Main conclusions

The primary goal of our research was to bring a contribution to understanding obstacles and facilitators of decision making transfer. Moreover, we wished to compare our findings with the ones related to open-ended problem solving transfer in order to understand if they applied to a different type of problem, thus to what extent they could be generalized. Furthermore, from a practical point of view, the research was intended to provide some tentative indications that could be applied in real-life learning environments in order to support transfer. Finally, although it was not one of the goals of the research but rather a by-product, we were nevertheless able to better understand the working of web-based experimenting.

We will now bring together experimental results and discuss them trying to answer the research questions that we had raised in the beginning. We will pin-point our findings on binary decision making transfer and underscore differences and similarities with open-ended problem solving transfer, identifying, where necessary, the limits of the research. We will be able to provide answers to research questions and to draw implications regarding what might support transfer in training, but we will, in some cases, also explain why we think the finding requires further investigation. We will however, at our best, try to circumscribe the open-issues and suggest what type of analysis would help in resolving them.

Before discussing the results of our research, since we chose to compare them with the ones from open-ended problem solving, and, particularly, with those from Gick and Holyoak's (1980, 1983) experiments, it is important to underscore a few differences between the types of tasks involved and how they were conducted. Disagreements between our results and the ones from open ended problem solving transfer experiments should be interpreted in the light of these differences, which may also help explain them.

To begin with, in Gick and Holyoak's 1983 problem solving experiments on analogical transfer, the target task was to find possible solutions for an open-ended problem (the most frequently used was Duncker's *Radiation Problem*) after reading one or two analogous stories drawn from different domains. The only link between learning problems and target was their structural analogy since they were from distant domains and there were no shared superficial cues.

The task we selected to study transfer, a binary decision, has some fundamental differences with open ended problems. The first and most obvious difference is that binary decisions are very "closed" tasks (only two clearly stated options to choose from) while open-ended problem solving

tasks are at the other end of the spectrum (not even the problem space is defined, solutions can be searched for anywhere).

Because analogous binary decision making stories must lead to a final closed question which has only two possible answers, they tend to share a higher degree of similarity even when drawn from semantically distant domains and carefully formulated to avoid verbal cues. Of course, even with binary decision making tasks, there can be differences in the degree of similarity of the final question. In the case of the story we chose as our medium semantic distance case, *Lucrezia,* the task was a *switch-or-keep* choice, exactly like the learning stories. In our other target story, *Prisoners*, which we classified as having a high degree of semantic distance, the final decision consisted in deciding if the probability associated to two items (the prisoners) was the same or different, which is clearly less similar to the learning story switch-or-keep task. Of course, the switch-or-keep decision in the learning stories is made based on the probabilities of the items (e.g. the probability of each ball in the *3 Balls* story), but since this is not reflected in the formulation of the problem ("the host gives the contestant the choice between sticking to the ball originally selected or changing selection…") we believe that cues are eliminated or minimal.

Another difference between our task and the open ended problem solving tasks is that ours requires some operations to be retrieved from memory: insight from the previous problem and correct mapping are not enough to solve the target. Our task also requires to adapt what was learned in order to solve the target problem (e.g. to respond correctly in the *Lucrezia* transfer task, participants needed to invert the logic learned in the analogue because the goal was reversed from "finding" to "avoiding"). As a consequence, in our transfer tasks, once the analogy had been recognized, there was still the possibility of making mistakes, much more so than in the case of Duncker's *Radiation* problem.

Moreover, Gick and Holyoak's 1983 experiments are based on same context immediate transfer (few seconds after the learning task). We never adopted same context immediate transfer because we feared a ceiling effect: our shortest interval between learning and transfer was five minutes, filled with a creativity distracter task. Given that the forgetting curve is logarithmic with a steep decrease in remembering performance in the first few minutes and that a change in mindset was found to impact transfer in open-ended problem solving transfer, these differences should be kept in mind while comparing our results to Gick and Holyoak's. In fact, since in our experiments context was always changed (even if in some experiments the shift was low, obtained just by interposing a short distracter task), we need to also refer to Spencer and Weisberg (1986) and Catrambone and Holyoak (1989) studies which used open-ended problems as transfer tasks, but introduced context and time shifts.

In addition, in their experiment 4 (1983), Gick and Holyoak used analogical encoding (i.e. brief summarization of the analogues followed by a comparison in search of similarities between the two) as learning method. This choice was driven by the fact that, with their type of problem, analogical encoding facilitated transfer, otherwise difficult to obtain. In our experiments we made participants learn either by summarization, guided-comparison or free-comparison (which is very close to analogical encoding). If this choice allows us to base conclusions on a wider selection of learning methods, on the other hand, it makes our results not always straightforwardly comparable to theirs.

