Mind & Society (2008) 7:1–19
DOI 10.1007/s11299-007-0029-3

ORIGINAL ARTICLE

The role of cognitive and socio-cognitive conflict in learning to reason

Katiuscia Sacco · Monica Bucciarelli

Received: 28 February 2005 / Accepted: 26 April 2006 / Published online: 24 March 2007

Abstract The mental model theory claims that the ability to falsify is at the core of human rationality. We assume that cognitive conflicts (CCs) and socio-cognitive conflicts (SCCs) induce falsification, and thus improve syllogistic reasoning performance. Our first study assesses adults' ability to reason in two different conditions in a single experimental session. In both conditions the participants are presented with conclusions alternative to their own. In the CC condition they are told that these conclusions are casual, in the SCC condition they are told they have been produced by another individual. The second study is analogous to the first, with the exception that the participants deal with the two conditions in two different experimental sessions. The overall results reveal that falsification is enhanced by conflicts experienced at the cognitive level. The results also reveal that learning to reason occurs in adults, when tested in two distinct experimental periods.

Keywords Reasoning · Falsification · Socio-cognitive conflict · Learning · Mental model theory

1 Introduction

A long and consolidated tradition, both philosophical and psychological, identifies the heart of human rationality in the ability to draw deductively valid inferences. The surprising finding is that people often fail in reasoning tasks, while in everyday life, they prove to be able to reason efficiently. This

K. Sacco (✉) · M. Bucciarelli
Centro di Scienza Cognitiva and Dipartimento di Psicologia,
Università di Torino, via Po 14, 10123 Torino, Italy
e-mail: sacco@psych.unito.it

contradiction was emphasized by Evans and Over (1996), who opened an ongoing debate about human rationality and by Stanovich and West (2000). The questions underlying the debate are: if humans are indeed rational, why might they fail in inferring logically correct conclusions in reasoning tasks?

On one hand, non-logic approaches suggest that people are not rational and that they are systematically biased by what catches their attention most in a given situation, neglecting the 'logically salient' data (for a review, see Evans 1982).

On the other hand, there are scientists who strongly believe in human rationality. Pragmatic rule theorists, for example, believe that people act according to previous experience: this provides them with a set of domain-dependent rules applicable to the current situation by analogy (Griggs 1983) or with specific schemas for sets of situations (Cheng and Holyoak 1985). However, pragmatic rules only seem to give a description or, at the best, a post hoc explanation of human cognition. The processes which lead people to succeed or fail in reasoning tasks appear to be out of their scope. Other theorists propose that individuals use abstract rules of inference similar to logic calculus (see Braine 1998; Rips 1994). These logical systems, when correctly applied to the premises of an argument, lead to the normatively correct conclusion. However, for these approaches it is not easy to explain the errors people make in reasoning tasks.

The mental model theory provides a plausible answer to the question of rationality (Johnson-Laird 1983; Johnson-Laird and Byrne 1991). According to this theory, people can mentally construct models of a state of affairs and manipulate such models in order to make inferences. According to this view, people are rational in theory, in that they are able to falsify their own putative conclusions, however they tend to fail in practice; they possess the basic competence to be rational, but they fail in the executive phase when rationality must be put into action. Given a sufficiently long time, strong motivation and an efficient working memory, they would manage to reach valid conclusions. The theory is able to predict and explain people's performances in all kinds of deduction, and in our opinion it is currently the best explanation of human reasoning. However, we think our comprehension of reasoning by models would benefit from a deeper investigation into the functioning of the falsification process. Bucciarelli (2007) claims that cognitive conflicts (CCs) and socio-cognitive conflicts (SCCs) favor learning to reason, as they make the subject falsify. She grounds her assumption on the following findings.

First, Bruner et al. (1966) find evidence in favor of the claim that the main mechanism of development is the conflict between different modalities of representation; when two systems of representation do not correspond, the child performs a clear-cut revision of his way of solving problems. Therefore, in giving an account of the development of children's ability to reason, they introduce the term 'conflict'. For example, a CC in the physical domain is a discrepancy between a child's expectations and an empirical outcome that contradicts those expectations; the confrontation of a subject's point of view with contradictory visual feedback would constitute an example of this category (see, e.g., Levin et al. 1990). As regards the SCC, from a developmental point of view, different theorists underline the importance of social interactions in cognitive development, by claiming that the first stages of thinking development are deeply social, since individual abilities become structured thanks to interaction with others (see, e.g., Vygotskij 1978). Doise and Mugny (1984) empirically found that the experience of the SCC assists children to reach the correct solutions to certain problems. In particular, when working alone, children at a certain stage in their development are unable to perform certain tasks because they cannot decenter, i.e., they focus on certain aspects of problems and exclude other, more important aspects; however, when exposed to differing points of view with regard to such problems, the same children find it easier to decenter. It follows that social interactions should provide an obvious forum for the possibility of exchanging perspectives.

Studies on the SCC follow two main approaches (Druyan 2001): they either focus on interactions between peers with different perspectives (see, e.g., Ames and Murray 1982; Druyan and Levin 1996; Perret-Clermont 1980), or on interactions between individuals of different status, such as an adult (e.g., a teacher) and a child, or an expert and a layman (see, e.g., Saljo and Wyndhamm 1990). In our study, we deal with SCCs of the first type, where a subject is pressed to compare his/her perspective with that of another individual having a comparable social status. 'This need to verify one's own perspective in coordination with other perspectives structures the process of interpersonal negotiations in ways that can promote cognitive growth' (Bearison 1986, p 136). On the other hand, in interactions of the second type, knowledge is imparted in a linear way, as one partner is the authority and the source of correct knowledge, which must be conveyed to the partner.

