Mental models and Local models semantics: the problem of information integration

Paolo Bouquet
Dipartimento di Informatica e Studi Aziendali
Università di Trento
Via Inama, 4 – I-38100 Trento (Italy)
Phone: +39-0461-882135
bouquet@cs.unitn.it

Massimo Warglien
Dipartimento di Economia e Direzione Aziendale
Università Ca' Foscari di Venezia
Dorsoduro 1075 – I-30123 Venezia (Italy)
Phone: +39-041-2578745
warglien@unive.it

Cognitive Science Laboratory
University of Trento
Via Tartarotti, 7 – I-38068 Rovereto, TN (Italy)
Phone: +39-0464-483515
lsc@inf.unitn.it

ABSTRACT

In this paper, we sketch a framework complementing the *theory of mental models* with the notion of *context*. A context is defined as a collection of *local mental models*, namely mental models which represent only part of a situation. We argue that most reasoning tasks essentially involve a stage at which the information represented in a set of initial "input" contexts is integrated in a more comprehensive "target" context, where a solution can be searched for. Our claim is that this more structured account of mental representations can provide insights in the constructive process through which mental models are generated and assembled. The preliminary experimental evidence reported in this paper shows that the organization of contexts matters in explaining how individuals perform disjunctive reasoning tasks, and that the difficulty in information integration complement the number of mental models as an explanation for deductive failures.

Keywords

Mental models, contextual reasoning, disjunctive reasoning.

INTRODUCTION

Two key assumptions lie at the heart of the *theory of mental models* (TMM): that human reasoning has a semantic nature, and that representations (sets of mental models) underlying the reasoning process are partial - i.e. they provide only a partial truth-functional evaluation ((Johnson-Laird 1983, Johnson-Laird & Byrne 1991); see also (Doherty 1996, Fenstad 1997) for a formal treatment of semantic partiality).

The TMM embodies the assumption that the "space" within which a problem is represented is unique - namely, there is a sort of mental blackboard where reasoners append the mental models they construct. Once such a collection of models is built, they are considered as a list that can be scanned in search for some conclusion and eventually checked out for validation. As such, mental representations are sets of models without further structure, and the number of models needed to provide a complete representation is the main proxy of the structural complexity of the reasoning task.

This contrasts with several works in cognitive science and artificial intelligence (e.g. (Dinsmore 1991, Fauconnier 1985, Giunchiglia & Bouquet 1998, Guha 1991, McCarthy 1993)) suggesting that there is more structure in human reasoning, and that the cognitive space where people represent and solve reasoning problems is better thought of as a structured collection of relatively small spaces (or *contexts*, as they are often called in AI: e.g. (Giunchiglia & Ghidini 1998, McCarthy 1993)), each of them providing a partial representation of a state of affairs. In other words, the idea is that – even for simple problems – reasoners do not construct a single representation, but instead they start with a collection of local representations and then try to integrate them in a comprehensive picture. This suggests an explanation of deductive failures complementary to the one provided by the number of models needed - namely *the difficulty in integrating local representations*.

There is a widespread intuition in the psychological literature that individuals may suffer from severe difficulties in integrating different pieces of information available on a problem. Well-known examples include many cases of disjunctive reasoning (see (Shafir 1993) for a review), mental accounting (Thaler 1985), meta-deductive tasks such as the three hats problem (see the works by George Erdos at the University of Newcastle), and the identification of dominance in combinations of lotteries (Tversky & Kahneman 1981). However, to the best of our knowledge, there has been little effort to explictly introduce into the theory of mental models the internal structure of mental representations and its processes of integration .

In this paper, we sketch a framework complementing the TMM with the notion of *context*. A context is defined as a collection of *local mental models*, namely mental models which represent only part of a situation. We argue that most reasoning tasks essentially involves a stage at which the information represented in a set of initial "input" contexts is integrated in a more comprehensive "target" context, where a solution can be searched for. Our claim is that this more structured account of mental representations can provide insights in the constructive process through which mental models are generated and assembled. The preliminary experimental evidence reported in this paper shows that the organization of contexts matters in explaining how individuals perform disjunctive reasoning tasks, and that the difficulty in information integration complement the number of mental models as an explanation for deductive failures.

LOCAL MODELS SEMANTICS

The notion of context we use in this paper has been intuitively defined in (Giunchiglia 1993): "A context is a theory of the world which encodes an individual's perspective about it. It is a *partial theory* as the individual's complete description of the world is given by the set of all contexts. It is an *approximate theory* [...] as we never describe the world completely". A context, in this sense, is quite close to the notion of *mental space* proposed by Fauconnier (1985) and even closer to the notion of *partitioned representations* by Dinsmore (1991). We adopted Giunchiglia's notion of context because it has been used as a basis for a formalization of *contextual reasoning*, that is reasoning involving facts from multiple contexts. Since in this paper we are mostly concerned with model-based reasoning, we will summarize the principles of contextual reasoning from a semantic perspective.

Local Models Semantics (LMS) is a general semantic framework for contextual reasoning (Giunchiglia & Ghidini 1998). It is based on two general principles:

- *principle of locality*: when reasoning, an agent does not use all s/he knows about the world, but only a subset of it. This subset is called a context;

- *principle of compatibility*: there are structural relations between contexts. These relations are called compatibility relations.

