An inquiry about students’ naïve knowledge of metacognitive strategies and the delayed JOL effect

Ivo Todorov
AN INQUIRY ABOUT STUDENTS’ NAÏVE KNOWLEDGE OF METACOGNITIVE STRATEGIES AND THE DELAYED JOL EFFECT

Ivo Todorov

Properly tuned metacognitive knowledge is important for setting up realistic learning goals. One of the more robust findings in metacognitive science, the delayed JOL effect, pertains to the fact that delaying judgments of learning (JOL) leads to more accurate monitoring of one’s learning. Thirty students were tested on their knowledge of metacognitive strategies. They were asked to study paired associates, make JOLs, and were later tested with a cued recall test, as well asked about the efficacy of strategies for making JOLs. There was a significant positive effect in monitoring accuracy, from delaying JOLs, yet the participants showed poor explicit knowledge of it, and neither did their choice of strategy improve with task experience. The results demonstrate the important role of correct assessment during ongoing learning, and that even experienced learners, such as, university undergraduates are seemingly unaware of which strategies lead to optimal monitoring.

The aim of the study is to; (i) replicate the delayed JOL effect, a term denoting the fact that delaying judgments of learning (JOL) leads to highly accurate monitoring of one’s learning, (ii) to investigate students’ knowledge, and naïve beliefs, about metacognitive strategies, and (iii) to explore whether the preference for metacognitive strategy changes as a function of task experience. The present study draws upon an earlier study conducted by Jönsson & Kerimi (2011), where students’ metacognitive knowledge about the delayed JOL effect and their knowledge about metacognitive strategies, were investigated. I plan to address the methodological limitations of the Jönsson & Kerimi (2011) study, and conduct a more systematic investigation of students’ metacognitive strategies with regard to the delayed JOL effect.

Background and short history of metacognition
In the late seventies John Flavell coins the term metacognition, which refers to thoughts about one’s own thoughts and cognitions (Flavell, 1979). This fact does not allude to metacognition being a new phenomenon. One could look much further back in time and trace its origins to the first human (there is still an ongoing debate whether animals exhibit metacognitions, for a short review see Smith, 2005) that has ever wondered about his own thoughts or knowledge. That of course is just a hypothetical contemplation that lies far beyond the scope of this paper. Instead, for the beginning of this short background, I choose to omit the first couple of hundred thousand years, of evolution of human thought and jump straight to René Descartes and his idea of self-reflection. It was an ingenious idea that gave him conviction of his own existence. His famous phrase “Cogito Ergo Sum - I think, therefore I am” has since become an influential thought of western philosophy (Baird et al., 2008). The ability to think about one’s own thinking was for him the one thing that he deemed as the undisputable proof of him being a conscious entity (Metcalfe, 2008). In hindsight, it can be said that, through metacognitive reflection Descartes was able to establish the starting point for humans, from where all further knowledge originates.
In what is widely regarded as the paper to truly spark the metacognitive research, Joseph Hart (1965) determines an ingenious method for investigating the accuracy of peoples’ metacognitive capabilities with the help of Feeling of Knowing judgments (FOK). FOKs are judgments about the likelihood of recognizing currently unrecallable answers on an upcoming test (Metcalfe, 2008). He uses the so called recall-judge-recognition (RJR) method, where participants are asked to make FOKs about stimuli that could be recognized, but not verbally recalled. Following on that paper, which focuses on memory monitoring and feeling of knowing judgments, other researchers begin using introspective judgments to explore peoples’ monitoring of memory processes (Kroll & Kellicut, 1972; Flavell, Friedrichs & Hoyt, 1970; Markman, 1977).

For the following years the state of the field remains somewhat isolated and fragmented up until Nelson and Narens (1990) propose a robust theoretical framework for understanding metacognition. This framework rests on the basis of three simple principles. First there are two separate levels of cognitive processing, the meta-level and the object-level. The object-level consists of mental representations of objects, i.e. cognition. The meta-level holds both a higher order mental representation of the task that the person is performing and the ongoing cognitive processes that are engaged in completing that task, such as reflecting upon ongoing memories and ongoing learning (Metcalfe, 2008). The two levels of processing interact by the means of bidirectional flow of information. The flow of information from the object-level to the meta-level constitutes the process of monitoring, whereas information flowing from the meta-level to the object-level is considered control. Through monitoring information about the state of the object-level is being brought to the meta-level. This allows the meta-level’s model of the object-level to be updated. Then, depending upon the state of this model, control can initiate, sustain, or terminate cognitive activities such as ongoing learning, problem solving or memory retrieval. The requirement for such metacognitive framework is that the object of the judgment has to be a mental event itself and not something that is present in the immediate environment, i.e. a stimulus being perceived at the same moment as it is being thought of (Dunlosky & Metcalfe, 2008).

**Metacognition and Metamemory**

In much of the literature, the term metacognition is broadly construed to apply to all self-reflective facets of cognition. Metacognition itself is not just about making judgments among different options, nor is it the product of a complex multi-step response to a difficult discriminative judgment. Instead, it is a rather special kind of judgment or mental commentary that involves the two different levels of processing that Nelson & Narens (1990, 1994) called the object-level and the meta-level.