In the literature, SCCs have mainly been used to induce children to solve conservation tasks. Doise and Mugny (1984), for example, added social interaction to a liquid conservation task by saying that two glasses of juice were a prize for two equally good children, thus introducing the social norm that the two children have the right to receive the same prize. The experimental subjects, who at first did not realize that the two glasses contained the same quantity of liquid, showed an understanding of the rules of conservation thanks to the social context. Siegler (1995) also obtained improved performance when asking the child to explain the point of view of the individual giving the correct answer to a conservation task.

Consistent with the above-mentioned literature, and in line with Bucciarelli (2007), we argue that humans' ability to assume different points of view is central to reasoning (also see Bucciarelli and Johnson-Laird 2001), and we aim to study possible different effects of CCs and SCCs on the ability to reason. Generally speaking, we assume that there is a strict relationship between mindreading and the falsification phase postulated by the model theory. Indeed, mindreading presupposes the assumption of different points of view, an ability which can be triggered by the experience of a conflict between one's own and other persons' beliefs. Such an ability plays an important role in the falsification phase of a reasoning process, during which temporarily abandoning the conclusion that has been reached, in favor of a different one, is crucial. Some evidence of the relationship between mindreading, falsification and SCC is presented in the next paragraph.

2 Falsification, socio-cognitive conflict and mindreading

Experimental data reveal that the main difficulty people encounter in making inferences from a given set of premises concerns the search for alternative representations of the problem; in terms of the mental model theory, people usually fail to falsify the conclusion deriving from the first model constructed. In the falsification phase, reasoners search for alternative models that might falsify the putative conclusion they previously reached. If there are no alternative models, then the putative conclusion is valid; otherwise, reasoners must find a conclusion which includes all the models of the premises. According to the mental model theory, people may fail to make inferences when they stop at the putative conclusion and do not search for alternative models. In those cases, if there are alternative models, the putative conclusion they reached cannot be said to be valid, i.e., logically correct. Thus, failure to search for alternative models might greatly affect reasoning with multiple-model problems. Indirect evidence that people tend not to falsify within the syllogistic domain is that they perform better and faster in tasks requiring the construction of only one model than in tasks requiring two or more models (Bucciarelli and Johnson-Laird 1999); moreover, invalid conclusions very often correspond to those which can be inferred from the first model constructed (see Bara et al. 1995, 2001). Bucciarelli and Johnson-Laird (1999) also find that, when dealing with syllogisms, individuals not trained in logic may search for alternative models if forced to do so, though they are less likely to search for counterexamples if the task does not explicitly request them to do so. However, at least within the domain of reasoning with non-standard quantifiers, Neth and Johnson-Laird (1999) find that individuals seem to search for counterexamples spontaneously. The participants in their study were presented with syllogistic premises involving the non-standard quantifier «more than half»; each series of premises was followed by a putative conclusion, and participants were asked if the conclusion was valid or not. Most of the participants found counterexamples to invalid inferences. The same authors admit, however, that the domain they investigated could elicit the search for alternative models more than other domains. All this considered, it would be interesting to find out which factors or circumstances tend to elicit falsification.

Bucciarelli and Johnson-Laird (2001) argue that deontic materials make the false instances of a rule more noticeable by making other people's mental states, such as beliefs and expectations, more noticeable. In particular, the authors investigate the contexts that facilitate young children in a RAST (Wason 1968). The task requires the reasoner to choose whether to inspect potentially confirming or violating cases when testing a rule. An example of the rules used to test a factual rule is 'All the squeaky mice are in the house', and an example of the rules used to test a deontic rule is 'All the squeaky mice must stay in the house' (Cummins 1996: Experiment 2). In the experiment by Bucciarelli and Johnson-Laird, 3-, 4- and 5-year-old children deal with factual and deontic versions of the task, both in a context provided by the possibility of a false belief (mental state protocol), and in the standard context where only the deontic task invites reasoners to consider the others' mental states (standard protocol). The results concerning the standard protocol show that 3-year-olds perform better in the deontic version than in the factual version of the task (also see Cummins 1996). However, the results concerning the mental state protocol show no difference in performance by the three groups of children in the factual and the deontic tasks. In line with the authors' expectations, in this protocol the deontic version has no advantage over the factual one, and this result also holds for the youngest group of participants. This result is consistent with the assumption that the need to take into account possible false beliefs of others might induce a person to envisage the false instances of a rule easily. Besides, this explanation is consistent with the facilitatory effect sorted by deontic materials in the standard RAST. Bucciarelli and Johnson-Laird claim that in their experiment they induced reasoners to use their mindreading ability (see below) to conceive possible alternative realities. As Bucciarelli (2007) claims, attributing beliefs to other agents may involve a SCC, which is also involved in mindreading tasks in general.

Mindreading is the ability to (i) acknowledge that human beings possess mental states, (ii) attribute mental states, such as beliefs and desires, to oneself and to others, (iii) recognize that the mental states of others do not necessarily correspond to one's own, (iv) understand that mental states determine external behavior, such as decisions and actions, (v) predict, describe and explain behavior on the basis of such mental states (Premack and Woodruff 1978; Leslie 1987; Baron-Cohen 1994).

In line with Bucciarelli and Johnson-Laird (2001) and Bucciarelli (2007), we believe that a crucial ability in falsification is the assumption of another person's point of view, which is part of our ability to mindread. The ability to mindread enables us to understand that not everyone's beliefs are the same and that there are often other possibilities which one has not yet explored. Since mindreading allows for the assumption of an alternative point of view, it produces a SCC that might play a role in the ability to falsify one's own conclusion in reasoning tasks. Bucciarelli and Johnson-Laird (2001) find that the need to take into account the other person's point of view induces a person to falsify in a testing hypothesis task. Our aim is to explore whether and how the need to take into account the other person's point of view induces a person to falsify in a deductive reasoning task.