Lastly, Gick and Holyoak (1983) made participants compare two stories even in the one analogue condition (i.e. between two non-analogous stories) and we did not. We avoided following the same procedure because looking for analogies where there are none might result in confusion, as Gick and Holyoak (1983) themselves commented (p.25). Consequently, we slightly modified the procedure making our one-analogue condition participants summarize the analogues they received.

Research question # 1. What is the effect of learning one analogue on transfer in binary decision making tasks?

To answer the question if one analogue is enough to induce transfer and explicit recognition of the analogy, we ran nine experiments (three lab based and six online) which compared solution rates of participants learning one analogue (by summarizing it, except in the first experiment where the learning also implied answering some questions) with transfer rates of control group participants. We also ran a simple binary logistic regression on all the relevant cases.

One analogue was found to lead to significant transfer in all online experiments and the result was confirmed by the binary logistic regression. The result held true even in challenging circumstances (such as long delays coupled with medium context shifts), as found in our online experiments. This evidence, however, should be better qualified completing the picture with two other results. Firstly, in our lab experiments N. 1 and 2 (one week delay coupled with a high level of context shift) one analogue did not lead to significant transfer. Although the number of cases was low, this result hints possible difficulties in transfer in presence of high context shifts. This possibility is further reinforced by another finding, this time from our online experiments. Analyzing correctness of solution at boundary conditions (i.e. considering the subset of cases which were on the boundary, such as participants transferring to a highly distant problem at several days of delay) we found levels not different from control. Once again, unfortunately, we did not have enough cases to be conclusive.

Based on these results, we can be rather optimistic on transfer of binary decisions based on one analogue even in presence of in conditions which are far less restricted than those tested by Gick

and Holyoak (immediate transfer, no context shift of any type, including type-face and experimenter).

Differently from Spencer and Weisberg (1986) and from Catrambone and Holyoak (1989), who failed to obtain transfer when learning and target stories were presented after a delay and in different contexts, we were able to see transfer happen, even from one analogue, in context shifted and delayed situations. However, optimism should be mitigated by saying that we do not know what would happen stretching conditions even further. Our online cases hint that, as conditions are made more challenging, one analogue might not be enough to trigger transfer and our lab based cases (although insufficient for any conclusion to be based on them) suggest that one analogue does not allow transfer when context shift is high and delay is several days. We formulate the hypothesis that high context shift *coupled* with delay hinders transfer and that context is the main responsible of the effect. We also would expect, given the relevance that we found semantic distance to have, that solution rates would drop to control level in conditions of high semantic distance and medium context shift (which we have not tested).

In sum, in transfer of binary decision making, one analogue is enough to trigger un-hinted transfer. Moreover, delay coupled with context shifts up to medium and delay coupled with high semantic distance do not, in general, impair un-hinted transfer. In this sense, our results would indicate that, in binary decision making, transfer from one analogue is possible under more difficult conditions than were found for open-ended problems. A possible explanation for this difference is the already mentioned higher degree of similarity between learning and target tasks, which could serve as a cue. However, this is probably not a sufficient explanation since Catrambone and Holyoak (1989) found that, in open ended problem solving, an increase in verbal cues is not enough to induce un-hinted transfer a delay.

Future research should explore both what limits transfer from one analogue and what causes them (i.e. which part of the transfer process is failing). In order to investigate where the boundaries are, experiments could be run (with different delays) varying number of analogues (0,1), levels of semantic distance and level of context shifts between learning and transfer in a factorial design. If high context shift and high semantic distance are confirmed to impair transfer from one analogue, a further experiment could be performed in which the number of analogues is increased to two, in order to assess the effect on transfer of an increase from one analogue to two in boundary conditions. Quality of the learning output and explicit recognition of the analogy should be analyzed seeking correlation with successful transfer in the various combinations of semantic distance and delay because they may play a different role depending on transfer conditions.

Future experiments should also extend the final questionnaire and/or add an interview at the end of it to seek better understanding of the causes of mistakes and, if the experiment is conducted via the web, to track context changes outside of the experimenter control to better classify cases (e.g. asking participants if they performed the two parts in the same room, if they were interrupted by something, etc.). Finally, future research should also investigate counter-measures that could help overcome no-transfer effects, such as providing detailed analogy-oriented questions in the learning phase and/or adding analogues to the learning phase and/or teaching meta-strategies (e.g. how to systematically search for analogies that might help in solving a new problem).

