Chapter 6

Cognitive Plausibility

In this chapter we investigate the cognitive plausibility of the conceptual framework for qualitative prediction of behaviour (cf. [18]). We compare think-aloud protocols of human subjects predicting the behaviour of a complex configuration of balances with a computer model of the same problem solving task, implemented in *GARP*.

The contents of this chapter are as follows. The first section explains why cognitive plausibility is important. The second section discusses the nature of the problems on balances in more detail. The third section focuses on potential strategic knowledge that subjects may use during a behaviour prediction task. The fourth section describes the protocol analysis of the problem solving task and discusses to what extent the framework for qualitative prediction of behaviour fits the think-aloud protocol data. The last section summarises and further discusses the important results .

6.1 Relevance of Cognitive Plausibility

Developing conceptual frameworks for modelling problem solving expertise is considered to be an important aspect of research on knowledge based systems. It is a generally accepted hypothesis that such frameworks can usefully support the knowledge acquisition process. However, little research has been published that investigates this hypothesis. In particular, the claim that these conceptual models are knowledge level models (cf. [109]) has seldom been investigated in experimental research.

The issue of cognitive plausibility is relevant for deciding on which parts of the problem solving competence should be described in the design model and which in the analysis model. In our research, for example, we constructed a complex algorithm for specifying new states of behaviour (section 5.2.3). The question is, how can we decide that this algorithm should not have been a part of the analysis model, i.e. of the knowledge level model of qualitative prediction of behaviour.

Erroneously representing design (or symbol) level issues in an analysis (or knowledge) level model, will lead to a knowledge acquisition process on the wrong level of abstraction. As a result, the expert will be confronted with inappropriate questions and the analysis of his expertise will be guided by the wrong model. We do, for example, not want to question an expert on diagnosis (cf. [59]) about the knowledge that he uses in his ATMS (cf. [54]). The ATMS is a computer specific mechanism that can be employed for keeping track of inference dependencies, but it is not something that is used by human experts.

In many cases the difference between knowledge level and symbol level is not that straightforward. It is therefore important that we investigate whether we represented the knowledge, involved in qualitative prediction of behaviour, at the right level of abstraction. This explains why testing the cognitive validity of the conceptual framework is important.

6.2 The Balance Problems

To investigate the cognitive validity of the theory underlying *GARP*, the qualitative reasoning task was operationalised with six balance problems (see figure 6.1). The problem is to predict the behaviour of balances with containers on each balance arm.

