – the correlation between the calculated difference and the other line in each pair was negative, namely **-0.277**, $p<0.01$.

This result could be interpreted as unambiguous contrast effect with respect to the second stimulus in the pair. The source of this contrast in JUDGEMAP was the fact that in all cases a comparison between the two stimuli in each pair was performed (the two target stimuli were the most active ones).

Because the experimental results from experiments with similar design are contradictory – some obtain contrast, others - assimilation, a psychological experiment was conducted with the same stimulus material as in the simulation (we did not find psychological experiments with lines on simultaneous presentation of two stimuli. Anderson has obtained assimilation but he used human faces as stimuli and it is possible that other factors related to the instruction played a role in obtaining this result).

6.4.6 Psychological Experiment 2 – Simultaneous Judgments

Stimulus material

A set of 112 line's segment was used. They were distributed into 14 length groups, with 8 lines in each group. The shortest lines were 7 pixels, the longest ones – 410 pixels, with increment 31 pixels. Thus, the longest lines are almost as long as the half of the screen's width. The lines were projected with red color on gray background on 17" monitor.

Procedure

The participants have to judge two lines on each trial – the first one was on the left side of the monitor, the second one – on the right side. The instruction was to judge the lengths of both lines on 7-point scale, and to report their ratings in any order they chose – immediately after they were ready with one of the lines, they had to report its rating.

Participants

6 students from New Bulgarian University participated in the experiment voluntarily, without payment.

Results and Discussions

First, the mean ratings for each participant and for each length were calculated. Then, from each single answer the mean ratings for the corresponding participant and length was subtracted. Then the correlation between the received difference and the length of the other line in the pair was calculated. This correlation was negative, namely **–0.133**, $p<0.01$, which gave us reason to conclude that contrast effect appeared. This result was consistent with the prediction of JUDGEMAP from the simulation 'simultaneous judgment'. However, because of the low value of the correlation coefficient, an additional experiment should be performed in order to test whether the negative correlation is robust.

The behavior of JUDGEMAP was tested when the distribution of the stimulus set varied, when the scale type varied, and when different contextual stimuli were added to the representation of the stimuli and the task. In all the cases the stimuli were simple – lines with different lengths (and possibly colors). The next step was to test the model with more complicated stimuli – when more than one dimension should be taken into account in the judgment task.

6.5 Judgment of Two-Dimensional Stimuli

6.5.1 Correlated Magnitudes

The items, given for judgment to JUDGEMAP in this simulation represent students. Each student had two exam scores and on the basis of these two scores the model had to judge the overall performance of the students.

Each student was represented with a head DUAL agent, connected to two its aspect – respectively its two exam scores. The correspondence relation attached to the GOAL represent the information that better students had to receive higher ratings. The concept 'better student' had two subclasses – comparison-relations respectively for higher score at the first exam, and higher score at the second exam. The scale, on which the students were judged, was again from 1 to 7.

The scores from both exams varied between 10 and 140. The two scores for each student correlated, i.e. if a student had a lower score at the first exam, he/she also had a lower score at the second one, and vice versa.

Totally 112 students, randomly ordered, were judged sequentially. The procedure was repeated 15 times, each time with a different random order of the stimuli.

Result and Discussions

The correlation between the score from the first exam and the given ratings was **0.780**, i.e. the model was able to deal with more than one dimension.

The distribution of the ratings again was non-uniform – their standard deviation was **1.39** (min 1.26, max 1.51, std. dev. 0.07 on the aggregated for the runs of the program data) with expected for uniform distribution 2.03.

Sequential assimilation effect was also observed. The standardized regression coefficient for the previous ratings was **0.44** (sequential assimilation), for the previous lengths was **-0.42** (contrast from the previous magnitude).

Thus, the simulation demonstrated that JUDGEMAP is able to make judgments of complex stimuli, using more than one relevant dimension. The next simulation tested whether the frequency effect would appear when a skewed set of two-dimensional stimuli was judged.

6.5.2 Skewed Set of Two-Dimensional Stimuli

Design

In this simulation, part of the stimuli, designed for the previous simulation, was used, forming a positively skewed subset. Eight students with scores 10 at both exams, eight students with score 20, seven ones with scores 30 and 40, etc., one student with score 130, and one with score 140 (a total of 56 students), were judged in random order. The procedure was repeated 15 times, varying the order of the presentation.

Results and Discussions

The mean ratings for each score were compared with those from the previous simulation. The results are shown in fig.6.13.