Research question # 2. What is the effect of learning two analogues rather than one on transfer and explicit recognition in binary decision making tasks?

To answer the question if two analogues lead to significantly higher transfer and explicit recognition of the analogy than just one, we ran eight experiments (two lab based and six online) which compared solution rates and explicit recognition rates of participants who had learned one or two analogues by summarization. We also performed two binary logistic regressions, one for transfer and one for explicit recognition, including all key variables and their interactions, on all the relevant cases.

In all the experiments analyzed one-by-one, we never found two analogues to lead to significantly higher transfer than one. This result was confirmed by the logistic regression results, which shows that number of analogues does not have a significant main or interaction effect on transfer.

In sum, all our evidence indicates that the number of analogues learned does not significantly influence transfer within the conditions that we tested, which were rather wide (context shift up to medium, delays up to over 20 days, semantic distance up to high).

A few comments are, however, necessary. Firstly, as we mentioned earlier discussing one analogue versus no-analogue transfer, it is possible that more extreme conditions than the ones we tested would make the second analogue relevant. In fact, we have few cases in true boundary conditions (e.g. high semantic distance coupled with delays of 2 days and higher only had three participants learning with one analogue) and no cases in extreme conditions such as high semantic distance, medium context shift and long delays.

Moreover, connected to the first point, it should be mentioned that we observed, in the few cases in which transfer was more extreme, a trend towards significance. For example, taking all cases in which context shift is medium and delay was over 15 days (average delay was 18 days), results were not far from significance. On the one hand, although this result may very well be due to

chance, it is possible that increasing from one analogue to two would impact transfer significantly under some of the conditions which we haven't tested. On the other, it should be pointed out that, since the conditions under which the second analogue did not seem to add significantly to transfer were rather wide in terms of delays, context shifts and semantic distance, the importance of the second analogue appears greatly reduced compared to what was found in open-ended problem solving transfer.

Secondly, a comment should be made regarding Gick and Holyoak's (1983) results which showed that learning with two analogues led to significantly more transfer than learning from just one analogue. It should be underscored that they compared one analogue with two analogically encoded analogues. In their 1989 experiments, Catrambone and Holyoak compared transfer rates of two summarized analogues versus one and did not find a significant difference. They did, instead, find that two analogically encoded analogues were better than two compared analogues. This could mean the learning method was responsible for the difference that had been found. It should be further noted that the conditions under which they examined the relationship were quite restricted (only immediate transfer and no context shift).

Finally, Spencer and Weisberg (1989), who introduced different context shifts and delays, did not see any significant difference in transfer between participants who learned one or two analogues, although they used the same problem and the same procedure. This indicates that context shifts and delays, not only differences in the task, can explain the different results.

As for explicit recognition, evidence from single experiment and logistic regression analyses seems to indicate that number of analogues does not have a significant relationship with explicit recognition.

Future experiments should seek to investigate if adding a second analogue can benefit transfer in the more challenging situations. Varying semantic distance, context shifts and delays in a factorial type design should help identify those situations or rule out that the second analogue brings an advantage.

In practical terms, these results cast doubt on the general usefulness of adding an analogue to increase transfer probability. Without comparison between the two analogues the second analogue, even in open ended problem-solving immediate and same context un-hinted transfer the second analogue was not proven to be useful. Context and time shifts seemed to nullify the benefit of the second analogue in open ended problem and in our type of closed decision transfer.

A useful next step in solving this issue would be to run Gick and Holyoak's 1983 experiment 4 and Spencer and Weisberg's 1989 experiments using our binary decision making problems.

Research question # 3. What is the effect of learning method on transfer and explicit recognition of the analogy in binary decision making tasks?

To answer this question we ran eight experiments (two lab based and six online) which compared solution rates of participants who had learned with three different methods: *summarization* (writing a brief summary of each of the analogues*), guided comparison* (answering a set of questions highlighting structural analogies between the two analogues) and *free comparison (comparing the two analogues and identifying analogies).* We also performed two binary logistic regressions, one for transfer and one for explicit recognition, including all key variables and their interactions, on all the relevant cases.