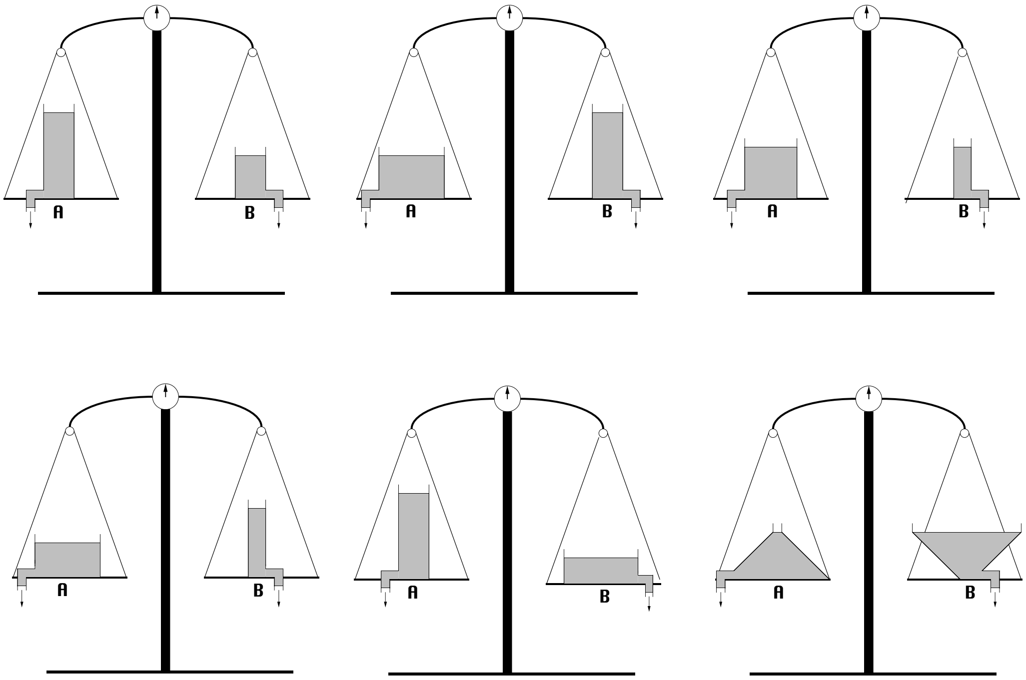

Figure 6.1: The balance problems

Both containers are assumed to be equal in weight. Depending on the difference in the mass of the water contained by the containers, one balance arm may be heavier than the other. Therefore, after releasing it from the starting position ($position = zero$), the balance may change its position.

Through outlets near the bottom of the containers the water gradually flows out of the containers. Depending on the pressure at the bottom, the $flow\_rates$ of the two containers may be different. As a result the balance may move to a different position, because the difference in weight between the two balance arms changes. Eventually, when both containers are empty, the balance will have reached an equilibrium.

Predicting the different states of behaviour that the balance goes through, after it is released, is the goal of solving the balance problems.

6.2.1 System Elements and Parameters

The structural description of the balance problems is depicted in figure 6.2. It consists of a

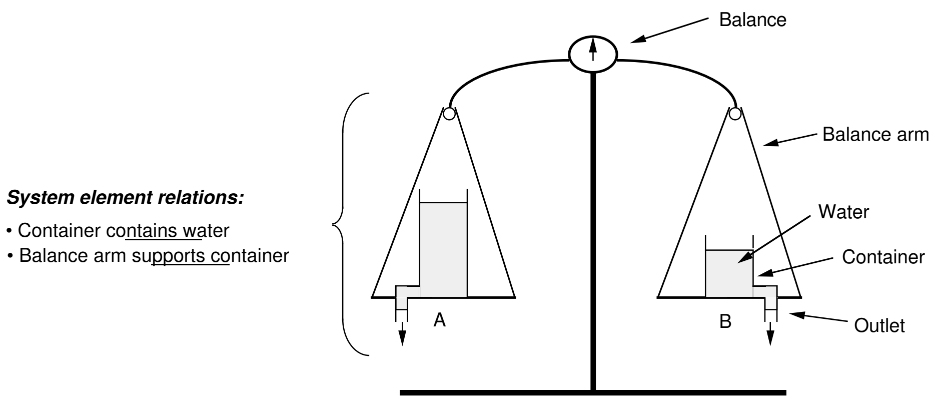

Figure 6.2: System elements of the balance problems

balance with two balance arms. Each balance arm supports a container. Both containers contain a certain amount of water. There are equally shaped outlets near the bottom of each container.

The parameters that describe the behaviour of the container and the water it contains, are shown in figure 6.3. For the behaviour of the balance it is sufficient to give each

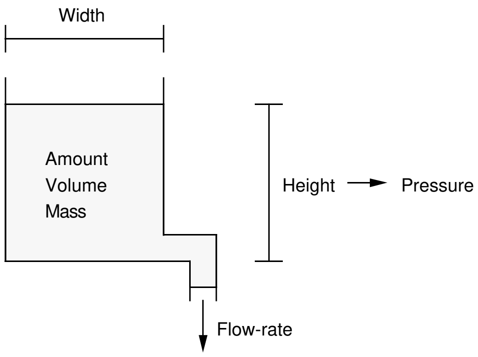

Figure 6.3: Parameters of a container

of these parameters values from the quantity space *zero-plus* (the quantity is present or not). In addition to the parameters for the containers, a *position* parameter has to be introduced for the balance (see figure 6.5). This parameter requires the quantity space *min-zero-plus*, respectively referring to the three possible positions of the balance. How the balance moves (the derivative of the position) is shown in figure 6.6 (see also below).