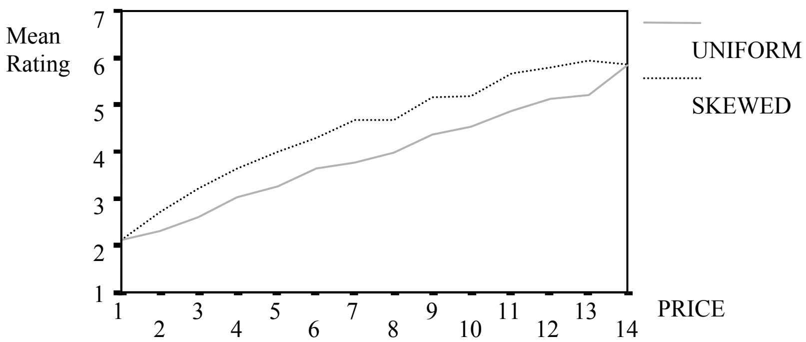

Fig.6.13. Mean ratings of the students' performance in two conditions – uniformly distributed set (solid line), and positively skewed set (dashed line)

The difference between the mean ratings in the two simulations for each score was estimated with t-test (t(13)=8.21, the mean difference was **0.59**, p<0.01).

The overall significant overestimation of the ratings pointed to contrast effect from the skew of the stimuli's distribution, exactly as is predicted by the Frequency Principle (Parducci & Perret, 1971).

6.5.3 Negatively Correlated Magnitudes

Often in everyday life people face stimuli that have two negatively correlated dimensions. For example, in the so-called trade-off judgments, one needs to estimate products that have a kind of benefit and a kind of price. In addition, lower benefits correspond to lower prices, and vice versa.

The next simulation tests the behavior of JUDGEMAP in a similar situation.

Design

In this simulation, the judged stimuli were different cheeses that vary in their price and quality. The correlation between the two dimensions, however, was negative. The cheeses with higher quality were more expensive, those with lower quality – cheaper. The magnitude of each quality was represented with a whole number between 100 and 1400. There were 14 different quality magnitudes – from 100 to 1400 with increment 100. The cheeses with quality 100 had price 1, these with quality 200 had price 2, etc., the cheeses with quality 1400 had price 14. The task of the model was to rate each cheese on a 7-point scale, with constrains that higher qualities should correspond to higher ratings, and that higher prices should correspond to lower ratings. Thus, a natural trade-off judgment situation was simulated.

Again, the whole judged set consisted of 112 stimuli, with 8 in each group. Again, the procedure was repeated 15 times, varying randomly the order of the presentation

Result and Discussions.

There were no correlations neither between the ratings and prices, nor between the ratings and qualities. In other words, the task was too ambiguous – the price and the quality of each cheese justified different hypotheses, and thus the competition in the constraint satisfaction network was extremely cruel. Each hypothesis was inhibited from too many competitors, and, consequently, the activation received from its elements influences judgment with a relatively lower degree. Thus, the sequential assimilation disappeared, and was even inversed. The standardized regression coefficient for the previous ratings was **–0.17** (p<0.01), for the previous lengths it was not significant, i.e. a sequential contrast effect was observed.

In order to check this prediction of the model, additional experiments were designed. One problem, however, arose – people are very confused when they have to judge such type of stimuli on a numerical scale. Some people use only one dimension, neglecting totally the other one. Others give to all stimuli one and the same rating (usually after long thinking). Thus, the data were not valid. However, additional experiments that could test this prediction of the model should be designed.

6.6 Choice

The last simulation tests the capability of JUDGEMAP-2 to make choices, using the same mechanisms as in the judgment task. An experiment by Shafir, Siminson & Tversky (1993) served for a template for this simulation. They asked people to choose between two gambles on the basis of their probability to win, and the amount of the possible profit. Always one of the gambles was with higher profit but with lower probability. With series of pretests, they found out two psychologically equal gambles, i.e. half of the people chose the first one, and the other half – the second one. After that, the authors asked another group of people to choose between three alternative gambles – the same two, together with a third one, which was dominated by the first one according to both dimensions, and was dominated by the second one according to only one of the dimensions. The result was that now more people choose the first gamble than the second one.

An analogical situation was tested with JUDGEMAP.

Stimuli and procedure.

Two gambles, named **A** and **B** were designed. The probability of **A** was randomly chosen to be between 0.51 and 0.99, and the profit was randomly chosen to be between 1 and 49. Then the probability and the profit of the gamble **B** were chosen to be respectively between 0.1 and 0.49, and between 51 and 99 (see fig.6.14).

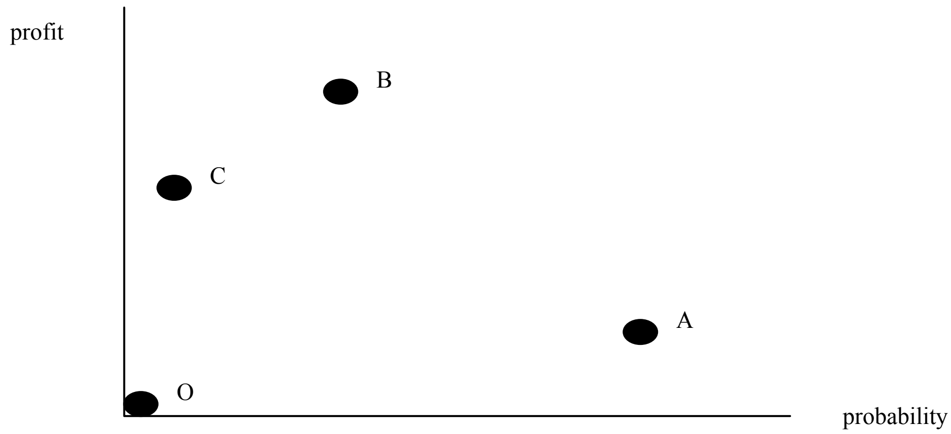

Fig.6.14. Relative positions of the stimuli, used in simulation for testing choice making.