Although in the single experiment analysis results were mixed (learning method did not have a significant impact on transfer in three experiments out of six), the binary logistic regression which we ran on all the data showed that learning method does not have a significant main effect on transfer. Learning method does, however, interact with other variables significantly, which might explain why we got mixed results. More specifically, these interactions happen as learning method is changed from summarization (which was our default learning) to free-comparison. The strongest and very significant interaction is between context shift from low to medium and learning method moving from summarization to free-comparison. The interaction is positive, which means that a change from low to medium context shift increases transfer significantly for free-comparison learners compared to summarization learners. This effect is difficult to explain with theory and indeed could be spurious (the number of cases on which the result is based are probably not large), but it is interesting to observe that the same interaction also boosts explicit recognition. The meaning appears to be that the increase in context shift is positive for this type of learning, possibly thanks to an increased recognition of the analogy.

On the other hand, the interaction between delay and learning method *free-comparison* seems to have a significantly negative effect on transfer: the probability of transfer decreases with time for learning method *free-comparison* and not for the other methods. Again, the result needs to be verified and the causes understood.

Furthermore, it appears that context shift and delay interact significantly and negatively in the case of type of learning free-comparison. This means that participants who learn with type of learning free-comparison and must transfer in a situation of medium context shift have an initial advantage over participants who learned with summarization and transfer with a low context shift. However, passing of time diminishes and eventually nullifies the advantage. The same comment

that we made discussing the previous interaction effect applies: the number of cases is not very high and the effect could be spurious.

Finally, semantic distance interacts with delay in the case of learning type *free comparison*. This interaction, significantly positive, means that the negative impact of high semantic distance, in learning type free comparison, is lessened by delay. In other words, it seems, that the effect of delay, with learning type *free-comparison*, is different depending on semantic distance being medium or high. This result should be investigated further (number of cases is limited) and its causes explained.

Comparing our results to the ones from open-ended problem solving studies requires a few comments. Although Catrambone and Holyoak (1989), confirming Gick and Holyoak (1983), showed with a direct comparison that analogical encoding leads to higher pre-hint transfer levels than just summarization in immediate and not context-shifted transfer, when they introduced context and time shifts the advantage of analogical encoding in un-hinted transfer disappeared (as had been found also by Spencer and Weisberg, 1986). On the other hand, superiority of analogical encoding over other methods was confirmed, even with time-context shifts, by further experiments in the domain of transfer of negotiation skills (Gentner, Lowenstein and Thompson, 2003). So, in reality, even in open-ended problem solving studies, when un-hinted transfer was considered, analogical encoding had been found superior only under certain conditions.

We found learning method even less relevant to transfer, since it does not have a main effect but only interactive effects. It is possible that the mapping and schema abstraction that comparison induces were not a significant advantage in our less-open type of task. It is also possible that the other mentioned differences (specially context shifts) were responsible for the results that we obtained. Moreover, we found *free-comparison* to be a case in itself: in this type of learning delay, context shifts and semantic distance interact significantly. These findings, if confirmed, also need to be further appreciated because it is unclear what causes free-comparison to be different form other learning methods.

In the case of this issue, future research should be aimed at collecting more cases in order to base conclusions on firmer grounds and also to understand what is making *free-comparison* different. To pursue the latter objective, initially think-aloud protocols should help formulate hypothesis, which could then be incorporated in a structured questionnaire.

As for explicit recognition, logistic regression analysis indicates that learning method does not have a significant main effect on explicit recognition, but that learning method free-comparison does have an effect in interaction with context. Moreover, close to significance values were reached for the difference between summarization and guided comparison learning methods.

Research question # 4. Does schema formation support transfer in binary decision making tasks?

There are two ways of attempting an answer to this question. The first is the one followed by Gick and Holyoak (1983) who searched for a correlation between quality of schema and transfer rate. As they were able to show in open-ended problem solving transfer, the better the schema, the higher the transfer rate. Consequently they were able to posit that schema mediates transfer.

Our results, however, show that quality of summaries, not only of schemas, was predictive of transfer. We interpret this to mean that, in binary decision making tasks, it is not necessarily the schema that mediates transfer, but rather learning level or a confound variable (note that this could be true for Gick and Holyoak's cases as well).

Another way to understand if schema fosters transfer is to compare transfer rates of participants learning with one analogue to those of participants learning with two analogues by comparison. The latter category should derive a schema by mapping of the two analogues (which is induced by the comparison instructions) while the former should not be able to. In fact, Gick and Holyoak (1983) tried unsuccessfully to support schema formation from one analogue augmenting it with a verbal statement of the underlying principle, in experiment 2, and adding a diagrammatic representation, in experiment 3. This made them conclude that schema formation required at least two analogues and comparison between them.

As our analyses show, it is never the case that transfer rates are significantly higher for participants learning two analogues by comparison than for those learning from one analogue by summarization. To ensure that we were considering cases in which learning had taken place following our instructions, we eliminated the cases falling in the no/little learning category, but still did not reach significance.