6.2.2 Parameter Relations

In each version of the balance problem, the two containers differ in shape and in the amount of water they contain (see table 6.1). The different instances of the balance problem have been constructed in this way in order to allow as much variation as possible with respect to

Balance problem (figure 6.1)	Width	Height	Amount
Left (top)	$A = B$	$A > B$	$A > B$
Middle (top)	$A > B$	$A < B$	$A = B$
Right (top)	$A > B$	$A = B$	$A > B$
Left (below)	$A > B$	$A < B$	$A > B$
Middle (below)	$A < B$	$A > B$	$A > B$
Right (below)	*unknown*	$A = B$	$A = B$

Table 6.1: Differences between parameters in the six balance problems

the factor(s) that determine(s) the $flow\_rate$ at the outlet. In particular, the parameters *height*, *amount*, and *width* will be important for capturing the different interpretations (including misconceptions) that the subjects may use. By varying these parameters for each of the balance problems, different behaviour predictions are to be expected depending on how the $flow\_rate$ is derived. Figure 6.4 depicts the possible interpretations.

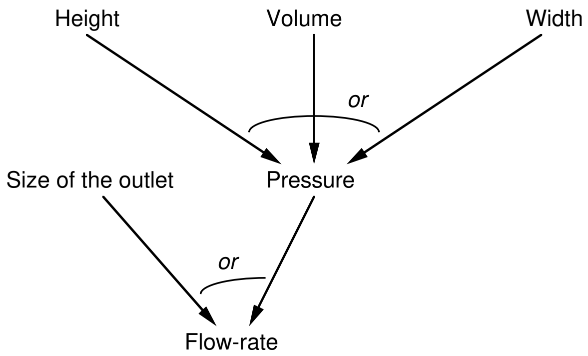

Figure 6.4: Possible ways of determining the $flow\_rate$ (including misconceptions)

Although none of the balance problems varies the outlet, it is possible that subjects use this as a factor for determining the $flow\_rate$.

6.2.3 Partial Behaviour Models

The partial behaviour models specify the different viewpoints that subjects may use in solving the balance problems. The different ways of determining the $flow\_rate$, as discussed above, will be represented as alternative models for establishing the behaviour properties of a $contained\_liquid$ (composition model). In addition, partial behaviour models are needed for:

- water flow from container to world (process)
- balance arm supporting container (composition model)
- position of balance configuration (composition model)
- movement of balance configuration (process)

The flow of water from the container into the world applies to a 'contained liquid' with an outlet near the bottom of the container. If this configuration exists and the outlet is open, then a liquid flow process introduces a $flow\_rate$ that influences the amount of water negatively. The balance arm supporting a container is used to establish the total weight for each of the balance arms. The behaviours introduced by these composition models is used by the balance configuration to determine the position of the balance. Three possible positions are depicted in figure 6.5. The movement of the balance depends on the

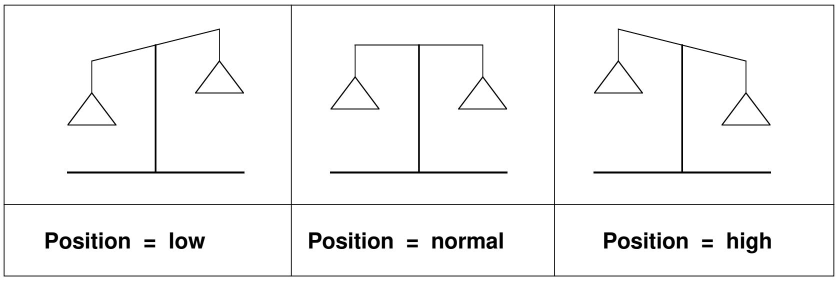

Figure 6.5: Partial behaviour models for the position of the balance

difference in $mass\_loss$ between the two contained liquids. The three possible movements of the balance are: (1) left side down and right side up ($movement = plus$), (2) balanced ($movement = zero$) and (3) left side up and right side down ($movement = min$) (see also figure 6.6). The movement represents the derivative of the position.