We used the mental model theory as our framework; thus, we assume that reasoning involves the construction and manipulation of the mental models of the premises, and that incorrect conclusions are mostly due to the fact that people tend to accept the first model constructed without attempting to falsify it. Table 1 exemplifies the phases involved in reasoning through models with both one-model and multiple-model valid and invalid syllogisms.

In the construction phase, each linguistic premise is translated into an analogical representation, i.e., a mental model. In the integration phase all the models of the premises are integrated into a single mental model by overlapping their identical 'tokens'. In the conclusion phase reasoners extract the relevant information from the model produced in the integration phase and formulate a putative conclusion. In the falsification phase, reasoners [should] attempt to falsify the putative conclusion by searching for alternative integrated models which are inconsistent with the conclusion: if the search fails, then the original conclusion is valid; if the search is successful, the reasoner formulates a new conclusion which is consistent with all the models produced; if it is impossible to formulate a conclusion, the reasoner claims that there is 'no valid conclusion' (nvc).

Contrary to the assumptions of model theory, some studies support the claim that people do not normally search for counterexamples at all in syllogistic reasoning (see, e.g., Klauer et al. 2000; Newstead et al. 1999). Evans et al. (1999), for example, investigate the predictions of model theory for syllogisms in evaluation tasks. In their study, participants are asked to evaluate whether a given conclusion is necessary and, in addition, to state whether that conclusion is possible. The results confirm three main predictions based on the mental model theory. However, they also reveal an unexpected effect: two sorts of problems supporting possible conclusions. Some (retrospectively identified as 'possible strong') are regularly taken to imply necessary conclusions, whereas others (retrospectively identified as 'possible weak') are

Table 1 The mental processes involved in reasoning with 'All A are B. All B are C' (one-model syllogism) and 'Some A are B. No B are C' (multiple-model syllogism)

Phases	Construction		Integration			Conclusion	Falsification		
One-model									
First premise									
All artists	[a]	b	[a]	[b]	c	All artists	No alternative		
are bandits	[a]	b	[a]	[b]	c	are chefs	model of the		
							premises falsifies		
							the conclusion		
Second premise									
All bandits	[b]	c							
are chefs	[b]	c							
Multiple-model									
First premise									
Some artists	a	[b]							
are bandits	a								
		[b]							
Second premise									
No bandits	[b]	–c	a	[b]	–c	No artists	a		[c]
are chefs	[b]	–c	a			are chefs	a	[b]	–c
		[c]		[b]	–c			[b]	–c
		[c]			[c]		a		[c]
					[c]		Some artists are		
							not chefs		

rarely taken to imply necessary conclusions and indeed are often not even taken to imply possible conclusions. The authors suggest that this phenomenon can be interpreted by assuming that in evaluation tasks individuals only tend to consider a single model, and possible strong conclusions are those supported by the first model that occurs in their mind. On the basis of this hypothesis, people are willing to endorse any conclusion supported by the first model of the premise they have thought of. Thus, the authors conclude, falsification plays no role in reasoning with syllogisms in evaluation tasks.

The idea that in evaluation tasks people construct a single model of the premises constitutes the core of the Selective Processing Model of belief bias, a revised version of the mental model theory for syllogistic reasoning in evaluation tasks, proposed by Evans et al. (2001) in a subsequent study. The Selective Processing Model acknowledges that the conclusion to be evaluated has a profound influence on the evaluation process: when the conclusion is believable or neutral, reasoners attempt to construct a model that supports it; when the conclusion is unbelievable, they attempt to construct a model that refutes it. The use of falsification in evaluation tasks is still a matter of debate (C. Becchio et al. , unpublished data, obtained experimental results that support the claim that falsification plays a relevant role in both evaluation and production tasks). In our opinion, much depends on the experimental procedures adopted. The tasks we used in the experiments presented here are not simple evaluation tasks: participants are first asked to produce their own conclusion and then to revise their reasoning process comparing another conclusion with their own. By asking for a comparison between two different conclusions, we aim to force reasoners to falsify the conclusion they spontaneously produced.

Experiments 1 and 2 investigate the role of CC and SCC in syllogistic reasoning. Experiment 2 is a replication of Experiment 1, with the only difference that the second experimental session occurs 1 week after the first, rather than following on immediately. The rationale for this procedure is that the results of the first experiment were not in line with the literature on reasoning, according to which individuals develop strategies to deal more efficiently with deductive problems, both within the same experimental session and between consecutive sessions (see, for a review, Schaeken et al. 2000). As strategies are thought processes that, besides being systematic, goal directed and under explicit conscious control, are elaborated in time (Evans 2000), in the second experiment we investigated whether a time lag of 1 week improved individuals' performance in the second experimental session compared to the first.

3 Experiment 1

In the present experiment participants deal with syllogisms; for each syllogism, once the participant has stated her own conclusion the experimenter presents an alternative conclusion (the correct one if the participant's conclusion is incorrect, an incorrect one if the participant's conclusion is correct). The participant is then asked to choose the best of the two conclusions. Each participant is tested on two protocols: in the CC protocol the experimenter states that each alternative conclusion is drawn by lot; in the SCC protocol the experimenter states that each alternative conclusion has been reached by another participant. Note that, in the SCC protocol, we told the participant that the conclusion had been produced by '*another participant*' in order to underline that there was no difference in the social status/authority between the participant and the other person who produced the alternative answer. We proceeded in this way because we wanted to avoid the effects of authority. The core assumption underlying our experiments is that CC and SCC induce reasoners to revise their way of solving a problem, thus favoring an improvement in their reasoning performance.