Based on these results, we cautiously hypothesize that the role of schema in mediating transfer is less relevant in binary decision making than it is in open-ended problem solving, possibly due to the more circumscribed type of question that our transfer problems require to answer.

Both methods of assessing the importance of schema in transfer, however, have limitations and we think future research should further investigate the issue, which has theoretical relevance. On the one hand, future experiments should find a way to classify accurately cases in which a schema was formed and those in which it wasn't. Think-aloud protocols might help in understanding if participants are recalling the learning stories via a schema probably better than final questionnaires as it is difficult for participants to accurately self-report what happened during the analogical transfer process. On the other hand, future research should be very careful about inferring that schema mediates transfer based on the fact that schema quality predicts successful transfer because

there are several possible confounds (including cognitive ability which could favour production of high quality schemas as well as transfer).

Moreover, we found that it is difficult to evaluate learning simply form participant's output because some learners appear to learn far more than they care to put in writing, thus potentially distorting results. To contain this issue, think-aloud protocols in the learning phase may be useful as they would allow to better assess the level of learning. Furthermore, offering an incentive to good performance in the learning output (although potentially distorting) might help ensure that participants reflect their learning as much as possible in the output.

Finally, to control individual characteristics which might act as confounds, participants could be tested in order to identify and counterbalance relevant individual differences in conditions.

Research question # 5. What is the effect of analogue diversity on transfer and explicit recognition of the analogy in binary decision making tasks?

We studied this relationship in a somewhat more limited way than we investigated the others (we just considered transfer after 5 minutes, with low context shift and high semantic distance), but nevertheless managed to base our conclusions on 189 cases.

Confirming Gick and Holyoak's findings, diversity does not seem to impact transfer nor explicit recognition of analogy. It should however be pointed out that we were not able to study interaction effects, thus we can't exclude that diversity would become relevant in interaction with other variables. It should be noted that it is a practical and widely accepted practice in training to provide diverse examples in order to allow a more flexible use of what is being learned. It is therefore interesting to observe that this seems not to make a difference, at least in terms of main effect. If future research decides to pursue this issue further, it would make sense to vary the other variables (context shift, semantic distance and, possibly, delay) in a factorial design experiment so that interaction effects can be detected.

Research question # 6. What is the effect of semantic distance on transfer and explicit recognition of the analogy in binary decision making tasks?

We base the answer to this question mainly on our two binary logistic regressions, one for transfer and one for explicit recognition, including all key variables and their interactions, which we performed on all the relevant cases. As semantic distance was never varied *within* experiments, all we could examine at an experiment level was if significant transfer and explicit recognition had

taken place with that level of semantic distance i.e. if our task was solved at levels significantly above control.

Our logistic regression shows that semantic distance was the only variable having a significant main effect (negative) on transfer. This is consistent with research in the domain of open ended problem solving (Gick and Holyoak, 1983) who found lower transfer from their military and fire-fighting stories to the *Radiation* problem than they did from the *Cord* problem to the semantically closer *Birthday party problem* (in which cords are substituted by decorative party ribbons).

Moreover, as reported discussing the relationship between type of learning and transfer, there is a significant three-way positive interaction effect of semantic distance and delay in the case of learning type *free comparison*. This means that the negative impact of high semantic distance, in learning type free comparison, is lessened by delay. In other words, it seems, that the effect of delay, with learning type *free-comparison*, is different depending on semantic distance being medium or high: the effect of high semantic distance on transfer is different depending on the moment in time i.e. that it is more negative right away and then, as time goes by, it is reduced.

Despite its negative effect on transfer, semantic distance never impaired it in our experiments: transfer levels were always above control levels, even in the experiment with several days of delay between learning and transfer. It should be underscored, however, that we varied semantic distance only in a low context shift situation, thus we do not know anything about its interaction effects with context.

As for the relationship between semantic distance and explicit recognition, it was not significant. This result, in part surprising since there is a significant effect of semantic distance on transfer, may be due to the fact that semantic distance does not affect the recognition part of the transfer process, but some later stage, for example retrieval or mapping.