We can explain the two principles with a simple example. Suppose there are a red (R), a green (G), and a blue (B) blocks in a room. The room has two doors (D_1 and D_2), each of which allows an agent A to see only part of the room. Suppose that the three blocks are lined on the floor, so that only R and G are visible through D_1, and only G and B through D_2. Now imagine that A is asked the following questions about what she sees:

1. Is R to the left of G? Is B to the right of G?
2. What happens if you swap R and G? What happens if you swap G and B?

In order to answer the first two groups of questions, A needs to use only a subset of what she knows. For example, answering the question "Is R on the left of G?" only requires a model of what A sees through D_1, whereas answering the question "Is B on the right of G?" only requires a model of what A sees through D_2. Let us call M_1 and M_2 these two models. They are depicted in Figure 1 (left and right, respectively). In LMS, M_1 and M_2 are called *local models*, since they represent only a portion of what A knows about the room (and *a fortiori* about the world in general). Semantically, a *context* can be viewed as a collection of local models of the same portion of a situation[1]. Thus in our example, we have two contexts: the first, C_1 representing the portion of the room that A can see through D_1, and C_2, representing the portion of the room that A can see through D_2. A reasoning process that uses only local models of a single context is called *local reasoning*. This provides an illustration of the principle of locality.

Figure 1: The models M_1 and M_2

Apparently, the situation is about the same for the questions in (2). After all, the consequences of swapping R and G can be derived by local reasoning in C_1, and analogously for G and B in C_2. But this is not the whole story. Indeed, swapping R and G will have the side effect of changing what A can see through D_2. This is precisely what the principle of compatibility says: there may be compatibility relations between (sets of) local models belonging to different contexts. In our example, the compatibility relation is the following: the block occupying the right position in any model in C_1 and the left position in a model in C_2 must be the same (because

[1]Intuitively, a context is defined as a collection of local models – and not just a single model – because in general an agent has partial information about a state of affairs. Therefore, having a collection of models allows modelling the fact that there are propositions about a state of affairs that an agent neither believes nor disbelieve. Technically, this means that for each of these propositions there are local models of a context in which it is true and others in which it is false.

the portions of the room visible through D_1 and D_2 are partially overlapping); otherwise, two models are incompatible. Figure 2 represents the complete compatibility relation corresponding to our example. Models in the left column are all the possible "views" through D_1 (and therefore all the possible local models in C_1); models in the left column are all the possible "views" through D_2 (and therefore all the possible local models in C_2)[2]. The dashed lines connect compatible local models. Notice that a local model in a context can be compatible with more than one local model in the other. This represents the fact that what is true in a context does not completely determine what is true in the other. Thus, if A sees no blocks through D_1 and cannot see through D_2 (imagine the door is closed), then she can only infer that either there are no blocks in the other portion of the room, or B is on the right, or G is on the right (just follow the dashed lines connecting the last model in the left column with models in the right column).

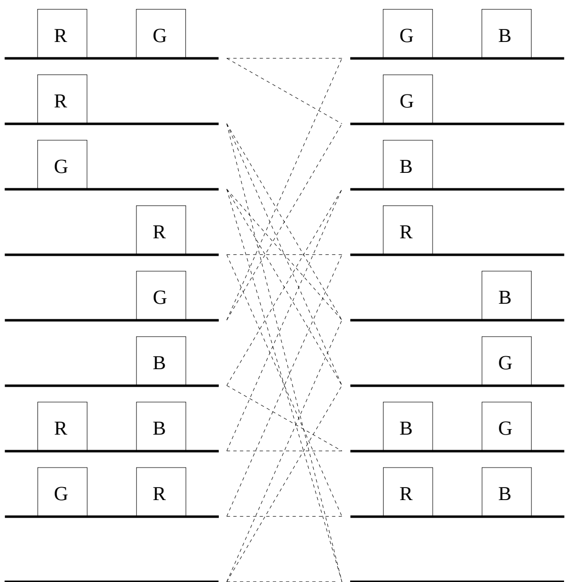


Figure 2: The compatibility relation between C_1 and C_2

Compatibility relations can be used to integrate information about portions of a more comprehensive situation. But they can also be used to construct a model of such state of affairs by putting pieces together. In our example, this would mean the construction of a more general context C_{1-2} in which the three blocks are represented together starting from the local models in C_1 and C_2. Figure 3 shows one model of this context (namely the model corresponding to the initial situation in Figure 1). Semantically, building C_{1-2} is still an instance of a compatibility relation. Only, the compatibility relation is now a 3-place relation between local models of C_1, C_2 and C_{1-2}. For example, the model in Figure 3 is not compatible with any other pair of models in Figure 2, but the ones on top of the two columns.

[2]For the sake of simplicity, we assume that R cannot move to the far right and B to the far left of the room. Otherwise, we would get more local models for each context and more links between them.