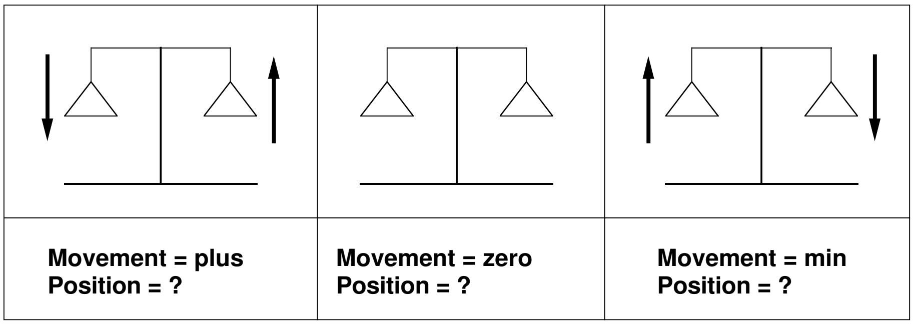

Figure 6.6: Partial behaviour models for the movement of the balance

Using all the behaviour models, the total set of dependencies between the different parameters of the balance configuration turns out to be complex (see figure 6.7). The $width$ of the container determines the $width$ of the liquid column. The $height$ of the latter depends on the $width$ of the column and the $amount\_of$ liquid that is present. The $height$ determines the $pressure$ at the bottom of the container and as such the $flow\_rate$ of the liquid flow out of the container. The $flow\_rate$ is qualitatively similar to the $mass\_loss$ except that the latter is a property of the liquid and not of the liquid flow process.[1] The $mass\_loss\_difference$ between the two arms of the balance depends on the individual $mass\_loss$ of each container.

[1]Notice, that the $mass\_loss$ could differ from the $flow\_rate$ if there were other $flow\_rates$ influencing the $amount$ of liquid. In the balance problems this is not the case.

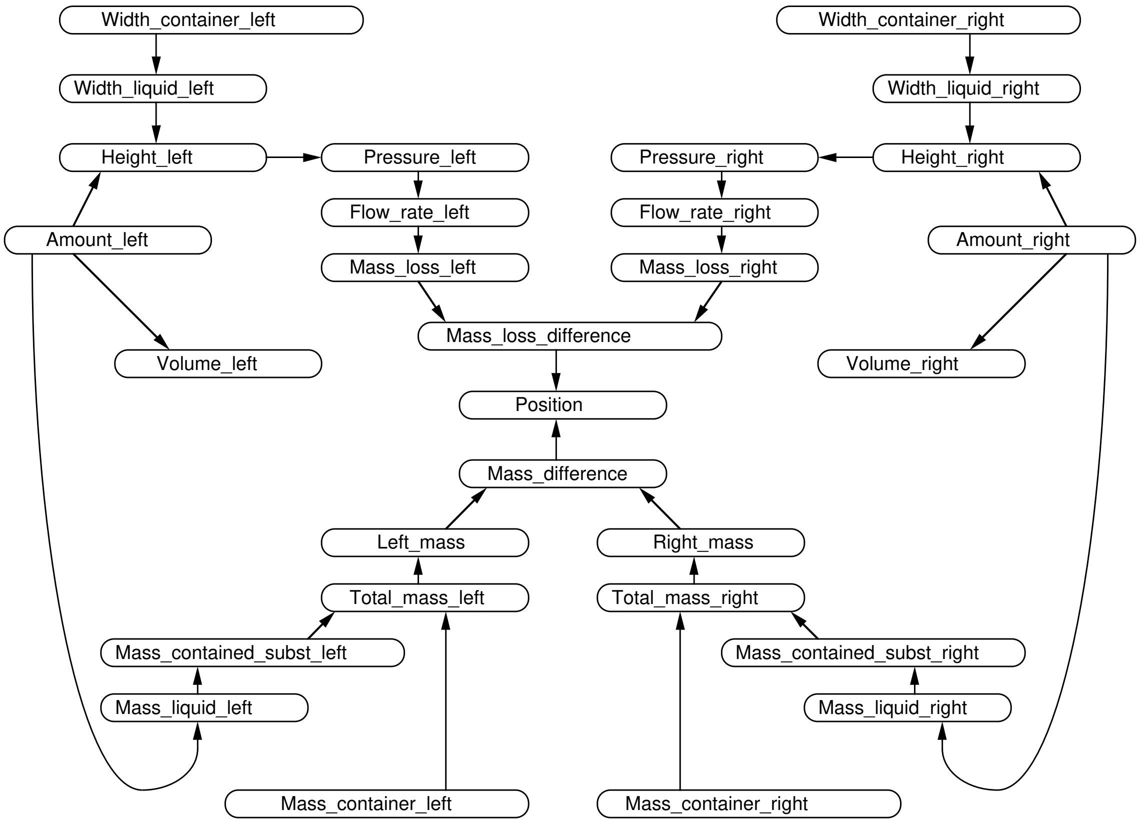

Figure 6.7: Detailed model of the parameter relations

The *amount_of* liquid determines the *volume* and *mass*.[2] The *mass* of the liquid and the *mass* of the container (and the *mass* of the balance arm itself), determine the *total_mass* for a balance arm. The difference between the *total_mass* of the two balance arms determines the position of the balance (and the initial movement of the balance after it is released from its starting position.