Unfortunately, we did not have cases of medium context shift and high semantic distance, which are interesting because they are both challenging and realistic (i.e. close to what happens in naturalistic learning). Further research, to begin with, should widen the cases investigated. The experiments should vary semantic distance and context in a factorial type design to study possible interactions between semantic distance and context shift. Furthermore, it would be interesting to run experiments aimed at understanding what can be done to counter the negative effect of an increase in semantic distance. Increasing number of analogues from 2 to 3 does not seem promising (given that we found no difference between two and one) nor does using a higher number and/or more detailed analogy-oriented questions to lead the comparison (guided-comparison was not significantly better than the other learning methods), but nevertheless should be tried. Another

possible approach could be to give some participants a short training on how to encode learning in a way that helps bridge semantic distances (e.g. at a more abstract level).

Research question # 7. What is the effect of delay on transfer and explicit recognition of the analogy in binary decision making tasks?

To study the relationship between time delay (from 5 minutes to over 20 days) and transfer and between delay and explicit recognition, we ran and analyzed nine experiments (three lab-based and six online). We also performed two binary logistic regressions, one for transfer and one for explicit recognition, including all key variables and their interactions, on all the relevant cases.

In the single experiments, the relationship was tested under different conditions (number of analogues, learning method, context shift and semantic distance) and was never found significant with a minor exception (free-comparison learning in experiment 8 with medium context shift).

The logistic regression on all experimental data confirmed non significance of the relationship, but highlighted a couple of significant interactions (with semantic distance and with context) in the case of learning method free-comparison.

In sum, our web experiments indicate that delay (between 5 minutes and one day up to around 20 days) does not have a significant impact on transfer except in limited cases, occurring in interaction with other variables.

In the two lab experiments in which context shift was high and the delay was a week long, however, experimental conditions did not transfer at significantly higher levels than control, indicating a possible interaction effect with that level of context shift (a main effect seems unlikely as our online experiments show that even long delays alone cannot stop transfer).

Delay was also not found to have a significant main effect on explicit recognition both in our logistic regression and at a single experiment level (in five out of six cases). This appears consistent with our finding on transfer. Delay, however, has a significant interaction with context: the negative effect on explicit recognition due to context shift is less strong as time goes by. So, not only does delay not have a negative effect, but it can even, under certain circumstances, have a positive one. Again, it should be said that these types of interaction effects are found on subsets of the total participant population therefore should be verified with more cases.

Our findings differ from those in the domain of open-ended problem solving transfer where delays, which inevitably bring a change in context (at least mental context) had a severe impact on un-hinted transfer. Spencer and Weisberg (1986), later confirmed by Catrambone and Holyoak (1989), showed that introducing even a short time delay (6 minutes) between learning and transfer eliminates transfer, if coupled with a context shift which was very similar to the one we categorized

as *medium*. In our experiments, on the contrary, even long delays and a medium context shift did not impair transfer significantly.

Once again, it would be useful to have more data, particularly in high context shift situations. To confirm cases in which delay is significant and to better understand time effects, targeted experiments should be performed. These experiments should adopt free-comparison as a learning method and vary in a factorial design variables whose interaction effects are to be investigated. Moreover, it would be interesting, from a theoretical point of view, to understand if what we observed is due to the shape of the forgetting curve or if other explanations are necessary. It would also be useful to study if the pattern of mistakes is the same or changes with delay. To do this, an increase in the quality of data is necessary. This in turn would require a categorization of mistakes to allow an in-depth analysis of missed transfer, think aloud protocols and more detailed final questionnaires.

Taking a practical perspective, there is not very much that can be done to counter delay effects, with what we know. The first step would be to understand the type of mistakes that time induces, what causes them and what, if anything, reduces them.

Research question # 8. What is the effect of context shift on transfer and explicit recognition of the analogy in binary decision making tasks?

We base the answer to this question mainly on our two binary logistic regressions, one for transfer and one for explicit recognition, including all key variables and their interactions, which we performed on all the relevant cases. As context was never varied *within* experiments, all we could examine at an experiment level was if transfer and explicit recognition had taken place despite the context shift we had applied.

Based on our two lab experiments with high context shift, where there was no significant transfer effect, we could cautiously say that, probably, high context shift, alone or in interaction with delay, impairs transfer (if the effect was not spurious). We attribute the drop in performance to high context shift or to high context shift in interaction with delay since we know that delay alone – well over a week- is not enough to stop transfer.

On the other hand, in all other experiments (web-based), context shifts up to medium level were not sufficient to impair transfer, even in presence of long delays. It could be hypothesized that context shift becomes critical as it changes from medium to high.