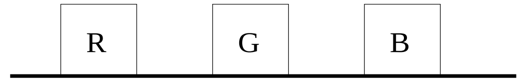


Figure 3: The model M_{1-2}

If, from a logical point of view, the construction of M_{1-2} is relatively straightforward, not in every situation real reasoners are able to make such an integration. On the contrary, sometimes the integration can prove very hard. In this sense, LMS can be used as a complement to the TMM to represent reasoning process and explain some kinds of deductive failure.

THEORY OF LOCAL MENTAL MODELS

The framework we propose in this paper is an extension of the TMM that uses the principles of contextual reasoning to model the process of integrating information from multiple sets of mental models.

One of the most important motivations for the TMM is the idea of partiality. Given a problem – for example a deductive problem – reasoners tend to construct only a subset of the possible models of the premises, and therefore they may fail to see a valid conclusion. However, the TMM seems to disregard another important notion of partiality, namely that a single situation can be represented by constructing multiple sets of mental models, each of which is a partial representation of a portion of the original situation. In other words, there are two different levels of partiality:

1. on the one hand, an articulated state of affairs can be partitioned into smaller portions, each of which is represented with a distinct set of mental models (see the example of the section above). We propose to call these sets *contexts*, and the mental models within a context, *local mental models* (LMMs);

2. on the other hand, each of these portions is only partially represented (in the sense of the orginal TMM). So a context is partial in two ways, since it is a partial representation of a portion of a state of affairs.

In order to take into account this second form of partiality, we propose to modify the semantic procedure described in (Johnson-Laird 1983, Johnson-Laird & Byrne 1991) as follows:

- when confronted with a problem, reasoners construct one or more input contexts, each of them containing a set of LMMs (in the sense defined above). The way a representation is partitioned into multiple contexts depends on multiple factors, among which an important rôle is played by those related to the linguistic formulation of the problem (see (Fauconnier 1997) for a convincing discussion of this point);

- if the problem requires to integrate information from more than one input context, then a target context is constructed that provides a more general representation of the problem. The target context is the result of integrating LMMs according to some compatibility relation which holds among LMMs belonging to different input contexts;
- the formulation of a putative conclusion is obtained in a target context and then checked by searching for alternative models that can falsify it. If such a model exists, the conclusion is retracted and the reasoners either (i) formulate a new putative conlusion within the target context, or (ii) go back to the second stage and checks the application of the compatibility relation.

This extended version of the TMM, which we call *Theory of Local Mental Models* (TLMM), provides us with another dimension along which reasoning failures can be explained beside the number of mental models that reasoners must take into account. Indeed, we claim that the way in which contexts are constructed and organized affects the difficulty of the reasoning task. In particular, the TLMM entails two general hypotheses:

1. the organization of contexts required to solve a problem may affect the type of errors individuals make in drawing their conclusions;
2. organizations of contexts that reduce the number (or complexity) of constraints between LMMs correspondingly reduce the difficulty of the task.

The next section provides some preliminary experimental evidences that support these hypotheses.

A CASE STUDY: DISJUNCTIVE REASONING

Disjunctive reasoning is an ideal testbed for our framework: disjunctions are basic logical and linguistic devices for generating representations of alternative states of affairs. Reasoning through disjunctions implies both separating and integrating the truth values of propositional variables. In this paper we restrict our attention to the exclusive disjunction. One should expect reasoning with such connective to be rather straightforward. Its semantics is rather simple (according to TMM, only two, symmetric models need to be explicited) and its interpretation in logics is very close to its use in natural language. Thus, experimental evidence that individuals frequently fail in basic deductive tasks with the exclusive disjunction is especially puzzling. We suspect that part of the difficulties involved in reasoning with such connective are not due to the complexity of the task (i.e. the number of explicit models needed to make a correct deduction) but instead to the peculiar nature of disjunction, that may induce the editing of disjoint local representations (contexts) of the alternative states of affairs. In order to test the impact of the structure of local representations on deductive thinking, we revisit some classical experiments on disjunctive reasoning (Johnson-Laird & Byrne 1991, Johnson-Laird, Byrne & Schaeken 1992). Our revisitation consists in simple manipulations of the presentation format of the original experiments. These manipulations are based on the assumption that language contains operators that "initialize" local representations (e.g. person names), inducing the opening of specific contexts within which mental models are located. The intuition is that, keeping constant the number of mental models needed to flesh out a complete representation, differences in the local organization of mental models might account for differences in deductive performance.

Experiment 1: negative disjunction

The first experiment reconsiders a negated deduction based on exclusive disjunction task. Experimental work by Evans and Newstead (1980) and Johnson-Laird, Byrne and Schaeken (1992) has shown high failures rates in this kind of reasoning task. We have tried to reproduce these original results, while checking for the effect of alternative presentations facilitating the integration of contexts. The between-subjects experiment involved 62 third-year undergraduate students of psychology at the University of Padova. In the basic treatment, in analogy with (Johnson-Laird et al. 1992), subjects are asked to evaluate what follows from a disjunctive premise when a constituent of the premise is negated.

Treatment 1A Suppose the following is true:

Either Mary is in Brussels or Gino is in Rome, but not both
Gino is not in Rome

What follows?