One more stimulus, named **O** was added. Its probability and profit were always zero. Without it, the system would be unaware of the range of the given magnitudes. Recall that the model captures the absolute range with comparisons of second order.

The procedure of creating stimuli was repeated 100 times, varying randomly the exact value of the probability and the profit of each gamble, but keeping their relative position. Thus, a hundred different knowledge bases were created and the model was run on each one of them. The task of the model was to judge simultaneously the three gambles on 2-point scale. In addition, the :t-driver tag was given not to the stimuli, but to the rating 2. Thus, only the highest rating (rating 2) could promote winner hypotheses. The winner hypothesis for rating 2 was interpreted as an answer of the system.

After the runs over all 100 knowledge bases (KB), a fourth gamble **C** was added to each KB. Its values for probability and profit were randomly chosen, satisfying the constraints that its probability should always be lower than that of gamble **B**, and its profit should always be between the profits of **A** and **B**. The task of the system was to judge simultaneously the four stimuli, following the same procedure for choice. Again 100 runs under each KB were performed.

All procedures were repeated 10 times in order for the results to be tested statistically.

Results.

When choosing between **O**, **A**, and **B**, the system picked up **A 49.30** times on the average (out of 100 runs), and **B** – **50.70** times. In the second session, when the choice was between four alternatives, JUDGEMAP preferred **A 36.80** times, **B** – **60.30** times, and **C** – **2.90** times. The existence of the alternative **C** turned out to be a significant factor for the number of choices of **A** and **B**, tested with ANOVA analysis (respectively, $F(1,18)=44.828$, $p<0.01$ for **A**; and $F(1,18)=34.525$, $p<0.01$ for **B**).

Thus, JUDGEMAP replicated the results obtained by Shafir, E., Siminson, I., Tversky, A. (1993), and demonstrated that the same mechanisms could be used for the tasks of judgment and choice.

6.7 Summary of the Simulations

Many simulations were performed with JUDGEMAP in order to test its behavior when judging different stimuli, in different contexts, and with different scales.

The results showed that the model is capable of judging one- or multi-dimensional stimuli, on various scales. The same mechanisms could also be used for making choices. JUDGEMAP reproduces sequential assimilation effect, Range and Frequency effects, the effect of non-uniform distribution of the ratings, even when the stimuli are uniformly distributed. It also reproduces various contrast effects when anchors are presented together with the judged stimuli.

The model made several predictions that were experimentally tested and confirmed. The not so obvious and even strange prediction that JUDGEMAP made was that people do not neglect the role of the irrelevant dimensions in judgment. Instead, the model produced small shifts in judgments, depending on these irrelevant properties, and this prediction was confirmed by a psychological experiment.

The model gives rise to the observed fact that the role of the Frequency principle increases when the number of the scale labels decreases.

It is able to judge more than one stimulus simultaneously. A simulation, performed with such design made a prediction that was tested experimentally and was confirmed.

Finally, JUDGEMAP is able to use the same mechanisms to make choice, and replicates the results, proposed and tested by the theory for Reason – Based Choice (Shafir, Siminson & Tversky, 1993).

6.8 JUDGEMAP and the Other Models of Judgment

Various descriptive theories of judgment are discussed in literature. JUDGEMAP shares some ideas with many of them. For example, it assumes that one of the basic mechanisms in judgment is the creation of comparisons between the target stimulus and similar ones, like the two-Path Model (Manis & Paskewitz, 1984) and the theory for Comparison-Based Judgments (Mussweiler 2003). JUDGEMAP is also similar to the Norm Theory (Kahneman & Miller, 1986) with respect to the fact that both assume that memory is constructive. From a connectionist point of view JUDGEMAP is a localist neural network and from this perspective, it is close to the Change of Meaning Approach (Wyer, 1974).

I believe that that some descriptive theories that focus on aspects that were not treated in this dissertation (like perception and categorization processes) could also be integrated with DUAL architecture and particularly with the JUDGEMAP model in the future. Examples of such theories are Perceptual Learning Theory (Goldstone, 1998), and Integration Theory (Anderson, 1971).

JUDGEMAP differs, however, from all descriptive theories in that it proposes concrete cognitive mechanisms and algorithms for judgment. Thus, the behavior of the model can be tested more precisely, and novel predictions can be proposed and tested experimentally.

JUDGEMAP can be compared with normative theories and with other computational models based on the results from their simulations.