6.3 Task Knowledge and Strategic Knowledge

The task layer is used to represent typical chains of inferences that experts make in solving a particular, well-known task. It is to be expected that human subjects will concentrate only on attainable envisionments, i.e. a prediction of the actual behaviour of the balance (and not a description of all possible behaviours).

Strategic knowledge is rarely present in the original approaches to qualitative reasoning. They always execute the same task structure, are not able to monitor their own inference process, and as such are not able to modify or change their own reasoning process. However, it is likely that subjects will encounter difficulties in the reasoning process,

[2]Depending on the specific conceptualisation used by the subject *amount_of* and *volume* may represent the same entity or be different. In the case of the latter *amount_of* is used for referring to the number liquid molecules present.

because the problem solving goal cannot be reached. We expect two types of difficulties to emerge (cf. [6]):

- Problems with the available domain knowledge (=knowledge conflicts). In particular, we expect problems with missing, ambiguous, or contradictory facts in the domain knowledge.
- Insufficient processing capacity.

For the knowledge conflicts we expect repairs as listed in table 6.2. The shortage of processing capacity will not be discussed further in the research reported here.

Knowledge conflict	**Type of repair**
Missing knowledge	*Practical repairs:* read the question again ask for additional information *Repairs by reasoning:* continue reasoning after making an assumption try extreme values use analogy use different/other domain knowledge
Ambiguous knowledge	try one (randomly or according to an estimate of success) try all reason backwards from a known final state
Contradictory knowledge	try again check computations use other domain knowledge

Table 6.2: Repairs for knowledge conflicts

Once a repair is selected its plan is tuned to fit the specific instance of the impasse. This means that reasoning at the strategic level may involve a revision at the task level.

6.4 Protocol analyses

To test the cognitive plausibility of the conceptual framework we compared problem solving activities as manifested by human subjects in think-aloud protocols with those predicted by the framework. The problem solving model that was built and implemented in *GARP* was therefore translated into a *coding template* with which the think-aloud protocols could be analysed. In table 6.3 a coding template is shown that assigns a coding category to each of the knowledge sources in the inference structure (depicted in figure 4.11).

The subjects were ten psychology sophomores who had taken high school courses in physics. Three protocols were coded entirely, resulting in 673 coded expressions. All ten protocols were screened for the occurrence of the following three criteria:

- *The model is incomplete*
  There are expressions that cannot be encoded.

Knowledge sources in the inference process		**code**
Compound specification		I-1
	Assembling the mathematical model	I-1.1
	Computing parameter values/relations	I-1.2
Compound transformation		I-2
	Selecting terminations	I-2.1
	Sorting terminations	I-2.2
	Transforming into a new SMD	I-2.3

Table 6.3: Coding schemata for inference knowledge

- *The model is too detailed*
  There are coding categories that are not used.
- *The model is wrong*
  There are deviations in the expected order of expressions.

The results of this analysis are described for different layers of the conceptual model in the next three sections. An overview of the the coded expressions is given below:

- Domain and inference layer: 62 percent
  - Specification: 37 percent
    (assemble:compute ⇒ 2:3)
  - Transformation: 25 percent
    (mostly selecting terminations)
- Task and strategic layer: 40 percent
  - Detecting impasses: 7 percent
  - Repair impasses: 33 percent
    * Specification: 12 percent
    * Transformation: 11 percent
    * Others: 10 percent
- Not coded: 21 percent[3]
  (all referred to relating different versions of the balance problems)

6.4.1 The domain layer

In going through the sequence of problems, subjects seemed to have a tendency to abstract from the details of the system. While solving the first problems, specific *system elements*, like container and water, were referred to. In the later problems the behaviour of the system was described as if only one object was present, the balance, with a mass-loss on two sides, one losing mass faster than, or with equal speed to, the other. In a similar way, many *parameters* were used while solving the first problems. For example, *height* (or

[3] $21 = 100 - (62 + 40 - (12 + 11))$

amount etc.) was seen as determining the *pressure*, which in turn determines $flow\_rate$ which equals $mass\_loss$ (see figure 6.7). In the abstracted form the *height* determined the $mass\_loss$ directly (see figure 6.8).