One third of the subjects failed in this task (see Table 1). Their performance was somehow better then in the analogous previous experiment by Johnson-Laird et al. (1992), but still a relevant failure rate persisted. It is noteworthy that in such treatment many subjects answered that Mary can be in Brussels or somewhere else. This seems to rule out the explanation that subjects didn't flesh out enough mental models (on the contrary, one might argue that they fleshed out too many models, including illegitimate ones). Instead, subjects' answers suggest that they couldn't see how the truth value of the elementary sentence concerning Gino could affect the truth value of the elementary sentence concerning Mary. Our hypothesis is that local evaluation of the truth value of the negated disjoint (*Gino is in Rome*) might not affect the truth evaluation of the other disjoint (*Mary is in Brussels*) when the two disjoints are made explicit in different contexts. Treatment 1B has been designed to test the hypothesis that a simple language manipulation, i.e. reframing the premises in terms of properties of a same name, should facilitate the representation of both disjoints in a unique context, and thus make the deductive task easier.

Treatment 1B Suppose the following is true:

Either Francesco is in Brussels or Francesco is in Rome, but not both
Francesco is not in Rome

What follows?

As Table 1 shows, the results support our hypothesis: failure rates dropped from $1/3$ of the subjects to less than 10%.

Answer	**Treatment 1A**	**Treatment 1B**
Correct	66,7%	90,6%
Incorrect	33,3%	9,4%
N. of subjects	30	32

Table 1: Treatment 1A

However, one might argue that the improvement in the rate of correct answers in treatment 1B is mostly due to a pragmatic effect: anybody knows that in our world, "you can't be in two places at the same time". We will say more on pragmatic effects in the subsequent discussion of Experiment 2. We just notice at this stage that the pragmatic effect might directly explain why, if one assumes that Francesco is in Rome, he cannot be in Brussels. But the pragmatic scheme "you can't be in two places at the same time" doesn't directly help to tell where Francesco is when he is not in Rome.

Experiment 2: a thinking aloud experiment on double disjunction

Our second experiment explores a variant of Johnson-Laird's double disjunction task (see again Johnson-Laird et al. 1990). In the TMM, double disjunctions are a classical example of how the explosion of models to be kept in mind causes the breakdown of deductive performance. We don't deny that, but again we suspect that the organization of local representations may matter in affecting why and how individuals make incorrect deductions. Double disjunction tasks consist in couples of disjunctions that are simultaneously true and are made interdependent by the presence in the two disjunctions of a same name with conflicting properties. Treatment 2A reproduces the typical pattern of a double disjunction task (Johnson-Laird & Byrne 1991) (p.59):

Treatment 2A Please carefully consider the two following statements:

a) Either Mary is in London or Jack is in Paris, but not both.
b) Either Charles is in Rome or Mary is in Mallorca, but not both.

Both statements a) and b) are simultaneously true. What can you deduce?

Classical experiments on double disjunction present the subject common to both disjunctions in different syntactical positions (left or right of the connective) in each single proposition . This may be a source of difficulty for individuals, that might be induced to represent information related to a same name in different contexts. Thus, treatment 2B tries a simple syntactical variant of 2A, that doesn't change the logic structure of the task but makes the constraints arising from Mary location more visible:

Treatment 2B Please carefully consider the two following statements:

a) Either Mary is in London or Jack is in Paris, but not both.
b) Either Mary is in Mallorca or Charles is in Rome, but not both.

In both cases, the seemingly logically valid conclusion is that Jack is in Paris or Charles is in Rome, or both (or the less concise: Mary is in London and Jack is in Paris, or Mary is Mallorca and Charles is in Rome, or Jack is in Paris and Charles is in Rome). However, it is interesting to remark that in order to reach such conclusion one needs to introduce a hidden common sense axiom: that Mary cannot be in two places at the same time. Otherwise no connection can be established between the two propositions. This important detail somehow remained unnoticed by Johnson Laird and colleagues in their discussion of double disjunctions. It initially remained unnoticed to us as well; we recognized it only after the experiment was performed. The experiment was performed as a thinking aloud one with 16 undergraduate students of the University of Trento (8 for each treatment). Subjects had as much time as they wanted to give an answer, and were asked to verbalize as much as possible their reasoning processes. Despite the small number of experimental subjects, preventing statistically significant results, we expected that the quality of information would counterbalance the small quantity of cases. Thinking aloud protocols (Warglien & Bouquet 1999) reveal some striking effect of the two different treatments. Quite like in Johnson-Laird et al. (1992) experiments, Treatment 2A turns out to be difficult for most subjects. In contrast, Treatment 2B turns out to be much easier (see table 2).