Our logistic regression on all online experimental data (which included three types of context shift: low, medium-low and medium) resulted in a non-significant relationship between context and transfer (i.e. no main effect). We found however that context has significant impact in interaction

with other variables (learning method, delay, semantic distance) as already pointed out. There is a two-way interaction between learning method and context shift: in the case of learning method *free comparison*, context shifts from low to medium have a strong and significant positive effect on transfer. Moreover, there is a significant negative 3-way interaction effect between this context shift, type of learning free-comparison and delay which means that participants learning with free-comparison in a situation of medium context shift loose their initial advantage in transfer (the positive two-way interaction effect describe above) versus participants who learned with summarization as time elapses.

In conclusion, the role of context shifts up to medium appears perhaps less important in transfer of binary decisions than we were expecting based on previous literature (Spencer and Weisberg, 1986). A context shift consisting in a large change in presentation format, in source ("a college wishing to remain anonymous" vs. Warwick University) and in goal (pre-testing of materials vs. experiment) did not significantly impact transfer in our experiments. In Spencer and Weisberg (1986), with open ended problems, the change of person administering the experiment coupled with a different declared goal (experiment vs. class demonstration) were sufficient to impair transfer even if physical context did not change (i.e. both parts of the experiment were administered in the same room).

On the other hand, context has a relevant impact on explicit recognition since we found a significant relationship. This could, in part, explain why context appeared more important in the transfer process of open ended problems, where recognition was the main factor affecting transfer. With insight problems, the critical aspect of the transfer process is retrieval, which is based on recognition. If context acts as a cue, as it generally does, it can have a relevant facilitative effect on the process, as pointed out by Spencer and Weisberg who wrote: "Perhaps contextual similarity, rather than analogy, served as a basis for retrieval of the past problem in Gick and Holyoak (1980, 1983) and Bauer (1978) studies." (Spencer and Weisberg, 1986, p. 442).

Caution should be used in drawing conclusions, however, as we do not have a precise metric for context shifts (e.g. how important is the change in physical space versus the change in person administering the experiment? How relevant is the context "on the screen" in a web experiment versus the physical context surrounding the participant?) nor did we control context completely in the web-based experiments (e.g. we don't know if learning and transfer happened in the same room).

Much more needs to be done to understand the impact of context on transfer. To begin with, a metric of context shift is needed and should be based on a rigorous definition of what constitutes context. For example, we attributed a medium level of context shift to cases in which learning and

transfer were presented differently (college testing materials vs. experiment) and in very different formats. However, it may be that the physical context surrounding the participant counts much more than what appears on the screen and that the shift is smaller than we think. Moreover, most of our results are based on web experiments, in which the level of control over context is not complete and in which a high shift is difficult to create. Probably it would be useful, even if challenging from a practical point of view, to re-run some experiments using the lab because it allows to carefully monitor context. Furthermore, we were not able to fully investigate high context shifts. Again, although practical problems make this challenging, it would be useful to collect more data on transfer with high and with medium context shifts, particularly as it would seem that the change from medium to high is the most critical.

Finally, if further web-experiments were run to analyze the relationship between context and transfer, it would be useful to administer, after the second part of the experiment, a questionnaire asking about the context that was surrounding the participant (e.g. physical space, presence of other people in the same space, music, etc.) while performing the learning and transfer tasks in order to attribute a more correct level of context shift to each case.

From a teaching and learning perspective, it is useful to know how much context counts, with what it interacts in creating positive or negative effects on transfer and if its relevance is the same for different problems. If context is as relevant as problem solving literature reports it to be, then practitioners should consider investing heavily on two aspects. The first is trying to make people learn in a context which is as close as possible to the performance context (the other way around seems less possible). For example, e-learning performed at one owns desk should be better than classroom training if, presumably, the need to transfer will happen at the same desk. The second aspect would be to devise some training that would teach, as a meta-competency, how to abstract oneself from the external context.

Research question # 9. What is the role of quality of learning in transfer?

To study of the relationship between quality of learning (on three levels) and transfer, we analyzed the results of our 16 experiments (three in the lab and 13 online) and performed a binary logistic regression which had explicit recognition and quality of learning as independent variables on all the cases.

Although at the single experiment level the relationship between quality of learning and transfer presented mixed evidence (probably due to lack of power), the logistic regression performed on all online data clearly showed a significant relationship between quality of learning and transfer. The greatest and most significant effect of the internal variables on transfer was shift from low to high

quality. The results have been checked with binary regression performed on sub-sets of data (just one analogue, all learning methods separately) to always find very significant levels for the relationship between quality of learning and transfer, with the only (already mentioned) exception of free-comparison learning.