Cognition is defined as symbolic mental activities and mental representations, that is, the process of thought (Dunlosky & Metcalfe, 2008). Metacognition on the other hand refers to thoughts about one’s own thoughts and cognitions (Flavell, 1979). Metacognition differs from standard cognition in that the self is used as a reference point of the processing or the knowledge (Wellman, 1977). A similar distinction exists for memory and metamemory. Memory is defined as the ability to encode, retain and recall information and experiences (Bäckman & Nyberg, 2010). Metamemory on the other hand, describes the processes by which people are actually able to examine the content of their memories, either prospectively or retrospectively, and make judgments or mental notes about them (Metcalfe, 2008). Thus metamemory is not memory itself, although it seems to be critically dependent on memory.
**Monitoring and Control**

As previously mentioned, the flow of information from the object-level to the meta-level constitutes the process of monitoring. It can also be described as peoples’ assessments of the current state of their cognitive activity. Whereas metacognitive control is the process of regulating some of the aspects of the cognitive activity, that is, the flow of information from the meta-level to the object-level. In short, monitoring refers to making judgments about one’s memory and control is using those judgments to guide actual behavior (Metcalfe, 2008). From this, it becomes pretty clear that monitoring and control function in a tightly orchestrated interplay with each other.

During experimental memory tasks, such as the one used in the present study, monitoring and control are present in all three stages of the memory process, namely, encoding, retention and recall, (Nelson & Narens, 1994).

Metacognitive monitoring includes ease-of-learning (EOL) judgments, judgments of learning (JOL), feeling of knowing (FOK) and retrospective judgments (RC). EOLs are judgments pertain to how easy or difficult it will be to learn a given item, before actual learning has begun. JOLs are judgments about the likelihood of remembering recently studied items on an upcoming test, and are made either during or after learning. FOKs are judgments of the likelihood of recognizing currently unrecallable answers on an upcoming test. Finally, RCs are judgments about confidence in retrieved answers (for a review on different metacognitive judgments, see Dunlosky & Metcalfe, 2008).

Control processes include selection of the kind of processes, allocation of study time, termination of study, selection of memory search strategy, and termination of search (Nelson & Narens, 1994). When optimized, these two aspects make for accurate metacognitions, which can enhance students’ learning (Azevedo & Cromley, 2004; Pintrich, 2002). For example, Azevedo and Cromley (2004) demonstrated that a short training session in self-regulated learning, designed to foster their participants’ conceptual understanding, resulted in a significant shift in the learners mental models, which in turn, resulted in much better learned material, as well as deeper understanding of it. While studying, an individual monitors his learning, that is, assessing how well the to-be-learned material is being understood and whether a desired goal state is being approached. Monitoring might (or might not) lead to metacognitive control, namely, deciding to spend more, respectively less, time trying to learn the material (Metcalfe & Finn, 2008). The learning process can be ineffective even when a person’s metacognitions themselves are excellent, if these metacognitions are not converted into optimal control strategies (Metcalfe, 2008). Accurate metacognitions could also result in a realization that the current study strategy might not be fully adequate and as a reaction, the student can employ a new strategy in order to solve a problem (reach a goal state) that has proved to be rather difficult. Nelson & Narens (1994) write that metacognitive control is exerted whenever the meta-level modifies the object-level, i.e. information from the meta-level acts to influence the ongoing activity in the object-level.

As mentioned earlier, monitoring and control alone do not constitute metacognition. An important part of it is that a mental event must be the object of such monitoring and control. For example, if a person just makes a judgment about something that they see, hear or smell, it is not a metacognitive judgment they are making. It is instead, a judgment about a stimulus that is being present in the moment of judging and it is not a judgment about a mental representation. Metacognitive judgment is a judgment about an internal representation.
(Metcalfe, 2008), hence metacognition differs from mere judgment about the world insofar it is not bound by the presence of an immediate external stimulus.

**Judgments of Learning and Methods for measuring their accuracy**

Judgments of learning (JOL) are cognitive assessments that people often make in automatic fashion while learning something (Nelson & Dunlosky, 1991). JOLs are concerned with the extent to which the task that needs to be learned, is indeed learned (Schwartz, 1994). According to theory, such judgments are important in studying, because they tap in to peoples’ prospective knowledge and are presumably used to control further study (Dunlosky & Hertzog, 2000; Metcalfe, 2008; Nelson & Dunlosky, 1991; Thiede, Anderson & Therriault, 2003; Metcalfe & Finn, 2008). More closely, JOLs seem to be closely linked to termination of learning or making a decision to continue studying (Metcalfe, 2008). Thus, if such judgments are faulty, the individual’s study behavior will be non-optimal at best, or totally ineffective at worst. The general idea is that faulty JOLs lead to non-optimal learning, while accurate JOLs make possible for optimizing the learning process (Dunlosky & Metcalfe, 2008) because JOLs seem to be causally related to study choice (Metcalfe & Finn, 2008).