[3]"Partial answers" are answers of the kind: "Mary is in Mallorca and Jack is in Paris", or "Mary is London and

Answer[3]	**Treatment A**	**Treatment B**
Correct	0	3
Partial	3	3
Wrong	4	2
Premises repeated	1	0
Number of subjects	8	8

Table 2: Experiment 2

In 2B, subjects immediately found the connection between a) and b), identifying the conflicting nature of Mary's location. After a while, this brought them to identify the common sense conclusion (3 subjects out of 8), or at least (3 subjects out of 8) the part of it that includes Mary (Mary is in Mallorca and Jack is in Paris, or Mary is London and Charles is in Rome). Only 2 subjects out of 8 couln't draw any meaningful conclusion. It is even more intersting to notice how they arrived at a conclusion. Most subjects in treatment 2B initially got stuck in the seeming contraddiction that if a) and b) are both true, than Mary has to be both in Mallorca and London; 6 subjects out of 8 intially asked whether Mary has to be considered as being the same person in both propositions a) and b) (while no one asked this question in treatment 2A!). It took a while before they could see that they could use the disjunctive nature of a) and b) to figure out cases in which Mary is in London but not in Mallorca (and thus Charles is in Rome and Jack is not in Paris) or the reverse. Only at the end of this process, and only for some subjects, came the third possibility involving Charles in Rome, Jack in Paris and Mary in no one of its two named locations. It is as though subjects did intially construct a context for Mary, and only subsequently could edit contexts reflecting a) and b), thus making it possible to unlock their reasoning process. It is worth stressing that the two only subjects that failed to reach a meaninful conclusions both got trapped in this initial state and couldn't get out of the seemingly contraddictory position of Mary. One of them finally said:

> I get it! Mallorca is a quarter of London

The comparatively poorer performance of subjects in treatment 2A seems to be at least partly motivated by the difficulty to identify the constraint arising from Mary, or, when seeing it, to keep it active while reasoning on the premises. As an example of the first case, a subject, in the post-experiment debreefing session, remarked:

> Ah, in the second case there was Mary, I thought they were ... I considered 4 persons, I didn't see that Mary was Mary even here.

As a typical example of the second kind of difficulty, we can cite a subject clearly losing track of the constraints:

Francesco is in Brussels". "Premises repeated" are answers that just reproduce the premises.

> Mary is in Mallorca, thus Charles is not in Rome ... so in the second (sentence) Mary cannot be in London thus Jack is in Paris, but then nothing excludes that Charles could be in Rome ...

Another regularity emerging from thinking aloud protocols may help to understand why the different presentation formats of 2A and 2B matter. Almost all subjects reasoned tansforming the disjunctive propositions in conditional ones (e.g.: if Mary is in Mallorca, then Charles cannot be in Rome), always anchoring on the first disjoint of each proposition as the antecedent of the conditional. This induces a structural asymmetry between the two disjoints that may explain why most subjects found easier to identify the constraints arising from Mary in treatment 2B.
Although the intrusion of common sense reasoning in the experiment has been involuntary and undesired, it has accidentally generated an interesting example of how pragmatic considerations can indirectly, rather than directly, affect reasoning. In Experiment 2, the pragmatic scheme that you cannot be in two places at the same time acts as a source of constraints over the combinatorics of logical possibilities, limiting the number of mental models that need to be generated; however, its action is effective only when an appropriate organization of local representations makes the common sense axiom directly representable within a single context.

Experiment 3: double disjunction again

The third experiment was built on what we had learned from Experiment 2, avoiding some of its weaknesses. On the one hand, the experiment involved a larger number of subjects (102 third year undergraduate students in Economics at the University of Venezia). On the other hand, it avoids the intrusion of common sense reasoning assumptions into the thinking process. The experiment only reports subjects' answers; the lack of thinking aloud protocols can by the way be in part compensated by a detailed analysis of subjects answers, as we will see below. The two treatments are as follows:

Treatment 3A Please carefully consider the two following statements:

a) Either Charles is in Rome or Jack is in Paris, but not both.
b) Either Kathy is in Florence or Charles is not in Rome, but not both.

Both statements *a)* and *b)* are simultaneously true. What can you deduce?

Treatment 3B Please carefully consider the two following statements:

a) Either Charles is in Rome or Jack is in Paris, but not both.

b) Either Charles is not in Rome or Kathy is in Florence, but not both.

Both statements *a)* and *b)* are simultaneously true. What can you deduce?

The treatments exploit the same syntactical manipulation of the position of Charles used in experiment 2. At the same time, in this new experiment no common sense assumption is needed to make Charles locations in *a)* and *b)* incoherent. Table 3 summarizes the main results of the experiment.

Answer	Treatment A	Treatment B
Correct	12,96%	18,75%
1 over2	9,26%	8,33%
Incorrect, excluding Charles	25,93%	41,67%
Incorrect, including Charles	38,89%	18,75%
Charles incoherent	9,26%	6,25%
Premises repeated	3,70%	6,25%
Number of subjects	54	48

Table 3: Experiment 3

In order to discriminate subjects reasoning processes, answers to both treatments were coded as follows.

- "Correct" means that answers were correctly specified (including the case in which subjects provided the whole set of possibilities coherent with both *a)* and *b)*).
- "1 over 2" means that subjects provided only one of the two disjunct possibilities coherent with both *a)* and *b)* (e.g. Charles is in Rome and Katy is in Florence).
- "Incorrect, excluding Charles" are only of two types:
  - Jack is in Paris and Katy is in Florence;
  - Jack is not in Paris and Katy is not in Florence;

  Both have been often accompanied by remarks like: "Charles can't be everywhere ..." or "Charles location is contradictory ...".
- "Incorrect, including Charles" have much more variance, and include all kind of wrong deductions like:
  - Charles is not in Florence, Jack is not in Paris and Katy is not in Paris;
  - If Katy is in Florence Charles is not in Rome and Jack is in Paris;
  - Charles is in Rome, Katy is not in Florence, Jack is not in Paris;
  - ...
- "Charles incoherent" answers contain only statements like: "Charles can't be everywhere ..." or "Charles location is contradictory ..."
- "Premises repeated" means that the answer just repeats the premises.