Some theories from this group predict either always assimilation, or always contrast. For example, Range–Frequency Model (Parducci, 1965) predicts only contrast, whereas Adaptation Level Theory (Helson, 1964) – only assimilation. JUDGEMAP is able simultaneously to demonstrate both assimilation and contrast tendencies depending on the conditions of the simulations.

The ANCHOR model (Petrov & Anderson, 2000) is capable of demonstrating various effects, for example sequential assimilation, and non-uniform distribution of the judgments. It also models the learning processes and the task for categorization – two related capabilities, which JUDGEMAP cannot simulate in its current version. However, JUDGEMAP does not put any a priori limitations on the type of scale or on the complexity of the judged stimuli, as ANCHOR does. JUDGEMAP also elaborates the role of the irrelevant dimensions in the judgment process, and the possibility to use the same mechanisms in judgment and choice tasks.

Finally, in my opinion, the key innovation of JUDGEMAP is that it integrates the mechanisms that underlie judgment and choice on the one hand, and analogy-making on the other, in a single whole, using a common cognitive architecture and underlying mechanisms.

CHAPTER VII
Shortcomings and Future Work

7.1 Main Problems

The simulations, performed by JUDGEMAP, demonstrate many of the capabilities of the model.

However, one problem of the model is that it has no perception. Attaching different manually created stimuli on the INPUT node only mimics it. The magnitudes of the judged properties are predefined and filled as real numbers in special slots. This procedure, however, is not psychologically plausible. If the model has to be augmented with perceptual mechanisms for self-creation of the agent's micro-frames, its behavior would be more flexible than it currently is.

A second problem of JUDGEMAP is that the concepts are not dynamic enough. Their micro-frames are too frozen and do not reflect to a satisfactory extent the relevance of their features. One step in this direction was made by JUDGEMAP by using temporary links from the concept–agents to their most relevant instances. In order for the process of categorization to become more plausible, however, additional work in this direction is necessary.

Finally, the simulations, performed with JUDGEMAP do not highlight all its capabilities. Additional simulations for judging multi-dimensional stimuli are needed. In the fields of choice and decision-making only first steps were performed and the model has not been explored enough.

7.2 Perception

Goldstone (1995) demonstrated how people's perception influence judgments. He received both assimilation and contrast effects depending on the degree of similarity between the target and the contextual stimuli. Arieh & Marks (2002) found out that some contrast effects appear at the level of retina. Lockhead (1992) demonstrated that all cognitive levels of the judgment process produce sequential assimilation effect.

None of these experiments could be replicated by JUDGEMAP-2. Additional perceptual mechanisms should be integrated in the model in order to do so.

However, JUDGEMAP, together with DUAL architecture, exerts some constraints on these eventual mechanisms.

First, these mechanisms should be *integrated* with DUAL, and should be applicable to all models, based on DUAL.

Second, the perceptual mechanisms should not precede the other ones, but should be *overlapped* with them. One of the assumptions behind DUAL is that cognition is a continuous process. All mechanisms work in parallel and influence each other. Thus, the eventual perceptual mechanisms should influence the other ones, but the higher level processes should also influence perception.

Third, the *relevance* of the various objects, features, and relations should be reflected in perception. The more relevant items should be perceived faster and better, the less relevant ones – more slowly and only in broad strokes. However, the relevance may not be the only driving force for attention. Some items may be more *salient*. Moving objects, vivid colors, and other features also attract the attention, and this should be captured by the perceptual mechanisms.

Fourth, in order to be consistent with DUAL principles, perceptions should be emergent processes – without any central executors or controllers

First steps in this direction were made by Nestor (2004), who modeled representations, derived from visual stimuli in the TextWorld micro-domain. The model combines low-level parallel computations with DUAL's semantic memory. However, additional work to fully integrate the model with DUAL is needed.

7.3 Categorization (Recognition)

In DUAL architecture terms, to categorize an item means to fill its :inst-of slot of its respective agent with a reference to a concept–agent. In order for this process to become psychologically plausible, several constraints have to be satisfied.

First, recognition should be a c*ontext–sensitive* process. The same item can be categorized in different ways in different contexts. One possible starting point is to think about the categorization process as a result of *mapping*. Different concepts compete with each other to capture a new instance, and this competition reflects the current relevance of the concepts and the current relevance of the justifications for each possible categorization. An example of recognition is implemented in JUDGEMAP, using a different starting point. The comparison-relations actively seek their manifestations and create their temporary instances.

Second, the content of the concepts changes dynamically in response to the environment and the current goals. This is not implemented yet in DUAL architecture to the required extend. The weights of the links from the agents to their different aspects should change dynamically in response to the relevance of these aspects.

Third, eventual augmentation of JUDGEMAP with such mechanisms should give rise to the 'hallo' effect (Anderson, 1971), and to the expectation–driven assimilation (Manis & Paskewitz, 1984) in judgment task.

Goldstone (1988) proposes four possible mechanisms for integration of perceptions and categorization – intentional weighting, stimulus imprinting, differentiation, and unitization. A careful implementation of some of these principles into computational mechanisms could serve as a basis for modeling these processes.