One method for studying peoples’ metamemory in experimental setting is by instructing them to make judgments of learning while, or shortly after, studying. Usually, JOLs can be made on an item-by-item basis or in a cumulative manner - aggregate JOLs, where after being presented the whole stimulus list without interruption the participants are asked to predict how many items and then JOLs are being made about each item on the list in consecutive fashion. In the present study, both methods were used for assessing students’ metacognitive beliefs.

Absolute and relative accuracy are two different measures for calculating JOLs diagnostic qualities. Absolute accuracy pertains to the degree to which the level of judgment ratings corresponds to the actual level of performance on a cued recall test (Metcalfe, 2008). Relative accuracy is a measure of the degree to which a person’s metacognitive judgments predict the likelihood of correct performance on one item relative to another (Metcalfe, 2008). According to Pintrich (2002), absolute accuracy is the accuracy that seems to be most crucial for learning. In the present study we have only calculated JOLs absolute accuracy, namely, the degree to which the participants’ judgments corresponded with their performance on the cued recall test.

**The delayed JOL effect**

Earlier studies on the subject showed that JOLs are rather inaccurate at predicting actual recall (Vesonder & Voss, 1985). In these studies JOLs were always made directly after the study trial and are therefore called immediate JOLs. A condition in which JOLs for paired associates have a near perfect correlation with results on a subsequent cued recall test was discovered by Nelson and Dunlosky (1991), when they asked their participants to study paired associates and make JOLs after a short delay. In the aforementioned study, the mean gamma (G)\(^1\) correlation between JOLs that were done immediately after the study trial and correct recollection on a subsequent cued recall test was .38. In comparison, the mean gamma correlation for the delayed JOLs was .90. Thus, when made after a short delay, and when the context in which the JOL is made allows for a diagnostic test of long-term memory (i.e., only

---

\(^1\) In the field of metacognition, a non-parametric Goodman-Kruskal gamma correlation has become the standard measure to correlate metacognitive judgments with memory test performance (for a comprehensive rationale on using gamma, see Nelson, 1984). Gamma ranges from -1.0 to +1.0, where 0 represents nil accuracy and +1.0 annotates perfect accuracy.
the cue word is shown, and not both words in a pair), JOLs seem to be far better at discriminating between items with high or low probability of being recalled. By deliberately delaying the JOLs, Nelson and Dunlosky (1991) discovered one of the most robust phenomena in metacognitive literature and labeled it the delayed JOL effect (see Sikström & Jönsson, 2005; Rhodes & Tauber, 2011, for reviews).

Delivering the judgments seems to grant them much higher predictive value, but there is another important factor as well. Nelson and Dunlosky (1991) had their participants study paired associates and make either immediate or delayed JOLs, while presented with either cue-only stimulus (e.g. dog-) or both cue and target stimulus (e.g. dog-spoon). Thus the choice was between four separate strategies: immediate cue-only JOL (iJOLc), immediate cue-target JOL (iJOLct), delayed cue-only JOL (dJOLc) and finally delayed cue-target JOL (dJOLct). The delayed JOL effect occurred in the dJOLc condition only and it was, as showed in their previous study, highly accurate (median G = +.93) in terms of absolute accuracy. This is because only the dJOLc strategy lets the participants perform a diagnostic retrieval attempt from long-term memory before making their JOL rating. In sum, among these four monitoring strategies, only the dJOLc strategy leads to optimal monitoring, and as better monitoring leads to better control of study (Metcalfe & Finn, 2008), it is important that the participants have knowledge of the superiority of this strategy (which is investigated in the present study).

Often, in studies dealing with the delayed JOL effect, the relationship between stimulus and JOL type is comprised by the experimental design. In other words, the choice about which items within a stimulus list will be followed by either an immediate or a delayed JOL is predetermined. Thus making a choice about metacognitive strategy is not left up to the participants but recently, Jönsson and Kerimi (2011) conducted two experiments where students were prompted to choose the most efficient JOL strategy while studying paired associates. The metacognitive strategies used in the study were the same as the ones used by Nelson and Dunlosky (1991). In both experiments the procedure was done on an item-by-item basis and was dubbed as a self-regulated judgment task, as opposed to the self-regulated learning task method used in most metacognitive studies (where participants get to choose which items, and for how long, should be studied).

In the first experiment by Jönsson and Kerimi (2011), participants began by scoring all forty Ukrainian-Swedish paired associates on perceived EOL. Then, after studying the first word-pair, they were prompted to choose one of the four available JOL strategy alternatives, and depending on their choice were either presented with the next study trial (if they chose one of the two delayed JOL strategies) or with a JOL prompt (if they chose one of the two immediate JOL strategies). After the whole stimulus list was studied and judged, and after a short filler task, the experiment ended with a cued recall test.

In the second experiment, the procedure was identical, but was repeated two times. This served as a way of measuring whether students do update their metacognitive knowledge as a function of task experience (i.e., demonstrate increased awareness of the delayed JOL effect).