The strategy for the revision process varies depending on the individual's cognitive resources, and the context within which reasoning occurs, i.e., CC or SCC. In particular, we expect adults to have the cognitive resources necessary to attempt a falsification of their conclusion. We distinguish between two sorts of possible falsification strategies: *falsifying one's own* and *falsifying another person's*. The *falsifying one's own* strategy implies the search for counterexamples to one's own conclusion, without taking the alternative conclusion into account, i.e., it is limited to the model previously constructed by the participant. The *falsifying another person's* strategy involves considering both one's own conclusion and the alternative conclusion. The strategy involves mapping the model that supports the putative conclusion into the model that supports the alternative conclusion, taking into account the premises of the syllogism. Consider, for example, the multiple-model syllogism in Table 1: 'Some artists are bandits. No bandits are chefs'. According to our proposal, the strategy *falsifying another person's conclusion* operates as follows. Assume that a participant constructs the model:

a	[b]	–c
a	[b]	–c
		[c]
		[c]

and produces the (incorrect) conclusion 'No artists are chefs'. In our experiment, the participant would be presented with the alternative (correct) conclusion 'Some artists are not chefs'. Following the strategy, reasoners may proceed in one of the following ways. One consists of manipulating the original model to check whether the alternative conclusion also holds in such a model. Alternatively, reasoners attempt to construct a new model of the premises where the alternative conclusion holds, and then check whether this can also support their previous conclusion. In both cases, the strategy requires the reasoner to consider two integrated models of the premises, and to map one model into the other. Once the differences between the two models have been detected, the participant evaluates which conclusion, if any, they actually support. The crucial test for this prediction are multiple-model syllogisms, for which falsification is most relevant.

In our view, the use of *falsifying another person's conclusion* strategy is not plausible in the CC protocol, as it is not worth spending time and effort when the alternative conclusion is drawn by lot; instead of engaging in a difficult mapping process, participants will opt for simply revising their own conclusion (*falsifying one's own* strategy). On the contrary, in the SCC protocol, the fact that the reasoner is presented with a conclusion different to his/her own, one which has been drawn by another agent, can be a good way of motivating him/her to spend more time and to make a cognitive effort. Thus, we expect the CC protocol to induce the *falsifying one's own* strategy, and the SCC protocol to induce the *falsifying another person's* strategy.

To sum up, we hypothesize that adults follow the '*falsifying one's own*' strategy in the CC protocol and the '*falsifying another person's*' strategy in the SCC protocol; thus, in the CC protocol they do not take into account the alternative that is given, while in the SCC protocol they do. If we are correct, then we should see a bigger improvement in individuals' performance in the SCC protocol than in the CC one.

3.1 Method

3.1.1 Participants

We tested a sample of 20 adults (age 20–28). They were students of Psychology at Turin University. All of them took part in the experiment voluntarily and none of them had attended any courses in logic. All participants were female.

3.1.2 Design

We devised two protocols: the CC protocol and the SCC protocol. Each participant dealt with both protocols.

3.1.3 Materials

For each protocol we used 20 syllogisms (5 one-model; 15 multi-model). The syllogisms in the two protocols were the same but with different contents. All contents were as neutral as possible with respect to participants' beliefs. Here is an example: 'All wine-tasters are geometers. Some geometers are runners'.

In particular, for each protocol we used the following materials:

– twenty separate sheets of paper, each reporting a couple of premises, i.e., a syllogism;
– twenty separate sheets of paper, each reporting the valid conclusion for each syllogism;

– fifteen separate sheets of paper, each reporting an invalid conclusion for each of the multi-model syllogisms. In particular, for each multi-syllogism, of the invalid conclusions predicted by the mental model theory, we chose the one given by the majority of participants in the study by Bara et al. (1995). For one-model syllogisms—that require the construction of a single model and for which the model theory only predicts the correct conclusions—the experimenter presented the correct conclusion in any case (thus, one-model syllogisms cannot be considered proper control problems with respect to multiple-model syllogisms);
– a sheet of paper reporting all the possible types of conclusions: *all...are; some...are; no...are; some...are not; nvc.*
– two boxes

3.1.4 Procedure

The participants were tested individually in a quiet room, in a single session. Half of the participants dealt with the CC protocol first, then with the SCC protocol, vice versa for the other half. The syllogisms were presented in two different balanced orders. Participants were invited to think aloud.

Warm-up: In the warm-up participants were presented with two relational non-spatial tasks, such as 'Albert is taller than Mary; Mary is taller than John; what, if anything, follows?' and 'Juliet is richer than Luke; Charles is richer than Luke; what, if anything, follows?'. The participants were told to link the two terms, recurring only once in the premises, in order to reach a conclusion. The second problem was used to show that some problems have nvc. The experimenter then explained to the participant that the following problems would be a little different and that the conclusions could be of one of the following types (the experimenter showed the participant the sheet of paper reporting all the possible types of conclusions): *all...are; some...are; no...are; some...are not; nvc.*

CC protocol: The participant was presented with each pair of premises and asked to reach a conclusion. After she had reached her own conclusion, the experimenter wrote it down on a piece of paper visible to the participant, and then presented an alternative conclusion, introduced by the formula 'The conclusion we casually paired with this problem is...'. In actual fact, if the participant reached an invalid conclusion, then the experimenter presented the valid conclusion; vice versa, if the participant reached the valid conclusion, then the experimenter presented the invalid conclusion predicted by the mental model theory. The experimenter then asked the participant if she would exchange her own conclusion for the alternative that had been presented.