7.4 Additional Simulations for Multi-Attribute Judgments and Choice

Even using only the currently implemented mechanisms, more extensive testing of JUDGEMAP can be performed.

Cooke et al (2004) demonstrated that under special circumstances when people make trade-off judgments, greater preference to the stimuli in the middle of the set's range could be observed (an inversed U-shape curve). New simulations with JUDGEMAP can be performed in order to test the behavior of the model when a task for trade-off judgments is presented.

The work on the topic of choice is just the beginning. JUDGEMAP is based on mechanisms that may capture many of the phenomena, observed in the field of choice, but much work has to be done in that direction.

Tversky & Kahneman (1974) propose several *heuristics*, on which people base their judgments and choices. *Representativeness* is an assessment of the correspondence between instances and categories; *availability* reflects how easily the system finds instances of a certain category; *anchoring and adjustment* are mechanisms for creating starting points for judgments and adjusting them to reach a final answer. Tversky (1972) and Shafir, E., Siminson, I., Tversky, A. (1993) propose the theory for Reason–Based Choice, which highlights the importance of the justifications when choosing a certain alternative. Kahneman & Tversky (1979) demonstrate the role of the context in judgment and choice.

The work of JUDGEMAP is based on comparisons between the target stimulus and memory traces, and this is in synchrony with the availability heuristic. The comparisons serve for creating justifications for alternative hypotheses and thus the model implements the ideas for reason–based choice and for anchoring and adjustment heuristics. Finally, one of the main principles in all DUAL-based models is context–sensitivity, thus giving rise to the role of prospect in decision–making.

However, all these phenomena should be carefully modelled and tested with many additional simulations.

CHAPTER VIII
Conclusions

8.1 Main Ideas of the JUDGEMAP Model

A model for judgment, JUDGEMAP, based on the cognitive architecture DUAL was presented.

It proposes a new approach – it treats judgment as a process of mapping between a dynamic comparison set and the set of available ratings. The model inherits the basic principles of DUAL, and many of the mechanisms of AMBR model for analogy–making. The results, reported by JUDGEMAP emerge from the common work of several overlapping processes.

The Working Memory is formed by the spreading activation. The activation level of each single unit represents its relevance to the current context. The overall pattern of activation changes continuously in response to the environment and the current goals.

Some of the relations actively seek and create their new instances due to a kind of recognition process. These relations represent comparisons between items.

Correspondence relations create justifications for new correspondences between stimuli and ratings. Their work reflects the constraints for homomorphism, defined in the judgment task.

The justification agents combine the work of several agents and create new hypotheses for correspondences or justify old ones.

At the end, the hypotheses are organized in a constraint satisfaction network, and the final result emerges from its relaxation.

8.2 Contributions of this Thesis

A model of judgment has been proposed and implemented. The model is based and integrated within the general cognitive architecture DUAL. Simulation experiments have been performed with the computer program.

JUDGEMAP integrates judgment with analogy-making, proposing that the same mechanisms may underlie both of them. In addition, it demonstrates that the same mechanisms could potentially be used for making choice.

JUDGEMAP replicates many of the results, observed in various experiments. It gives rise to sequential assimilation effect, to range and frequency principles, to the role of the irrelevant dimensions. The type of the stimuli and the scale does not limit the work of the model. It is capable of reproducing various contrastive effects with respect to some contextual elements. In addition, by changing the items assumed to have the driving role, the model demonstrates that it is able to make a choice between alternatives. It replicates some contextual effects, observed in people's choices.

The model made several predictions about new (yet unknown) psychological phenomena. Three of the model's predictions were tested with psychological experiment and confirmed – the influence of irrelevant features on the judgment outcome, contrast effects were observed when judging two line's lengths simultaneously, and that anchors improve the precision of judgment.

8.3 Open Issues and Future Research

JUDGEMAP is just at an intermediate stage in the long-term DUAL-research program.

First, many more simulations could be performed with the existing version of the model. The influences of the time delays between some contextual manipulations and judgments could be explored in more detail. Various simulations with priming could be tested. Much more complex stimuli could be given to the model in order for its behaviour to be tested.

It was demonstrated that the mechanisms of JUDGEMAP could be used for making choices. However, this is only the starting point for the essential work in this direction.

Testing of the model requires psychological experiments, too. For example, JUDGEMAP makes the prediction that the sequential effect disappears and even reverses in one special case of trade-off judgments. This prediction could be tested empirically.

A great challenge will be to add capabilities for perception and categorisation in DUAL. All models, based on the architecture, are subject to some constraints. In the least, these models should be integrated with JUDGEMAP.

ACKNOWLEGMENTS

I would like to express my special gratitude to my research supervisor Boicho Kokinov who has guided every aspect of my work.

I would like to express my very special thanks to my wife and collaborator Penka Hristova for her support and proposals during the whole project.

I would like to thank the other members of the DUAL research group – Moris Grinberg, Chavdar Naydenov, and Alexander Stoimenov for their continuous support and important discussions.