The two experiments were designed to investigate students’ knowledge of the delayed JOL effect, since strategy choice was left to the participants. The reported use of dJOLc strategy ranged from 25 to 34% across the two experiments. Data shows that the students were mixing different strategies throughout the experiment and choice of strategy was, to a degree, correlated with perceived item difficulty. Experiment 2 showed that such unawareness of the best judgment strategy could not be attributed to inexperience with the task, given that no
significant increase in the preference for dJOLc strategy was observed across both sessions. The authors conclude that whereas some students seem aware of the delayed JOL effect, the majority is not. Also, in line with results from a previous study conducted by Hertzog and colleagues (2009), students generally did not update their judgment strategy adequately through task experience. These results show that there seems to be no general consensus on which the best judgment strategy is, since no single metacognitive strategy was preferred reliably more often than any other strategy.

Metacognitive knowledge
The question of optimizing learning has been a hot topic in educational and metacognitive psychology for quite a long time. Most modern, introductory course books come with instructions on how one should read the respective book, in order to get the most out of it. Unfortunately, these chapters seldom include guidance about metacognitive strategies, and on how to optimally assess one’s learning, even though such knowledge of strategies can play an important role in learning. Metacognitive skills and knowledge, as important as they seem to be, are not often taught in most areas of the curriculum (Kornell & Bjork, 2007).

Metacognitive knowledge involves knowledge about cognition in general, as well as awareness of and knowledge about one’s own cognition (Pintrich, 2002). It is also highly dependent on language since it pertains to peoples’ declarative knowledge about cognition (Metcalfé, 2008). Metacognitive knowledge entails relevant information, about the cognitive task at hand or potential strategies for achieving that task, that one tries to access when thinking about particular target cognition (Serra & Metcalfe, 2009).

Flavell (1979) suggests that there are three distinct kinds of metacognitive knowledge. First, there is knowledge of strategic tasks. It includes knowledge of general study strategies, learning and problem solving, such as knowledge about various strategies that people might use during learning. Knowledge about different metacognitive strategies, which is the focal topic of this article, falls within this category as well. Metacognitive strategies are useful in planning, monitoring and regulating one’s cognition and the learning process (Pintrich, 2002). It is important to note that just having knowledge about different strategies and which one of them are better suited for different kinds of learning does not mean that such successful strategies will be automatically employed while learning. Just to give a practical example: in a situation where something has to be learned within a certain period of time, without the appropriate knowledge about what one doesn’t know yet, it is hard to employ a proper strategy to remedy the situation one is in, namely, actually learning the thing that has to be learned. Yet, with the proper metacognitive knowledge, which is reliant on correct Judgments of Learning (JOL), one might eventually allocate more of the study time to areas that were not properly understood and memorized. Just proper JOLs per se, are not a guarantee for successful outcome, as one can simply understand that he or she doesn’t comprehend the material to be learned but lacks the motivation, interest or the time, to be able to make things right. For practical results to take place, an additional step of controlling behavior is required for turning good JOLs in to optimal control strategies (Dunlosky & Metcalfe, 2008), otherwise one could know what one knows (or does not know), but still continue to do the wrong thing.

Second, metacognitive knowledge is about cognitive tasks that can be amassed, in addition to heuristic knowledge, about different cognitive strategies (Flavell, 1979). It involves the knowledge about whether a certain strategy is applicable, or not, as a remedy to a particular problem. Since not all study strategies are best suited for all possible learning scenarios, a
student must develop certain knowledge about the different conditions and cognitive tasks that call for different strategies. For example, for a student it is rather vital to know whether certain problem demands specific procedural knowledge to be employed, or if general problem-solving heuristics are better suited. Such insight entails knowledge about cognitive tasks.

The third component of metacognitive knowledge is self-knowledge, which includes knowledge of one’s strengths and weaknesses (Flavell, 1979), for example a student knowing that he or she is much better with numbers and analytical thinking, compared to literature and learning languages. Such personal knowledge might prove useful while studying for different classes as it might help the student to adapt to the situation.

It becomes evident that, among a myriad of other factors, knowledge about how to accurately choose an optimal study strategy and about how well something is learned can be used to dramatically improve one’s learning (Dunlosky & Metcalfe, 2008; Rhodes & Tauber, 2011; Pintrich, 2002).

**Research aims**
In a way, the present experiment continues where the study done by Jönsson and Kerimi (2011) ends. They investigated students’ knowledge of the delayed JOL effect (experiment 1 and 2) and whether that knowledge increases as a function of task experience (experiment 2). The purpose of the present experiment was to replicate the delayed JOL effect, as well as to lead an inquiry about students’ metacognitive knowledge and beliefs, and at the same time, address the methodological limitations present in the Jönsson & Kerimi (2011) study. One shortcoming of their experimental design is that by letting participants choose strategy, relative JOL accuracy could not be properly measured. The participants seemed to show preference for certain metacognitive strategies (however, there were no statistically significant preferences of one strategy over another), which led to some strategies not being chosen as often as others, and in some cases, not chosen at all. As a result, stable gamma correlations for each participant, for each of the four strategies could not be measured, and could only be computed on a group level. Another caveat of the study, as mentioned by the authors, was the use of a relatively easy stimulus list, which seems to have led to a relatively high predictive value of all JOL strategies, thus not really isolating the delayed, cue-only JOL strategy as the most optimal one.