SCC protocol: The procedure was the same as in the CC protocol but the alternative conclusion was introduced by the formula 'The conclusion a previous experimental participant gave to this problem is...'.

3.1.5 Scoring

We assigned two different scores: Conc1 is the score for the first conclusion reached by the participant: 1 for the correct conclusion, 0 for the incorrect conclusion; Conc2 is the score for the participant's final conclusion: 1 if it is correct, 0 if it is incorrect. Thus, for example, if the participant drew a valid conclusion and chose to keep that conclusion after being presented with the alternative, she scored 1 for Conc1 and 1 for Conc2. If the participant drew an invalid conclusion and then preferred the alternative conclusion (which in the specific case is valid), she scored 0 for Conc1, and 1 for Conc2. Thus, we only considered that an improvement in performance occurred in the latter case.

3.1.6 Predictions

Improvement from first to second conclusion: We expected an improvement in participants' performance from the first to the second conclusion on multi-model syllogisms, due to the occurrence of a conflict at the cognitive level. We expected this to concern both the CC and the SCC protocol.

Improvement in CC as compared with improvement in SCC protocol: For multi-model syllogisms the facilitating effect of being presented with an alternative conclusion is greater in the SCC protocol than in the CC protocol. Indeed, the SCC protocol, which involves a SCC, was expected to elicit the *falsifying another persons'* strategy.

3.2 Results

The results are shown in Tables 2 and 3.

The order in which the two protocols were presented did not affect participants' performance, either in the CC protocol (65% of correct responses by participants dealing with the CC protocol in the first session vs. 66% of correct responses by those dealing with the CC protocol in the second session; *t*-test for unpaired samples: $t = -0.199$, $p = 0.844$), or in the SCC protocol (62% vs. 70% of correct responses; $t = -1.379$, $p = 0.185$); thus, the data were pooled together.

As expected, there was a statistically significant improvement in performance at the second conclusion level compared to the first conclusion level with multi-model syllogisms, both in the CC protocol (43% vs. 57% of correct responses, $t = 4.803$, $p < 0.0001$) and in the SCC protocol (49% vs. 60%,

Table 2 Percentage of correct responses with multiple-model syllogisms in CC and SCC protocols in Experiment 1

Syllogisms	CC protocol		SCC protocol	
	First conclusion	Second conclusion	First conclusion	Second conclusion
MM	43	57	49	60

Table 3 Percentage of responses with multiple-model syllogisms in Experiment 1 as a function of the four possible patterns of performances (i.e., both first and second conclusion correct, first conclusion correct and second conclusion incorrect, first conclusion incorrect and second conclusion correct, both first and second conclusion incorrect) and the type of protocol

Correct_1 and correct_2		Correct_1 and incorrect_2		Incorrect_1 and correct_2		Incorrect_1 and incorrect_2	
CC protocol	SCC protocol	CC protocol	SCC protocol	CC protocol	SCC protocol	CC protocol	SCC protocol
40	44	3	5	17	16	40	35

$t = 3.611$, $p = 0.002$). Also with one-model syllogisms, there was a statistically significant improvement in performance at the second conclusion level compared to that at the first conclusion level both in the CC (88% vs. 92%, $t = 2.179$, $p = 0.042$), and in the SCC (84% vs. 90%, $t = 2.854$, $p = 0.010$).

In order to test whether the increase in performance at the second conclusion level compared to that at the first conclusion level was greater in the SCC protocol than in the CC one, we computed an 'increase variable' (the difference between performance at the second conclusion level and that at the first conclusion level) for each protocol. Contrary to our expectations, for multi-model syllogisms, the comparison between the two increase variables indicated that the improvement was no greater in the SCC protocol than in the CC one ($t = 0.611$, $p = 0.548$). Also, with one-model syllogisms the increase in performance at the second conclusion level compared to that at the first conclusion level was no greater in the SCC protocol than in the CC one ($t = 1.748$, $p = 0.097$).

A content and statistical analysis of participants' comments, produced while solving the tasks, is not possible as such comments are too few. However, think-aloud protocols suggest the use of mental models in reasoning (see also Bucciarelli and Johnson-Laird 1999).

Some individuals seemingly construct mental images representing sets for the entities mentioned in the premises, often through the use of circles, similar to those of Euler. Here is a typical protocol. The premises of the syllogism were: All painters are tennis players; Some painters are photographers. P1 said: 'I imagine an element which indicates a category...in this case a painter with his palette, and I create a set where each painter also has a tennis racket. Then...as I have to divide the group, the painters become two...one of them has a camera and the other one has not...'. While saying that, P1 drew circles with her hands and made other circles inside these. Most participants described their reasoning in similar terms.

Some participants appeared to change their strategies during the experiment. As this paper focuses on the revision strategies adopted when faced with a CC vs. an SCC alternative, we will now report some participants' comments with respect to the differences between the two protocols. Participants' comments, made at the end of the experiment, seem to confirm both our assumptions about the facilitating effect of the presentation of an alternative conclusion and our assumptions about the strategies used during revision. As far as the general improvement in performance after the presentation of the alternative is concerned, in both protocols some participants made considerations of the following sort: ‘When I have to evaluate the alternative conclusion, I feel I am able to concentrate more. The alternative pushes me to reason better’. As far as the differences in the revision strategies used in the two protocols are concerned, it seems that, when presented with a CC alternative, participants tend not to consider the alternative in their revision; on the contrary, they compare their reasoning with that of the ‘other person’s’ in the SCC protocol. Consider for example the following comments: ‘...I wanted to understand how she (the other person) reached that answer, and so I checked which conclusion (mine or hers) was more correct... When the alternative was drawn by lot I thought less and I only changed if my reasoning was not convincing’ (P13), and ‘When the alternative was CC, I felt I was less in discussion’ (P15).