As we expected, finding the right answer was easier under treatment 3B. However, the most interesting result concerns the distribution of incorrect answers. Our hypothesis is that under 3B more individuals should generate a Charles context, perceiveing Charles as incoherently located, while under treatment 3A more subjects should edit only contexts corresponding to the *a)* and *b)* statements, failing to see how Charles' location constrains the space of possible states of affairs. Subjects' answers support this hypothesis. In particular, the Incorrect answer excluding Charles, which is the modal answer under treatment 3B, clearly reflects the fact that, just like in 2B, the location of Charles has been perceived as incoherent (as often stressed by the remarks accompanying the answers) and thus Charles has been canceled from the conclusions. However, individuals have not given up after recognizing a contraddiction (this is a peculiar feature of local reasoning!). Two dual reasoning processes have been then activated: one taking both Charles locations as false as if the contraddiction had falsified both - and deducing that Jack is in Paris and Katy is in Florence; the other one taking them as simultaneously true, and deducing that Jack is not in Paris and Katy is not in Florence. On the converse, the modal answer in treatment 3A is the Incorrect answer including Charles, which doesn't seem to reflect the perception of an incoherence between Charles locations, but instead reflects difficulties in keeping track of the compatibility relations between contexts, leading to incompatible statements.

MODELLING DISJUNCTIVE REASONING IN THE TLMM

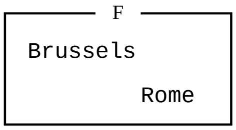
In this section we show how LMS can be used for modelling some relevant features of the experimental results presented in the former section. The goal is not to present the technical details of the formalization, but to provide an illustration of how LMS can be used as a formalization of the TLMM[4].

Modelling Experiment 1

In the discussion of Experiment 1, we suggested that Treatment A is harder than Treatment B because the disjuncts in the first premise are not interpreted in the same context. We suggested that the two proper names mentioned in treatment A, Mary and Gino, initialize two separate contexts, whereas in treatment B only one context is initialized, as only one agent, Francesco, is named. Here is how it works in LMS.

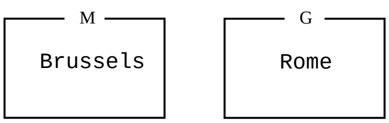
[4]For a general discussion on the use of LMS as a formal framework for mental representation, see (Giunchiglia & Bouquet 1998).

In Treatment B, corresponding to the first premise, a contexts is initialized, that is the context associated with Francesco (F):

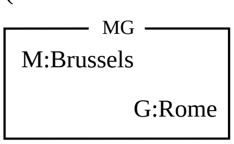


This context contains two LMMs, `Brussels` and `Rome`, one for each possible location of Francesco (we use the graphic conventions of (Johnson-Laird & Byrne 1991) for disjunctive models). When the categorical premise is added, it is easily integrated within F (after all, it is still about Francesco) and the model `Rome` is discharged. As a consequence, the conclusion that Francesco is in Brussels is easily derived in F. The reasoning associated with treatment B is a simple form of local reasoning within F. In this case, the source context is used as target context as well, and thus there is no need for any integration of models across different contexts.

In treatment A, mental representation is quite different. The names 'Mary' and 'Gino' trigger the initialization of two input contexts, M and G, each containing a single model representing the location of Mary and Gino:



The two models are LMMs, since each of them represents a portion of the information provided in the first premise. When the categorical premise is added (i.e. "Gino is not in Rome"), local reasoning in M or in G is not sufficient to locally derive the deductive conclusion we would expect. Indeed, adding the corresponding model ¬`Rome` to M does not allow to conclude anything about Mary's location (this corresponds to the answer that Mary can be in Brussels or anywhere else); and adding the model to G produces an inconsistent context. Only the construction of a more comprehensive target context, MG, allows the subjects to solve the problem. And building the target context requires for the subjects to integrate LMM from M and G in such a way that the intended relation (i.e. mutual exclusion) is represented[5]:



Failures are explained as the inability of constructing MG in the right way. In other words, subjects who fail the task do not see how the information locally represented in the two contexts M and G can be integrated in a single picture.

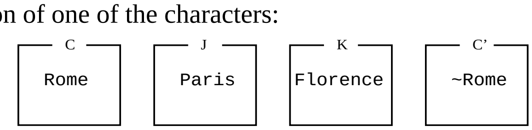
The target context MG is very similar to the single context F of Treatment B. However, MG is the result of integrating two local representations using the suitable compatibility relation, whereas F is "given" from the beginning. The experimental results show that Treatment A is intrinsically harder than Treatment B, and support the two hypotheses underlying TLMM. In particular they show that organizations of contexts that reduce the number (or complexity) of constraints between LMMs correspondingly reduce the difficulty of the task.