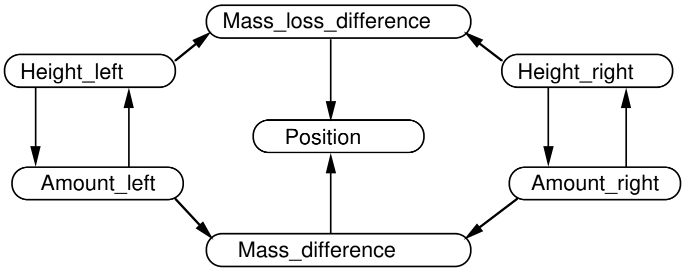

Figure 6.8: Abstracted model of the parameter relations

The detailed model can be regarded as a *context independent* model, because it represents all the information for reasoning about the behaviour of the system. The abstracted model, on the other hand, can be seen as a *context dependent* model. It abstracts from irrelevant details and focuses on those aspects of the system that are important for the context in which it is used (cf. [18]). Although both models can be represented in $GARP$, the abstraction process itself cannot.

The ten subjects used different *partial behaviour models* for determining the $flow\_rate$. One subject used the *size* of the outlet, but shifted, in a later version of the balance problem, to *width* of the water column. Only one subject used *height* of the water column (although three other subjects also considered it). Six subjects used *volume* of the water column, three of whom after showing doubts about other possibilities. Only one subject used the *width* of the column water from the outset. The remaining subject used an alternative viewpoint: the *weight* of the air over the water-surface determines the $water\_pressure$.

6.4.2 The inference layer

Compound specification made up 37 percent of the protocol statements, with a ratio of 2:3 between *assemble* and *compute*. Assembling the mathematical model typically appeared in the protocols as expressions in which was stated that a parameter had a value or that a relation or inequality was applicable. Computing values made implicit use of a qualitative calculus. It appeared in remarks like: 'because the pressure in A is higher the water flows out with more force'. Below some examples of these inferences from the protocols are listed:

- *Specify*

  It just matters ...

  In a balance it just matters what the weights are

- *Compute*

  Well, I think first it goes down at the side of container A

  Because at the outset it is heavier of course

  Because here [A] is more water

- *Specify/Compute*
  Let me think if it flows here [A] faster than here [B]
  Here [A] the pressure is, let me think
  Yes, here [A] it is higher of course at the bottom of the container
  Does it influence the flow rate, let me think
  Yes, the force of outflow will be bigger, undoubtedly

The sequence of assembling followed by computation was regularly violated and as such differed from what the *GARP* model predicted. First, for one part of the system, applicable partial models were found and computable parameter values were calculated. This was repeated for the next part of the system. It seemed as if applicable partial models focused the search for new ones.

Compound transformation made up 25 percent of the protocol statements. *Selecting* terminations was expressed in ways a state of behaviour ends: 'then the amount of water in container A becomes equal to that in B'. The *sorting* of terminations was only present in the protocols when the reasoning process was disrupted, i.e. if ambiguity in a situation was detected. Below some examples of these inferences from the protocols are listed:

- *Select (terminations)*
  A goes down ...
  Until B is entirely empty ...

- *Sort (terminations)*
  Yes, they either are empty at the same time or ...
  B is already ...
  I cannot imagine A being empty sooner, but why can't I ?
  I just, it's more a feeling, let me think
  A has got greater pressure but ...
  No, I wouldn't know if they'd become empty at the same time

The limited occurrence of the sorting inference can be explained in two ways. First, sorting occurs only if more than one termination is found. If so, the subjects did not find impossible terminations and terminations that were aspects of the same change, they just found possible ones (sometimes comprising different aspects). Second, impossible terminations and different aspects *were* found but the sorting of terminations in unambiguous cases is an automated process and therefore not reportable. The fact that the knowledge used to make this inference is not domain specific, makes an automated process plausible. However, additional research is needed to resolve this issue.