I would like to especially thank Armina Janjan and Marina Hristova for her valuable suggestions on this dissertation.

References:

Anderson, J., Lebiere, Ch. (1993) The Atomic Components of Thought. *Lawrence Erlbaum Associates, Hillsdale, NJ.*

Anderson, J. (2003) The Newell Test for a Theory of Cognition. *Behavioral and brain science, 26, pp. 587-637.*

Anderson, N. (1996) Component Ratings in Impression Formation. *Psycon.Sci, vol.6 (6), pp 279-280.*

Anderson, N. (1971) Integration theory and attitude change. *Psychological Review*, May, Vol.78 (3).

Arieh, Y., Marks, L. (2002) Context effects in visual length perception. Role of ocular, retinal, and spatial location. *Perception & Psychophysics, vol. 64 (3), pp. 478-492.*

Armstrong, L., Marks, L. (1997*)* Differential effects of stimulus context on perceived length: Implications for the horizontal – vertical illusion. *Perception & Psychophysics, vol. 59 (8), pp. 1200-1213.*

Cooke, A., Mellers, B. (1998) Multiattribute Judgment: Attribute Spacing Influences Single Attributes. *Journal of Experimental Psychology: Human Perception and Performance, vol. 24 (2), pp 496-504.*

Cooke, A., Janiszewski, C., da Cunha Jr., M., Nasco, S., de Wilde, E. (2004) Stimulus Context and Formation of Consumer Ideals. *Journal of Consumer Research, vol. 31.*

Goldstone, R. (1995) Effects of Categorization on Color Perception. *Psychological Science, vol. 6 (5), pp. 298-304.*

Goldstone, R. (1998) Perceptual Learning. *Annual Reviews Psychology, vol. 49, pp. 585 – 612.*

Jesteadt, W., Luce, R., Green, D. (1977). Sequential Effects in Judgment of Loudness. *Journal of Experimental Psychology: Perception and Performance, vol. 3 (1), pp. 92-104.*

Helson, H. (1964) Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior. *NY: Harper and Row.*

Herr, P., Sherman, St., Fazio, R. (1983) On the Consequence of Priming: Assimilation and Contrast Effects. *Journal of Experimental Social Psychology, vol. 19, pp 323-340.*

Hofstater, D. (2001*)* Analogy as the Core of Cognition. *In Dedre Gentner, Keith J. Holyoak, and Boicho N. Kokinov (eds.)* The Analogical Mind: Perspectives from Cognitive Science. *Cambridge, MA: The MIT Press/Bradford Book, pp. 499-538.*

Hsee, Ch. (1996) Elastic Justification: How Unjustifiable Factors Influence Judgments. *Organizational Behavior and Human Decision Processes, vol. 66 (1), pp 122-129.*

Kahneman, D. and A. Tversky (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica, vol. 47, pp 263-291.*

Kahneman, D., Tversky, A. (1984) Choices, Values and Frames. *American Psychologist, vol. 39, pp 341-350.*

Kahneman, D., Miller, D. (1986) Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review, vol. 93 (2), pp 136-153.*

Kenrick, D. T., and Gutierres S.E. (1980) Contrast Effects and Judgments of Physical Attractiveness: When beauty becomes a social problem, *Journal of Personality and Social Psychology, vol. 38, pp. 131-140.*

Kokinov, B. (1988) Associative Memory-Based Reasoning: How to Represent and Retrieve Cases. *In O'Shea, T.; Sgurev, V. (eds.)* Artificial Intelligence III: Methodology, Systems, Applications. *Amsterdam, Elsevier.*

Kokinov, B. (1994a) A Hybrid Model of Reasoning by Analogy. *In Holyoak, K., Barnden, J. (eds)* Advances in Connectionist and Mental Computation Theory, vol. 2: Analogical Connections. *Norwood, NY: Ablex, pp 247-318.*

Kokinov, B. (1994b) The DUAL Cognitive Architecture. A Hybrid Multi-Agent Approach. *In A. Cohn (ed.) Proceedings of the Eleventh European Conference of Artificial Intelligence. London: John Wiley & Sons, Ltd, pp. 203-207.*

Kokinov, B. (1994c) The Context-Sensitive Architecture DUAL. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: LEA.*

Kokinov, B. (1998). Analogy is like Cognition: Dynamic, Emergent, and Context-Sensitive. *In: Holyoak, K., Gentner, D., Kokinov, B. (eds.) – Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences.* Sofia: NBU Press.

Kokinov, B., Hristova, P., Petkov, G. (2004) Does Irrelevant Information Play a Role in Judgment? *In: Proceedings of the 26th Annual Conference of the Cognitive Science Society, 2004. Erlbaum, Hillsdale, NJ.*

Kokinov, B. & Petrov, A. (2001) Integration of Memory and Reasoning in Analogy-Making: The AMBR Model. *In: Gentner, D., Holyoak, K., Kokinov, B. (eds.) The Analogical Mind: Perspectives from Cognitive Science, Cambridge, MA: MIT Press.*

Kokinov, B., Zareva-Toncheva, N. (2001). Episode Blending as Result of Analogical Problem Solving. In: *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ.