In order to investigate students’ metacognitive knowledge in a more precise manner, several measures were taken. First, the experiment was designed in a way, that each participant had to make the exact same amount of immediate and delayed JOLs, and in exactly the same order. Second, only two metacognitive strategies, cue-only dJOL and iJOL, were allowed in order to maximize the contrast between the two JOL conditions, and turn participants’ focus to the temporal aspect of making JOLs, and its key role on their accuracy. Thus, it was possible to compute absolute JOL accuracy for each individual and for each of the two strategies. Third, the stimulus lists were carefully selected with regard to item difficulty and semantic nearness, thus minimizing the moderating function of item difficulty (detailed description follows in the method section). Jönsson and Kerimi (2011) reported an unusually high gamma score for the immediate JOL condition due to relatively easy stimulus list. The authors speculate that it could have led, for at least some of the participants, to use a learning goal as a criterion for choosing judgment strategy (i.e. concentrating on actually learning the material in order to perform well on the recall test, instead of concentrating on choosing the most diagnostic metacognitive strategy), subsequently reducing the chance for replicating the delayed JOL
effect. Finally, the experiment consisted of four successive sessions (i.e. two more than Jönsson & Kerimi had used), in line with study done by Kornell & Son (2009), who investigated students’ use of, and beliefs about, self-testing, and reported that they observed a significant a shift in strategy choice after the third and the fourth sessions of their procedure. In line with their study, predictions are that knowledge about metacognitive strategies will improve as a function of multiple sessions

Method

Participants
Thirty (10 males) participants, with a mean age of 24 years ($SD = 4.62$, $range$ 17-33), were recruited from various academic disciplines and different universities, institutes and colleges throughout the municipality of Stockholm. For their participation in the study they received either course credit or a movie voucher.

Materials
The experiment was designed and was run on PC-computers, via E-prime 2.0 professional software (Psychological Software Tools, Pittsburgh, PA).

One hundred and twenty Swedish paired associates (e.g. snigel – slibbig) taken from Swedish Associations Norms (Shaps, Johansson & Nilsson, 1976) were used as stimuli. The word-pairs were evenly divided in eight lists consisting of 30 words each. The words in each stimulus list had similar association value (varying from one to three). The association value was computed by Shaps, Johnsson & Nilsson (1976), by having participants report the first word they associate with a certain word they were presented with. The mean overall association value and the standard deviation over the eight stimulus lists were equal. This way, precautions were taken that none of the eight stimulus list was harder or easier in terms of its associative value, since perceived semantic relatedness in a word-pair is closely related to how easy or difficult a word-pair is to learn (Carroll, Nelson & Kirwan, 1997), and it also serves as a powerful cue for peoples’ JOLs (Koriat, 1997). By keeping this heuristic fairly constant within the two lists, and hence, decreasing the discriminability between the items, we argued that the diagnostic function of the retrieval attempt associated with the delayed JOL condition, but not the immediate JOL condition, would be even more salient.

Procedure
On average, including instructions, the experiment took around 60 minutes to complete and the participants were either tested alone or together with another participant. The reason for this was purely logistical, as it was not always possible to book two participants for the same day and time. In case two participants were tested simultaneously, they sat by two different computers, separated by a sound-absorbing divider. In addition, they were instructed to not talk with each other for the duration of the experiment.

Upon welcoming the participants, they were presented with a written consent form and general instructions were given. After starting the computer script, their age and gender was entered, and further instructions displayed on the computer screen. The following prompt was presented (all the italicized words in the method section, describing what was being displayed on the screen, are translated from Swedish): *We are interested in which strategy you think is best for making correct judgments of your learning.* It was then followed by a general explanation of the experiment: *In this experiment, you will first study word-pairs, than you will judge how likely it is, at a later memory test, that you will remember the second word in the pair if the first one is being displayed on the screen. The memory test will start a couple of
minutes after you have made your judgment. You will try two different ways of assessing: 1. The judgment will be done directly after each word-pair is presented. 2. The judgment will be done a few minutes after each word-pair was presented. Which one of these two methods for judging will let you make the most correct estimation of how you will actually perform on the memory test later on?

Each session consisted of two parts. The first part involved a study trial with an item-by-item cue-only immediate JOL (iJOLc) session followed by a 30 second filler task, and a cued recall test. The second part consisted of study trial, filler task, item-by-item cue-only delayed JOL (dJOLc), another filler task and a subsequent cued recall test. Each part ended with an aggregate JOL question before the next part commenced. The whole procedure (four consecutive sessions consisting of two parts each) was repeated a total of four times.