Little evidence was found for the third inference that constitutes the *compound transformation* (transforming into a new system model description).[4] We must, however, assume that subjects make this inference in order to preserve the continuity between one state and the next: values and inequalities that do not change in the next state are propagated. The lack of expressions confirming this inference can be explained by regarding it as a default mechanism. It provides a good example of how subjects deal with the frame

[4] Notice, that this inference, which follows sort, is also called transformation.

problem, that is, unless something explicitly changes, they consider things to be constant over states.

6.4.3 The task and strategic layer

A total of 40 percent of the protocol expressions was encoded as strategic reasoning, of which 7 percent was devoted to the detection of impasses, 33 to the overcoming of these impasses. This includes 12 percent specification and 11 percent transformation tasks that were executed in, for example, reasoning backwards from the derived final state, starting all over again and re-checking the computations. The remaining 10 percent covered other repairs such as making assumptions, comparing competing knowledge, using analogies and reading the question again. Below follow examples of these repairs from the protocols:

- *Extremes*
  I wonder if in the case of a narrow, very narrow and very high column ...
  If the pressure is higher than in a very wide container with a very thin layer of water
  Just to imagine if it matters, if it's volume that matters or just the height of the column

- *Assumption*
  I think it's volume that matters
  So they flow out with equal speeds and the balance remains equilibrated

- *Analogy*
  Maybe it's like, a bit strange maybe:
  Taller people have bigger feet, so the pressure is more spread

- *Back from end*
  So, after a while it will ...
  Yes, in the end it will come back to equilibrium, no doubt, but ...
  Let me see

The expressions that could not be coded, were all of one type: they referred to an activity that related different problems of the sequence. Apparently, the subjects were able to see similarities in two versions of the problem. This also explained the occurrence of the immediate production of an answer without any apparent specification and transformation in a few places. When the model, built for one version of the problem, was applied to another in the sequence, the complete behaviour description was applicable so the answer could be produced immediately. This type of inferencing could not be accommodated in the model represented in *GARP*.

6.5 Concluding Remarks

The framework presented in this thesis extends previous descriptions of qualitative prediction of behaviour by distinguishing between domain, inference, task, and strategic knowledge. The conceptual framework implemented in *GARP*, which is based on these knowledge types, appears to be appropriate for describing and interpreting the reasoning processes involved in this problem solving task. Both the different viewpoints subjects have on the domain knowledge as well as their reasoning process can be modelled by the framework. The canonical inferences and the meta classes defined in the model provide strong means for interpreting the steps of the reasoning process in the protocols. The notion of strategic reasoning explains disruptions and changes in the order in which new states are determined.

However, from the experiment presented here it also becomes clear that the model can be refined in some places. In particular, it is unclear what determines the order of the assembly and the computation inference during the specification of a system model description. The notion of *p-primes*, as introduced by DiSessa [62], can be used for explaining this phenomenon. P-primes can be regarded as a measurement of the likelihood that certain partial behaviour models are good candidates for augmentation of the current system model description. Behaviour models that have a high score on their p-prime should be favoured above those with a low score.

The modelling framework also does not account for learning over a number of problem solving sessions. As the subjects moved from one problem to another they abstracted from irrelevant details in the description and the analysis of the problem. The framework does not support this abstraction process. It seems, however, that this lack of support is part of a larger problem in $KADS$, namely the lack of structuring principles at the domain layer. The theory for modelling problem solving (chapter 3) does not provide normative support for chosing a level of abstraction at which the knowledge in the domain layer should be modelled.

The notion of task structures was less useful in analysing the protocols. Experts performing a particular problem solving task may develop, through the repeated execution of the same sequence of inferences, a trace of this sequence and as such learn a task structure. That is, an instance of a strategic reasoning process. Non-experts, like the subjects in the experiment reported on here, just use general strategic reasoning.

In conclusion, it is fair to assume that the conceptual model presented in chapter 4 does describe this problem solving task at the right level of abstraction, i.e. it constitutes a knowledge level model of this problem solving expertise. In addition, the presented research provides a basis for further research on the way humans perform qualitative prediction of behaviour (cf. [146; 97]). Such research should in particular focus on the learning aspects and the knowledge structuring principles that people use for developing their domain knowledge.