3.3 Conclusions

The results of the experiment reveal that the presentation of an alternative conclusion improves participants’ performance.

An unexpected result concerns the improvement in adults’ performance at the second conclusion level compared to that at the first conclusion level when dealing with one-model syllogisms. One-model syllogisms are quite easy to solve even for 9-year-olds (see Bara et al. 1995), therefore we did not expect to detect any improvement. As a matter of fact, adults’ performance with one-model syllogisms did not reach a ceiling effect at the first conclusion level, and both CC and SCC improved their performance at the second conclusion level. However, it has to be noticed that for one-model syllogisms participants were always presented with a correct conclusion.

The results of our first experiment, where participants dealt with the two protocols in a single session, did not reveal a learning effect from the first experimental protocol to the second experimental protocol. These results seem to contradict the literature on reasoning, according to which individuals develop strategies to deal more efficiently with deductive problems trial after trial. However, many studies have documented that learning improves more when there is a lag between one study trial for a particular item and the next study trial for that item (see, for example, Dempster 1996). In particular, the studies on the effects of time on the reiteration of new information concern the learning of declarative knowledge, rather than of procedural knowledge. Ebbinghaus (1985), cited in Schacter (1989), noticed that the distribution of the study sessions in time influences the consolidation of information in the long-term memory. Bahrick and Phelps (1987) observed that individuals tend to remember information longer when they acquire such information through distributed practice (i.e., several sessions spaced in time), rather than through massed practice (i.e., sessions accumulated in a single period): the longer the learning periods are spaced, the better individuals retain the information. In an in-depth study of the spacing effect, i.e., spaced learning, Glenberg (1977, 1979) found that it is associated with the process through which memories are consolidated in the long-term memory (also see Leicht and Overton 1987). The aim of our second experiment was to see whether a learning effect occurs when individuals are invited to deal with the two protocols in two different experimental sessions, the second experimental session occurring 1 week after the first. Thus, in Experiment 2, we arbitrarily choose to introduce a temporal interval of 1 week between the two experimental sessions. The aim is to verify whether a spacing effect occurs in case of learning to reason, which is a case of learning of procedural knowledge.

4 Experiment 2

In Experiment 2 we aimed to verify whether the results in Experiment 1 hold when adult participants deal with the two protocols in two different experimental sessions. The general predictions were the same as those for Experiment 1.

4.1 Method

4.1.1 Participants

We tested a sample of 20 adults (age 20–28). They were all female students of Psychology at Turin University, and took part in the experiment voluntarily. Participants had not attended any courses in logic.

4.1.2 Design, materials, procedure and scoring

The same as Experiment 1 with the only difference that the experiment was performed in two sessions, the second session occurring 1 week after the first. Half of the participants dealt with the CC protocol in the first session and with the SCC protocol in the second session, vice versa for the other half of the participants.

4.2 Results

A session effect emerged: the order of presentation of the two protocols affected participants' performance in the CC protocol (66% of correct responses for participants dealing with the CC protocol in the first session vs. 79% of correct responses for those dealing with the CC protocol in the second session; t-test for unpaired samples: $t = -2.626$, $p = 0.017$), but not in the SCC protocol (76% vs. 72% of correct responses; $t = 0.896$, $p = 0.382$).

The presence of this session effect prevents us from pooling the data and, consequently, from computing the statistics for the comparison between the CC and SCC protocols. However, we can analyze the predicted improvement

Table 4 Percentage of correct responses with multiple-model syllogisms in CC and SCC protocols (when participants dealt with them both in the first and second session) in Experiment 2

Protocols' order of presentation	CC protocol		SCC protocol	
	First conclusion	Second conclusion	First conclusion	Secnd conclusion
CC1st SCC2nd	41	51	45	53
SCC1st CC2nd	60	67.5	53	68

Table 5 Percentage of responses with multiple-model syllogisms in Experiment 2 as a function of the four possible patterns of performances (i.e. both first and second conclusion correct, first conclusion correct and second conclusion incorrect, first conclusion incorrect and second conclusion correct, both first and second conclusion incorrect) and the type of protocol

	Correct_1 and correct_2		Correct_1 and incorrect_2		Incorrect_1 and correct_2		Incorrect_1 and incorrect_2	
Protocols' order of presentation	CC protocol	SCC protocol	CC protocol	SCC protocol	CC protocol	SCC protocol	CC protocol	SCC protocol
CC1st SCC2nd	36	42	5	3	14	12	45	43
SCC1st CC2nd	57	47	3	6	11	21	29	26

from the first to the second conclusion level. The detailed results are in Tables 4 and 5.

As expected, there is a statistically significant improvement in performance at the second conclusion level compared to the first conclusion level for multi-model syllogisms, in both the CC and SCC protocols. In particular, we found a significant improvement with multi-model syllogisms when participants dealt with the CC protocol in the first session (41% vs. 51% of correct responses, $t = 3.676$, $p < 0.003$), and when they dealt with the CC protocol in the second session (60% vs. 67.5% of correct responses, $t = 2.246$, $p < 0.03$). We also found a significant improvement for multi-model syllogisms when participants dealt with the SCC protocol in the first session (53% vs. 68% of correct responses, $t = 3.705$, $p < 0.003$), but not when they dealt with the SCC protocol in the second session (45% vs. 53% of correct responses, $t = 3.179$, $p = 0.055$).