Lockhead, G. (1992) Psychophysical scaling: Judgments of attributes or objects? *Behavioral & Brain Sciences, vol. 15, pp. 543-559.*

Manis, M., Paskewitz, J. (1984) Judging Psychopathology: Expectation and Contrast. *Journal of Experimental Psychology, vol. 20, pp 217-230.*

Manis, M., Nelson, Th., Sheder, J. (1988) Stereotypes and Social Judgment: Extremity, Assimilation and Contrast. *Journal of Personality and Social Psychology, vol. 55 (1), pp 28-36.*

Marks, L. (1988) Magnitude estimation and sensory matching. *Perception & Psychophysics, vol. 43, pp. 511-525.*

Medin, D., Goldstone, R., Markman, A. (1995) Comparison and Choice: Relations between Similarity Processes and Decision Processes. *Psychonomic Bulletin & Review, vol. 2 (1), pp. 1 – 19.*

Mellers, B., Birnbaum, M. (1982) Loci of the Contextual Effects of Judgment. *Journal of Experimental Psychology: Human Perception and Performance, vol. 8 (4), pp 582-601.*

Mellers, B., Birnbaum, M. (1983) Contextual Effects in Social Judgment. *Journal of Experimental Social Psychology, Academic Press, pp 157-171.*

Mellers, B., Cooke, A. (1994) Trade-offs Depend on Attribute Range. *Journal of Experimental Psychology: Human Perception and Performance, vol. 20 (5), pp 1055-1067.*

Mussweiler, T. (2003) Comparison processes in Social Judgment: Mechanisms and Consequences. *Psychological Review, vol. 110, pp. 472-489.*

Nestor, A. (2004) Active vision in the DUAL cognitive architecture. *Ph.D. thesis, New Bulgarian University.*

Nosofsky, R., Palmeri, Th. (1997) An Exemplar Based Random Walk Model of Speeded Categorization and Absolute Judgment. *Psychological Review, vol. 104 (2), pp. 266-300.*

Parducci, A., Marshall, L. (1962) Assimilation vs. Contrast in the Anchoring of Perceptual Judgments of Weight. *Journal of Experimental Psychology, vol. 63 (5), pp 426-437.*

Parducci, A. (1965) Category Judgment: A Range-Frequency Model. *Psychological Review, vol. 72 (6), pp 407-418.*

Parducci, A. (1968) The Relativism of Absolute Judgment. *Scientific American, vol. 219 (6), pp 84-90.*

Parducci, A., Perret, L. (1971) Category-Rating Scales: Effects of Relative Spacing and Frequency of Stimulus Values. *Journal of Experimental Psychology, vol. 89 (2), pp 427-452.*

Parducci, A. (1973) A Range-Frequency Approach to Sequential Effects in Category Rating. *In Kornblum, S. (ed.)* Attention and Performance IV. *NY: Academic Press, pp 303-311.*

Petrov, A. (1997) Extensions of DUAL and AMBR. *M.Sc. thesis, New Bulgarian University.*

Petrov, A. (1998) A Dynamic Emergent Computational Model of Analogy–Making Based on Decentralized Representation. *Ph.D. thesis, New Bulgarian University.*

Petrov, A., Anderson, J. (2000) ANCHOR: A Memory-Based Model of Category Rating. *In L. R. Gleitman & A. K. Joshi (Eds.) Proceedings of the Twenty-Second Annual Conference the Cognitive Science Society. Hillsdale, NJ: LEA, pp. 369-374.*

Petrov, A., Anderson, J. (2005) The Dynamics of Scaling: A Memory-Based Anchor Model of Category Rating and Absolute Identification. *Psychological Review, vol. 112* (2), pp. 383-416.

Petrov, A., Kokinov, B. (1998). Mapping and Access in Analogy-Making: Independent or Interactive? A Simulation Experiment with AMBR. In: Holyoak, K., Gentner, D., Kokinov, B. (eds.) – Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences. Sofia: NBU Press.

Restle, F. (1978) Assimilation Predicted by Adaptation – Level Theory with Variable Weights. *In Castelan, N., Restle, F. (eds)* Cognitive Theory. *Hillsdole, NY: Erlbaum, vol. 3, pp 75-81.*

Sarris, V., Parducci, A. (1978) Multiple Anchoring of Category Rating Scales. *Perception and Psychophysics, vol.24 (1), pp. 35-39.*

Schwarz, N., and H. Bless, (1991) Constructing reality and its alternatives: assimilation and contrast effects in social judgment. *In L.L. Martin and A. Tesser, (eds.)* The construction of social judgment. *Hillsdale, NJ: Erlbaum, pp. 217-245.*

Shafir, E., Siminson, I., Tversky, A. (1993) Reason – Based Choice. *Cognition, vol. 49, pp. 11-36.*