There are no significant interaction effects between quality and explicit recognition. Fisher's exact test was also performed on the same data, confirming that quality of learning is highly predictive of transfer. The fact that, at a single experiment level we had found mixed results, can probably be interpreted as a lack of power of the relationship due to smaller number of cases.

Our results therefore confirm and extend the importance of quality of learning as a predictor of transfer that has been found by Gick and Holyoak (1983). As we mentioned, the conditions under which quality of learning was found significant in our experiments are wider than those found in problem solving transfer studies because the cases we examined include learning from summarization and not only from comparison-based learning, long term as well as 5 minute delay, different levels of contextual shift and of semantic distance.

Spencer and Weisberg (1986) had found that quality of schema is not predictive of transfer in the case of contextually shifted transfer. It should be remembered that we found the same result when context shift was high (lab experiments 1 and 2). Although the low number of cases and consequent lack of power may be a plausible explanation, the finding could also be taken to mean that, as context shift becomes high, the relationship between quality of learning and transfer changes, becoming non-significant.

It is possible, in conclusion, that there are conditions in which quality of learning looses significance. Further experiments might explore this aspect varying the conditions in which transfer takes place, testing the limits of the relationship's significance.

Research question # 10. What is the role of explicit recognition in transfer?

To answer this question, we analyzed one-by-one our 16 experiments (13 online and three in the lab), we ran a binary logistic regression which had explicit recognition and quality as independent variables on all experimental data except experiments 5 and 6 which had used the less reliable version of the questionnaire. While at the single experiment level evidence was mixed, the logistic regression on all or data showed a very significant relationship between explicit recognition and transfer and there were no interactions with quality of learning.

In all our experiments recognition of the analogy (perhaps inflated by self-reporting) was rather high (approximately 70% in online experiments) and higher than transfer (which was around 30%),

meaning that insight was not the only critical factor in transferring this type of problem. On the other hand, only rarely participants who did not recognize the analogy managed to transfer, which would indicate that transfer of binary decision making is mostly an explicit process.

One of the points which remain beyond the reach of our data is the differentiation between feeling of knowing and actual recognition of the analogy (explicit recognition was self-reported in our experiments and thus we can't exclude that some participants might have declared that they recognized the analogy in hindsight). In order to assess explicit recognition with a higher degree of certitude, distinguishing it from feeling of knowing, further experiments should be conducted perhaps with a dual pass (seeking un-hinted transfer first and hinted transfer afterwards) or with think-aloud protocols.

From a practical point of view, given the importance of explicit recognition effort should be put in enhancing people's recognition capabilities (possibly a key individual difference for transfer), which could be perhaps be developed as a crucial meta-competency, through appropriate training.

How far the ability to recognise problems with similar structure is distinct from the more general ability to encode information abstractly is a question for future work. It is possible that the dependence on abstract encoding varies with the type of problem. For the less open type of problems, such as the one we used, abstract encoding might be less important. Possibly, if the learning and target problems share more similarities (in our case, the question at the end was similar), retrieval does not rely as heavily on the abstract representation of the learned problem. It seems (but is beyond the reach of our data to determine it) that the learning story/ies may be retrieved, rather than the schema, when the target problem is tackled. Our findings on the lesser role of schema in mediating transfer could be taken to support this view.

On the other hand, for those problems, as the open-ended ones used by Gick and Holyoak, in which recognition is made more difficult by the absence of any type of cue, an abstract encoding (e.g. for the military learning analogue of the *Radiation* problem: "a story about how to overcome a target when a sufficiently great force is available, but cannot be applied along one path") surely is more effective in supporting retrieval of the key principle to transfer than a non-abstract one (e.g. "a military story about a fortress that was conquered attacking from several points simultaneously because the roads were mined and passage of a large force would detonate them").

Clearly, if the ability to encode learning analogues abstractly is relevant to transfer, then it would be extremely useful to teach learners how to identify, to focus attention on and to memorize the key features of training exemplars, without becoming "distracted" by the context of what is learned. However, as we speculated earlier, this might be more useful in analogical transfer of open-ended problems than of binary decision making ones.

As we mentioned this research is to be considered a first step in better understanding transfer outside of the traditional open-ended problems which have been used in most classic experiments. It indicates that what was found to be true in open ended problems does not necessarily extend to other tasks, thus suggesting that further research should more widely vary the types of tasks used.

In sum, we were able to see un-hinted transfer happen in challenging situations of delay, semantic distance and context shifts. The greatest obstacle, we would tentatively conclude, could be high context shifts. If these results were to be confirmed, practical efforts in *training for transfer* should be addressed primarily in trying to overcome this type of obstacle.