As expected, there was no improvement from the first to the second conclusion level for one-model syllogisms: this result holds for all conditions (CC protocol in first session: 90% vs. 96% of correct responses, $t = 1.964$, $p = 0.081$; CC protocol in second session: 98% vs. 100% of correct responses, $t = 1.000$, $p = 0.343$; SCC protocol in first session: 98% of correct responses both at the first and second conclusion levels; SCC protocol in second session: 98% vs. 100% of correct responses, $t = 1.000$, $p = 0.343$).

4.3 Conclusions

The adults participating in our second experiment performed better in the second session than in the first. This result suggests that learning occurred and that its effects become stable and detectable after 1 week. Thus, at least for adults, conflicts experienced at the cognitive level improve reasoning abilities provided that individuals are subject to a non-massed practice. Indeed, the learning effect, as detected through an improvement in performance from the first to the second experimental session, did not emerge in the first experiment.

A related finding is that the hypothesis concerning an improvement in performance after the presentation of an alternative conclusion is not fully confirmed. In particular, in both the CC and the SCC protocols we found a significant improvement with multi-model syllogisms in all sessions, an exception being the SCC protocol in the second session: evidently, the optimal learning process (to learn to falsify) consists of encountering first the alternative conclusion in an SCC context, then encountering the alternative conclusion in the CC context.

5 Discussion

When considered on the whole, the results of our experiments suggest that there is a relationship between conflicts experienced at the cognitive level and the ability to solve reasoning problems involving alternative representations. In particular, CC appears to play a role in improving deductive reasoning processes. Also, as regards SCC, the results suggest a relationship between the ability to read other people's mental states and the ability to reason. In the literature there are few studies on the effect of SCCs in adults, and they are mainly concerned with inductive reasoning (see Legrenzi et al. 1991; Butera et al. 2005). Our results are more general in that they reveal an effect of CC, along with an effect of SCC, in reasoning with a deductive task.

It can be argued that dealing with two conclusions reduces the probability of error or, also, that changes from correct to incorrect conclusions are less probable than those from incorrect to correct conclusions. Indeed, participants who spontaneously produced the correct conclusion probably have a more stable opinion than participants who initially produced an incorrect conclusion; people tend to endorse 'possible' conclusions, i.e., conclusions which are compatible with at least one model of the premises, and to refuse 'impossible' conclusions, i.e., conclusions which are incompatible with the premises (Evans et al. 1999). Thus, if our incorrect alternative conclusions were impossible conclusions, then the critique would hold. However, this was not the case: our alternative incorrect conclusions were possible, but not necessary conclusions selected from among those produced by the majority of people in previous experiments (Bara et al. 1995). Besides, on the other hand, there is evidence that, when erring, people tend to produce possible conclusions; thus, when faced with an alternative correct conclusion, participants again have to choose between two possible conclusions. These specifications are important to exclude that the improvements we found are due to the experimental procedure we adopted.

Another important issue concerns the way in which the effects of CCs and SCCs interact in promoting the ability to reason. The results of Experiment 2, where adults improve their performance when presented with the SCC protocol in the first session and with the CC in the second session, though not vice versa, would suggest that dealing with SCCs teaches us to appreciate and consider purely CCs. This can be accommodated by theories stating the importance of the kind of conclusion presented; for instance, in terms of the Selective Processing Model, which predicts the acceptance of believable conclusions and the rejection of unbelievable ones, it can be said that encountering of a believable alternative (SCC) leads to appreciate a simple alternative (pure CC). Moreover, our results indicate that such a learning effect occurs when participants go through a distributed practice of reasoning with syllogisms, rather than through a massed practice. This result is in line with the literature concerning the best way to organize time in order to learn efficiently.

As a general consideration, in this study we contravened the principle according to which one of the goals of cognitive science is to divide the mind into a number of different cognitive systems (Lyons 1999). Indeed, by exploring the relationships between reasoning and the SCC involved in mindreading, we intended to demonstrate that the study of deductive competence can benefit if it is accompanied by the study of other abilities with which it constitutes what is usually referred to as ‘rationality’. The mental model theory is currently the best explanation of human deductive competence, though we believe it should be extended and take into account the role played by SCC, since deductive reasoning also develops within social contexts.

Acknowledgments The authors wish to thank Bruno Bara, Cristina Becchio and Maurizio Tirassa for their insightful comments and suggestions on earlier versions of this paper. This work was supported by M.I.U.R. of Italy, PRIN project (research code: 2004111320_002).

References

Ames GJ, Murray FB (1982) When two wrongs make a right: promoting cognitive change by social conflict. Dev Psychol 18:894–897

Bahrick HP, Phelps E (1987) Retention of Spanish vocabulary over eight years. J Exp Psychol Learn Mem Cogn 13:344–349

Bara BG, Bucciarelli M, Johnson-Laird PN (1995) Development of syllogistic reasoning. Am J Psychol 2:157–193

Bara BG, Bucciarelli M, Lombardo V (2001) The mental model theory of deduction: a unified computational approach. Cognit Sci 25:839–901

Baron-Cohen S (1994) How to build a baby that can read minds: cognitive mechanism in mind reading. Cah Psychol Cognit 13:1–40

Bearison DJ (1986) Transactional cognition in context: new models of social understanding. In: Bearison DJ, Zimiles H (eds) Thought and emotion: development perspectives. Erlbaum, Hillsdale, pp 129–146