Modelling Experiments 2 and 3

Reasoning about double disjunctions is an extension of reasoning about a single disjunction, as it involves joining together two disjunctions. However, the tricky part of the "double disjunction task" is that in the first premise there is a proposition that is inconsistent with a proposition in the second premise (e.g. "Mary is in London" and "Mary is in Mallorca" in Experiment 2, "Charles is in Rome" and "Charles is not in Rome" in Experiment 3). This means that there is a compatibility relation between the local models of the two premises, and that reasoners must take it into account in order to make the right deduction.

Though the task assigned in the two experiments looks very similar, there are important differences. In Experiment 3, the contradiction between "Charles is in Rome" and "Charles is not in Rome" is simply a matter of logic, as one proposition in the negation of the other. In Experiment 2 this is not the case: "Mary is in London" and "Mary is in Mallorca" are not logically contradictory, and a contradiction can be derived only when the common sense fact that no one can be in two different places at the same time is added. This means that Experiment 2 involves further reasoning which is not required in Experiment 3. Therefore, we first provide a model of Experiment 3, and then extend it to consider the common sense knowledge about people's location.

When presented with the premises of Treatment 3A, subjects construct four contexts, each representing the location of one of the characters:



Notice that Charles' location is represented twice. This is because in Treatment 3A the two occurrences of Charles are quite apart, and therefore some subjects may overlook the fact that Charles is the same individual in the two premises (see discussion of Experiment 2, thinkink aloud protocols, when a subject commented that "[he] didn't see that Mary was Mary even here"). Subjects that get stuck in this representation will con-

[5] In the target context MG models are denoted by adding the name of the context from which they are "imported". This convention, used throughout the rest of the paper, has technical reasons. Here it is used to distinguish local models in different contexts.

struct a target context where no contradiction arises and therefore no interesting conclusions follow.

Let us now consider the case of subjects who recognize that Charles is the same individual in both premises. Following the model for single disjunctions, they will construct two contexts where the exclusive character of each premise is represented. The result is the following (P1 and P2 stand for premise a) and b), respectively):

```
P1                  P2
C:Rome              K:Florence
        J:Paris             C:~Rome
```

These two contexts correspond to the stage when subjects start reasoning separately about the two premises. Some subjects, in fact, do not go beyond this stage, and the only conclusions they draw are derived by local reasoning within the two contexts P1 and P2.

The next step is the integration of the two disjunctions into a comprehensive and coherent target context (TC). TC will contain a set of models each accounting for the location of the three characters, while respecting the logical relation between LMMs of the two input contexts. Intuitively, this is modelled in LMS by defining a compatibility relation between LMMs of three contexts: the input contexts P1 and P2, and the target context TC. Let $\mathtt{m}_1$, $\mathtt{m}_1$ and $\mathtt{m}_{TC}$ be LMMs of P1, P2, and TC respectively. Then they are compatible (written $\mathcal{R}(\mathtt{m}_1,\mathtt{m}_2,\mathtt{m}_{TC})$) if and only if:

1. $\mathtt{m}_{TC} = \mathtt{m}_1 + \mathtt{m}_2$;
2. it is not the case that:

$$\mathtt{m}_1 \models_{\text{P1}} P \text{ and } \mathtt{m}_2 \models_{\text{P2}} \neg P \tag{1}$$

Intuitively, the first conditions expresses the fact that $\mathtt{m}_{TC}$ puts together the information associated with the two LMMs $\mathtt{m}_1$ and $\mathtt{m}_2$. The second says: do not pick up models from P1 and P2 that express inconsistent information (e.g. a model that locally satisfies `Rome` in P1 and ¬`Rome` in P2). Models such as `C:Rome`–`J:Paris` or `K:Flo`–`C:`¬`Rome` are obviously discharged by local reasoning in P1 and P2, respectively.

Starting from the information in P1 and P2, and applying the two constraints above, the admissible models in TC will be the following (Florence is abbreviated to 'Flo'):

```
TC
C:Rome   K:Flo
J:Paris  C:~Rome
J:Paris  K:Flo
```

Each line in the figure corresponds to a situation compatible with the premises. For example, the model `C:Rome`–`K:Flo` is obtained by merging the first model in P1 and P2 respectively, and does not violate the second constraint in the compatibility relation. The three pairs of LMMs in TC correspond to the complete solution ("Correct" in Table 3). Notice that Charles' contradictory locations does not seem to play any particular rôle in the reasoning process. The models `C:Rome` and `C:`¬`Rome` are combined with all the other models, and – as we pointed out – errors seem to reflect the difficulties in keeping track of the compatibility relations. This explains the high rate of "Incorrect, including Charles" answers in Treatment 3A.

In Treatment 3B, the salience of Charles in both premises seems to produce a sort of false start, where the incoherent context C (representing Charles' location) is constructed together with J (representing Jack's location) and K (Katy's location):

```
C                 J        K
Rome   ~Rome      Paris    Florence
```

At this point, many subjets get stuck with this initial representation and disregard information about Charles. The result is the following pair of contexts:

```
P1          P2
J:Paris     K:Florence
```

This models the high rate of "Incorrect, excluding Charles" answers in Treatment B. Information about Charles' location is perceived as contradictory and therefore excluded from the deductive process.