The Immediate JOL Session
Study and JOL phase
After the instructions, the first part of each session began with a study trial where each word-pair was displayed on the screen for 6s, immediately followed by a JOL rating on a six point percentage scale JOL (0, 20, 40, 60, 80, 100%). How sure are you that, in couple of minutes, you will remember the second word if shown the first one? After the judgment was made the next word-pair was presented, followed again (after six seconds) by iJOLc screen. There were a total of 15 word-pairs included in each study trial.

Filler activity
After fifteen study-JOL trials, a 45 second filler task consisting of simple mathematical equations was presented on the screen. The filler task was presented after each cued recall test lasted 60 seconds.

Metacognitive beliefs
After each JOL phase and filler activity, participants were presented with a prompt, designed to assess their metacognitive beliefs, since they were asked to give an estimation of the accuracy of their JOLs. The question was as follows. “For each word pair you had to judge the probability of remembering the second word, when presented with the first one. How well do you think, your judgments will correspond with you actual performance on the cued recall test in couple of minutes?” Answers were given on a seven point Likert scale, with 7 denoting perfect correspondence, and 1 denoting no correspondence between JOLs and performance on the cued recall test. This measure will hence be referred to as JOL correspondence rating.

Knowledge about metacognitive strategies
In order to evaluate students’ knowledge about which metacognitive strategy they thought is most diagnostic when assessing ongoing learning the following prompt was displayed at the end of each session (study/iJOL/dJOL/test): “In this experiment you had to choose between two different alternatives on how to judge your learning. Now, in retrospect, which of the two alternatives you think lets you make the most diagnostic judgment on how well you have learned the word-pairs?” The JOL strategy choices displayed on the screen were as follow: 1. It is best to assess learning immediately after the study trial; 2. It is best to assess learning some time after the study trial; 3. Both methods for assessment of learning are equally good.
Cued recall
The cued recall test commenced after the filler activity. Each cue was presented for 15s and the participants, after typing in their answer, could either press ENTER or wait the full 15 seconds in order to advance to the next study trial.

The Delayed JOL Session
Study and JOL phase
The second part of each session was identical to the one described above with the only difference that, the whole study trial was presented in a continuous fashion (again 6s for each word-pair). After the last word-pair was studied a sixty seconds filler task commenced, directly followed by the delayed JOL phase. Delayed JOLs for all 15 word-pairs were done in consecutive order followed by a filler activity task. The last phase of this session was a cued recall test followed by a question assessing knowledge about the delayed JOL effect.

Filler activity
After fifteen study-JOL trials, a 45 second filler task consisting of simple mathematical equations was presented on the screen. The filler task was presented after each cued recall test lasted 60 seconds.

Metacognitive beliefs
After each JOL phase and filler activity, participants were presented with a prompt, designed to assess their metacognitive beliefs, since they were asked to give an estimation of the accuracy of their JOLs. The question was as follows. “For each word pair you had to judge the probability of remembering the second word, when presented with the first one. How well do you think, your judgments will correspond with you actual performance on the cued recall test in couple of minutes?” Answers were given on a seven point Likert scale, with 7 denoting perfect correspondence, and 1 denoting no correspondence between JOLs and performance on the cued recall test. This measure will hence be referred to as JOL correspondence rating.

Knowledge about metacognitive strategies
In order to evaluate students’ knowledge about which metacognitive strategy they thought is most diagnostic when assessing ongoing learning the following question was displayed at the end of each session (study/dJOLc/test): “In this experiment you had to choose between two different alternatives on how to judge your learning. Now, in retrospect, which of the two alternatives you think lets you make the most diagnostic judgment on how well you have learned the word-pairs?” The JOL strategy choices were as follow: 1. It is best to assess learning immediately after the study trial; 2. It is best to assess learning some time after the study trial; 3. Both methods for assessment of learning are equally good. Thereafter, a thirty second countdown clock was displayed to signal the end of the session and the whole procedure started again from the beginning for a total of four times.

Cued recall
The cued recall test began after the filler activity. Each cue was presented for 15s and the participants, after typing in their answer, could either press ENTER or wait the full 15 seconds in order to advance to the next trial.

Aggregate JOLs and assessment of metacognitive beliefs
After each study and JOL phase, both for the immediate and the delayed session, the participants were also asked to give an estimation of the accuracy of their JOLs. By typing a
number between 0 and 15 they would predict how many out of the 15 word-pairs they would correctly remember, at the cued recall test.

**Results**

*Replication of the delayed-JOL effect*  
My first concern was to see whether the delayed JOL effect (Nelson & Dunlosky, 1991), was replicated. This is a prerequisite, as there would be no reason for the participants to prefer that strategy if it is not clearly better than the other JOL strategies. For all 30 participants Goodman-Kruskal gamma correlations (G) were calculated for JOLs and subsequent cued recall within each condition. In order for a response to be scored as correct, it had to be accurately typed on the cued recall test. For 10 participants, for at least one trial, gammas could not be calculated due one or the other variable (JOLs or recall), in the correlation being constant. This was a result of either a person making constant JOLs (i.e. giving 80% JOL on all 15 items in one trial), or having a constant recall value (either correct or incorrect) for all 15 items in one trial. The missing values were therefore replaced with the respective participants’ mean G value for each session. Data was entered into a 2 x 4 (JOL type x Session) repeated measures ANOVA, with JOL accuracy as the dependent variable. There was a clear delayed JOL effect (Figure 1). The delayed JOLs (G = .76, SD = .05) were associated with significantly higher predictive accuracy compared to the immediate JOLs (G = .16, SD = .06), as shown by a significant main effect of JOL type, $F(1,29) = 107.65$, $p < .0001$. Gamma for the immediate JOL condition was significantly lower across all sessions compared to G for the delayed JOL condition. There was no main effect of Session $F(1,29) = .61$, $p = .60$). Neither was there any interaction effect between JOL type and Session $F(1,29) = .35$, $p = .79$).