Braine MDS (1998) Steps towards a mental predicate logic. In: Braine MDS, O’Brien DP (eds) Mental logic. Lawrence Erlbaum Associates, Mahwah, pp 27–331

Bruner J, Oliver R, Greenfield P (1966) Studies in cognitive growth. New York, John Wiley & Sons

Bucciarelli M (2007) How the construction of mental models improves learning. Mind Soc 6(1) (in press)

Bucciarelli M, Johnson-Laird PN (1999) Strategies in syllogistic reasoning. Cognit Sci 23:247–303

Bucciarelli M, Johnson-Laird PN (2001) Falsification and the role of the theory of mind in the reduced array selection task. Curr Psychol Lett 4:7–22

Butera F, Caverni J-P, Rossi S (2005) Interaction with a high- versus low-competence influence source in inductive reasoning. J Soc Psychol 145:173–190

Cheng P, Holyoak KJ (1985) Pragmatic reasoning schemas. Cognit Psychol 17:391–416

Cummins DD (1996) Evidence of deontic reasoning in 3- and 4-year-old children. Mem Cognit 24:823–829

Dempster FN (1996) Distributing and managing the conditions of encoding and practice. In: Bjork EL, Bjork RA (eds) Memory. Academic, San Diego, pp 318–344

Doise W, Mugny G (1984) The social development of the intellect. Pergamon, Oxford

Druyan S (2001) A comparison of four types of cognitive conflict and their effect on cognitive development. Int J Behav Dev 25:226–236

Druyan S, Levin L (1996) The differential contribution of field-dependent-independent cognitive styles to socio-cognitive transaction in promoting scientific reasoning. Int J Behav Dev 19:831–850

Ebbinghaus H (1985) Uber das Gedachtnis. Duncker and Humblot, Leipzig

Evans JStBT (1982) The psychology of deductive reasoning. Routledge & Kegan Paul, London

Evans JStBT (2000) What could and could not be a strategy in reasoning. In: Schaeken W, De Vooght G, Vandierendonck A, d'Ydewalle G (eds) Deductive reasoning and strategies. Lawrence Erlbaum Associates, London, pp 1–22

Evans JStBT, Over DE (1996) Rationality and reasoning. Psychology Press, Hove

Evans JStBT, Handley SJ, Harper C, Johnson-Laird PN (1999) Reasoning about necessity and possibility: a test of the mental model theory of deduction. J Exp Psychol Learn Mem Cogn 25:1495–1513

Evans JStBT, Handley SJ, Harper C (2001) Necessity, possibility and belief: a study of syllogistic reasoning. Q J Exp Psychol 54A:935–958

Glenberg AM (1977) Influences of retrieval processes on the spacing effect in free recall. J Exp Psychol [Hum Learn] 3:282–294

Glenberg AM (1979) Component-levels theory of the effects of spacing of repetitions on recall and recognition. Mem Cognit 7:95–112

Griggs RA (1983) The role of problem content in the selection task and THOG problem. In: Evans JStBT (ed) Thinking and reasoning: psychological approaches. Routledge & Kegan Paul, London

Johnson-Laird PN (1983) Mental models: towards a cognitive science of language, and consciousness. Cambridge University Press, Cambridge

Johnson-Laird PN, Byrne RMJ (1991) Deduction. Lawrence Erlbaum Associates Ltd, London

Klauer KC, Musch J, Naumer B (2000) On belief bias in syllogistic reasoning. Psychol Rev 107:852–884

Legrenzi P, Butera F, Mugny G, Pérez JA (1991) Majority and minority influence in inductive reasoning. Eur J Soc Psychol 21:359–363

Leicht KL, Overton R (1987) Encoding variability and spacing repetitions. Am J Psychol 100:61–68

Leslie AM (1987) Pretense and representation: the origins of "Theory of mind." Psychol Rev 94:412–426

Levin I, Siegler RS, Druyan S (1990) Misconceptions about motion: development and training effects. Child Dev 61:1544–1557

Lyons JC (1999) Carving the mind and its (not necessarily modular) joints. In: Wilson RA, Keil F (eds) The MIT encyclopedia of the cognitive sciences. MIT, Cambridge

Neth H, Johnson-Laird PN (1999) The search for counterexamples in human reasoning. In: Hahn N, Stoness SC (eds) Proceedings of the twenty first annual conference of the Cognitive Science Society. Lawrence Erlbaum, Mahwah, p 806

Newstead SE, Handley SJ, Buck E (1999) Falsifying mental models: testing the predictions of theories of syllogistic reasoning. J Mem Lang 27:344–354

Perret-Clermont AN (1980) Social interaction and cognitive development in children. Academic, London

Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci 1:515–526
Rips LJ (1994) The psychology of proof: deductive reasoning in human thinking. MIT, Cambridge
Saljo R, Wyndhamm J (1990) Problem solving, academic performance and situated learning. A study of joint cognitive activity in the formal setting. Brit J Educ Psychol 60:245–254
Schacter DL (1989) Memory. In: Posner MI (ed) Foundations of cognitive science. MIT, Cambridge, pp 683–725
Schaeken W, De Vooght G, Vandierendonck A, d'Ydewalle G (2000) Deductive reasoning and strategies. Lawrence Erlbaum Associates, London
Siegler RS (1995) How does change occur: a microgenetic study of number conservation. Cognit Psychol 28:225–273
Stanovich KE, West RF (2000) Individual differences in reasoning: implications for the rationality debate? Behav Brain Sci 23:645–665
Vygotskij LS (1978) Mind in society: the development of higher psychological processes. Harvard University Press, Cambridge
Wason P (1968) Reasoning about a rule. Q J Exp Psychol 20:273–281