After puzzling for a while, other subjects grasp the disjunctive structure of the task and realize that Charles' location is mutually exclusive. They construct the following contexts:

```
C                P2                P1
C:Rome           K:Florence        C:Rome
     C:~Rome          C:~Rome          J:Paris
```

Context C, in itself, is not very informative. Indeed, it corresponds to the representation of the tautology $A \vee \neg A$ (for any proposition A). However, it has a significant effect on the task at hand: it makes useless the second semantic constraint which was used in Treatment 3A. Indeed, subjects only need to "sum" local models from the input contexts, whereas the second compatibility constraints is replaced by local reasoning within C, and this makes the problem easier to solve. The result is the same target context as that of Treatment 3A, where the right solution can finally be found.

Briefly, let us consider now Experiment 2. In Treatment 2B, where Mary is made salient, subjects construct a representation very similar to the initial representation of Treatment 3B:

```
M                   J        C
London  Mallorca    Paris    Rome
```

However, unlike Treatment 3B, there's no immediate contradiction between Mary's two locations. As a consequence, Mary's location cannot be (mistakenly) ruled out as in Treatment 3B. A possible reaction is trying to make sense of the context M (there are two Marys, Mallorca is a quarter of London). However, sooner or later, many subjects realize that the two locations can be represented as alternative possibilities (something similar to Treatment 3B), and find two solutions analogous to those of 3B. However, there are two important differences with Treatment 3B:

- first, the correct representation is not simply the result of applying a logical principle (*tertium non datum*), but is an instance of a common sense general law (two persons cannot be in two different places at the same time);
- second, as a consequence of the different nature of principles involved, in Treatment 2B we have a third model in the target context, namely the model where Mary is neither in London nor in Mallorca. This solution requires that subjects flesh out the local models in context M to consider the model where Mary is in neither locations (i.e. $\neg$`London` and $\neg$`Mallorca`).

FINAL REMARKS

We believe our experimental and modelling results encourage further research on how the local organization of mental models may affect deductive reasoning. Our own research plans go in two directions. On the modelling side, we are trying to model the representational asymmetries arising from transformations of disjunctions in conditionals during reasoning. On the empirical side, we plan to extend our experiments to syllogistic and metadeductive reasoning. We also believe that the understanding of the local organization of representations can provide interesting insights for extending the theory of mental models to multi-agents domains, in particular to games.

ACKNOWLEDGMENTS

We thank Phil Johnson-Laird, Paolo Legrenzi, Massimo Benerecetti, Chiara Ghidini, Luciano Serafini and the participants at seminars at the Rovereto LSC for helpful comments during the development of our research. Financial support by the Rovereto LSC is also gratefully acknowledged.

REFERENCES

Dinsmore, J. (1991). *Partitioned Representations*, Kluwer Academic Publisher.

Doherty, P. (1996). *Partiality Modality and Nonmonotonicity*, CSLI Publications.

Evans, J. & Newstead, S. (1980). A study of disjunctive reasoning, *Psychological research* **41**: 373–388.

Fauconnier, G. (1985). *Mental Spaces: aspects of meaning construction in natural language*, MIT Press.

Fauconnier, G. (1997). *Mappings in thought and language*, Cambridge University Press.

Fenstad, J. (1997). Partiality, *in* J. van Benthem & A. ter Meulen (eds), *Handbook of Logic and Language*, Elsevier Science B.V., chapter 11, pp. 649–682.

Giunchiglia, F. (1993). Contextual reasoning, *Epistemologia, special issue on I Linguaggi e le Macchine* **XVI**: 345–364. Short version in Proceedings IJCAI'93 Workshop on Using Knowledge in its Context, Chambery, France, 1993, pp. 39–49. Also IRST-Technical Report 9211-20, IRST, Trento, Italy.

Giunchiglia, F. & Bouquet, P. (1998). A Context-Based Framework for Mental Representation, *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society - CogSci'98*, Cognitive Science Society, Madison, Wisconsin (USA).

Giunchiglia, F. & Ghidini, C. (1998). Local Models Semantics, or Contextual Reasoning = Locality + Compatibility, *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, Morgan Kaufmann, Trento, pp. 282–289.

Guha, R. (1991). Contexts: a Formalization and some Applications, *Technical Report ACT-CYC-423-91*, MCC, Austin, Texas.

Johnson-Laird, P. (1983). *Mental Models*, Cambridge University Press.

Johnson-Laird, P. & Byrne, R. (1991). *Deduction*, Essays in Cognitive Psychology, Laurence Erlbaum Associates, Publishers.

Johnson-Laird, P., Byrne, R. & Schaeken, W. (1992). Propositional reasoning by model, *Psychological Review* **99**(3): 418–439.

McCarthy, J. (1993). Notes on Formalizing Context, *Proc. of the 13th International Joint Conference on Artificial Intelligence*, Chambery, France, pp. 555–560.

Shafir, E. (1993). Thinking and deciding through disjuntions, *Cognition* .

Thaler, R. (1985). Mental accounting and consumer choice, *Marketing Science* **4**: 199–214.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psyhology of choice, *Science* **211**: 453–458.

Warglien, M. & Bouquet, P. (1999). Experimental protocols. University of Trento (Italy).