In order to verify that the imputation for missing data did not alter the results of the experiment, a separate ANOVA ($n = 20$) for repeated measures, with JOL accuracy as a dependent variable, where missing data was excluded using case-wise deletion, also yielded a significant main effect of JOL type $F(1,19) = 60.5$, $p < .0001$ as well. Again, there was no main effect of Session $F(1,19) = 1.71$, $p > .05$, nor any interaction between JOL type and Session, $F(1,19) = .25$, $p > .05$. In sum, the results from all the analyses of variance, with or without imputed values, yielded similar results.
Figure 1. The figure illustrates the predictive validity of the JOLs (as measured with G) across the four sessions as a function of JOL type.

**Students’ metacognitive knowledge**

*Students’ metacognitive beliefs about the accuracy of their judgments*

After the last, immediate or delayed, JOL, in the respective session, participants were asked to judge how well they thought their judgments would correspond with their upcoming performance on a cued recall test. Here, predictions were that students’ naïve knowledge about metacognitive strategies would result in main effect of JOL type and an interaction effect with Session (if they update their knowledge with task experience). The data was entered into a 2 x 4 (JOL type x Session) repeated measures ANOVA, with the JOL correspondence rating as the dependent variable. There was no main effect of JOL type, $F(1,29) = 2.44, p = .13$, no main effect of Session, $F(1,29) = .23, p = .68$, nor any interaction effects between JOL type and Session, $F(1,29) = 1.82, p = .15$.

*Students’ knowledge about metacognitive strategies.*

At the very beginning of the experiment, that is, before the first study trial, and at the end of each session (five times in total) participants were asked to judge which strategy, immediate or delayed, lets them make the most accurate judgments of their ongoing learning. This analysis was conducted in order to examine whether students will gain an understanding of the superior monitoring accuracy of delayed JOLs with task experience. This variable was devised to assess students’ JOL preference for each session, and was hypothesized that the preference will change in favor of delayed judgments, with task experience. A single score, out of the two wrong choices (both strategies are just as good; immediate judgments) was computed and tested against the right answers (delayed judgments). A Friedman test for ordinal data, a non-parametric equivalent to repeated measures ANOVA, yielded no main effect of session on the preferred choice for making JOLs, $\chi^2(4) = 3, p = .56$.

**Discussion**

The present research aimed to investigate students’ knowledge of metacognitive strategies and whether they possess, either explicit or naïve (i.e. having a feeling that delaying the judgments gives them better diagnostic abilities, without having explicit knowledge about it), knowledge
of the delayed JOL effect. An additional purpose was to investigate whether prolonged experience with the task of making judgments and testing the accuracy of these judgments with a cued recall test, would sway participants who show rather poor knowledge of metacognitive strategies, to change their preference (i.e., realize that the cue-only delayed JOLs have much higher predictive accuracy) in favor of the delayed JOL strategy.

Since the original article by Nelson and Narens (1991) that first showed the much higher predictive validity of delayed JOLs, there has been a myriad of literature corroborating their findings (e.g., Connor, Dunlosky & Hertzog, 1997; Dunlosky & Nelson, 1992, 1994; Schwartz, 1994; Koriat, & Ma'ayan, 2005; Nelson & Dunlosky, 1991; Sikström & Jönsson, 2005; Thiede, Anderson, & Therriault, 2003), while adding new knowledge about metacognitive judgments in the process. While Jönsson and Kerimi (2011) introduced something new to the field of metacognitive research when they targeted students’ knowledge about optimal monitoring strategies, their methodology, which was necessary to see how the students updated their knowledge on an item-by-item basis, was not perfectly optimized to calculate for example, the JOL-recall correlation. In this study, I set out to replicate their experiment, but precautions were taken for improving the methodology, mostly by having forced strategy choices, as well as only having cue-only judgments (iJOLc and dJOLc) and four consecutive sessions. Also, the stimulus lists were carefully selected with regard to discriminability of the word-pairs, and instructions were designed with emphasis on memory monitoring.