Sherif, M., Taub, D., Hovland, C. (1958) Assimilation and Contrast Effects of Anchoring Stimuli on Judgment. *Journal of Experimental Psychology, vol. 55 (2), pp 150-155.*

Srull, Th. and Wyer, J. (1979) The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, vol. 37, pp. 1660-1672.*

Srull, Th., Wyer, J. (1980) Category Accessibility and Social Perception: Some Implications for the Study of Person Memory and Interpersonal Judgment. *Journal of Personality and Social Psychology, vol. 38 (6), pp 841-856.*
Stapel, D., Kooman, W., van der Pligt, J. (1996) The Referents of Triat influences: The Impact of Trait Concepts Versus Actor – Trait Links on Subsequent Judgment. *Journal of Personality and Social Psychology, vol. 70 (3), pp 437-450.*
Stapel, D., Kooman, W. (2000) How Far Do We Go Beyond the Information Given? The Impact of Knowledge Activation on Interpretation and Influence. *Journal of Personality and Social Psychology, vol. 78 (1), pp 19-37.*
Strack, F., Mussweiler, Th. (1997) Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility. *Journal of Personality and Social Psychology, vol. 73 (3), pp. 437-446.*
Strack, F., N. Schwarz, and E. Gschneidinger (1985) Happiness and reminiscing: The role of time perspective, mood, and mode of thinking. *Journal of Personality and Social Psychology, vol. 49, pp.1460-1469.*
Strack, F. (1992) The Different Routes to Social Judgments: Experiential versus Informational Strategies" *In: Martin, L. & Tesser, A. (eds.)* The Construction of Social Judgment. *Hillsdale, N.Y.: Erlbaum, pp. 249 – 275.*
Tversky, A. (1972) Elimination by aspects: A Theory of Choice. *Psychological review, vol. 79 (4), pp. 281-296.*
Tversky, A., Kahneman, D. (1974) Judgment under Uncertainty: Heuristics and Biases. *Science, vol. 185, pp. 1124 – 1131.*
Ward, L., Lockhead, G. (1971) Response System Processes in Absolute Judgment. *Perception and Psychophysics, vol. 9, pp 73-78.*
Ward, L. (1973) Repeated magnitude estimations with a variable standard: Sequential Effects and Other Properties. *Perception and Psychophysics, vol.13, pp. 193-200.*
Ward, L. (1979) Stimulus Information and Sequential Dependencies in Magnitude Estimation and Cross-Modality Matching. *Journal of Experimental Psychology: Perception and Performance, vol. 5 (3), pp. 444-459.*
Wedell, D., Parducci, A., Geiselman, R. (1987) A Formal Analysis of Ratings of Physical Attractiveness: Successive Contrast and Simultaneous Assimilation. *Journal of Experimental Social Psychology, vol. 23, pp 230-249.*
Wedell, D., Parducci, A., Lane, M. (1990) Reducing the Dependence of Clinical Judgment on the Immediate Context. Effects of Number of Categories and Type of Anchors. *Journal of Personality and Social Psychology, vol. 58 (2), pp 319-329.*
Wedell, D. (1996) A Constructive – Associative Model of the Contextual Dependence of Unidimentional Similarity. *Journal of Experimental Psychology, vol. 22 (3), pp 634-661.*
Wedell, D., Pettibone, J. (1999) Preference and the Contextual Basis of Ideals in Judgment and Choice. *Journal of Experimental Psychology: General, vol. 128, pp 346-361.*
Wyer, R.S. (1974), Changes in meaning and halo effects in personality impression formation. *Journal of personality and Social Psychology, Vol.29 (6), pp.829-835.*

APPENDIX A
List of author's publications, related to the dissertation:

Kokinov, B., Hristova, P., Petkov, G. (2004) Does Irrelevant Information Play a Role in Judgment? *In: Proceedings of the 26th Annual Conference of the Cognitive Science Society, 2004. Erlbaum, Hillsdale, NJ.*

Petkov, G. (2005a) Judgment as Mapping (JUDGEMAP). *In: Technical Report LIP 2005/007 of the Laboratoire d'Informatique de Paris 6, URL http://www.lip6.fr/fr/production/publications-rapport-fiche.php?RECORD_KEY%28rapports%29=id&id(rapports)=243 pp. 95-104*

Petkov, G. (2005b) JUDGEMAP2 –Cognitive Model for Judgment. *In: Proceedings of the Balkan Conference of Young Scientists, 2005, Plovdiv. In press.*

Petkov, G., Hristova, P., Kokinov, B. (2005c) How Irrelevant Information Influences Judgment? *In: Technical Report LIP 2005/007 of the Laboratoire d'Informatique de Paris 6, URL http://www.lip6.fr/fr/production/publications-rapport-fiche.php?RECORD_KEY%28rapports%29=id&id(rapports)=244 pp.75-78*

Hristova, P., Petkov, G., Kokinov, B. (2005d). Influence of Irrelevant Information on Price Judgment. *In: Proceedings of the International Conference on Cognitive Economics.* NBU Press.