Consistent with several decades of metacognitive research, the delayed JOL effect (Nelson & Dunlosky, 1991; see Rhodes & Tauber, 2011, for a review) was replicated. This was important, as it is a fundamental premise for the present inquiry. If the delayed JOL effect had replicated, it would have been hard to argue for the dJOLc strategy as superior over the iJOLc strategy. While the delayed JOLs did show to be associated with a significantly higher predictive validity, students seemed to be unaware of which is the best strategy for making judgments of learning. Such poor knowledge of metacognitive strategies is in line with results from previous studies (Jönsson & Kerimi, 2011; Kornell & Son, 2009). The students’ beliefs about the accuracy of their judgments were not significantly affected by the type of judgment or by extended experience with the task. Also, contrary to predictions, having four consecutive sessions did not result in improved knowledge, about which is the best strategy for making judgments of learning. This result is in line with previous research by Hertzog and colleagues (Hertzog & Dunlosky, 2000; Hertzog et.al., 2009) as well as Jönsson and Kerimi (2011), even though they only had two session and had argued that a possible change in choice of strategy might have occurred with additional sessions.

With regard to the methods used, one could argue that it is unusual for students to try to focus on memory monitoring, rather on actually focusing on learning the material to be learned. If that is the case, yet this is only a speculative thought, it is possible that students did concentrate on memorizing the word-pairs, disregarding the instructions designed to put emphasis on monitoring, therefore not changing their preference for metacognitive strategy. Another thinkable caveat of the study could be that, the participants did not really operate as usual, while they were participating in a laboratory experiment, which is, no matter how carefully designed the experiment is, not their usual setting while learning. Outside the laboratory, while learning new material, it is possible that different factors influence the study choices, as compared to a one-hour laboratory experiment with no real consequences. Given the results, my conclusion from this experiment is, that there was no general consensus among the students about which the best judgment strategy is, as they were generally unaware of the
right answer. It is also apparent, and contrary to predictions, that experience with the task did not help to elicit a change in this unawareness. These results are in contrast with previous finding by Kornell and Son (2009), but are in line with other findings reported by Hertzog and colleagues (2009), as well as Jönsson and Kerimi (2011).

Another caveat might be that, even though great measures were taken for optimizing the instructions, the participants were still thinking in terms of “learning and memorizing” and not in terms on “judging and predicting ongoing learning”. To them, it might seem somewhat abstract, odd, or even hard to attain, to think in terms of monitoring and not memorizing, since this is what one does as a student, one tries to memorize new material. Yet, as demonstrated by Metcalfe and Finn (2008), for students it is rather important to keep both, memorizing and judging, in mind while learning new material. A good memory strategy will lead to improved learning, while using optimal monitoring strategy will lead to better study time allocation, thus, also playing a part in better learning. Unpublished data from our laboratory seems to indicate that the way the instructions of the JOL prompt are phrased might play an important role for the realization of the predictive superiority of delayed JOLs. In a similar experiment the instructions in the JOL prompts were phrased in a way that puts focus on current state of learning (i.e., “…how sure are you that you, right now, remember the association between the two words right now…”), instead of predicting one’s learning. This small change in phrasing elicited a significant shift in preference for monitoring strategy in favor of the dJOLc. A possible explanation for the observed shift in preference could be that the focus on current learning might lead participants to think in terms relative to judging their own learning, thus paying more attention to memory monitoring, and as a result, changing preference with regard to which the most efficacious monitoring strategy it. So, future studies about metacognitive knowledge, should look in to the effects of different phrasing. Also, it is possible that future experiments might benefit from immediate feedback displayed on the computer screen during the cued recall test, as a way to present the participants with indirect measure of JOL accuracy.

Students are individuals who presumably spend large portion of their time learning new material and preparing for exams, in order to improve their learning, it seems that is rather important to be able to distinguish between optimal and sub-optimal strategies. For example, poor judgments lead to poor study-time regulation (Kornell & Bjork, 2007). Hence, even if the student resorts to using a proper memory strategy the outcome may be inferior, if it is not combined with a proper monitoring strategy. Optimal study choices are highly dependent on a combination of accurate monitoring of ongoing learning, a realistic mental representation of how learning happens, as well as appropriate use of study strategies. Therefore, knowledge about which memory strategies result in better learning is important (Hertzog et. al., 2009), but so is knowledge about which strategy is optimal for accurately assessing one’s ongoing learning. For efficacious, optimized learning to ensue, it is necessary for the student to have knowledge, or at least a feeling, about which metacognitive strategies allow for more diagnostic monitoring (Pintrich, 2002).

The delayed JOL effect is a very robust effect, and while it accounts for only a modest memory boost (Rhodes & Tauber, 2011), it is easily the most diagnostic strategy for monitoring one’s ongoing learning. Yet, what good does it make if students, the people who would benefit the most from this knowledge, show such blunt unawareness of it. It is of little use if only researches have understanding about the highly reliable, diagnostic abilities of delayed JOLs and other facets of effective metacognition, and this is something that should be taken in to consideration by authors of study books, as well as teachers and lecturers. The
efficacy of various instructional approaches and techniques that students might not be aware of, such as delaying JOLs, the benefits of spacing learning, effective study time allocation, and the benefits of testing for subsequent retention should be discussed together with the more known learning strategies. This is where I find the present research to be highly relevant, as well as needed, since it is a step in the right direction to helping students become more knowledgeable and responsible about their own cognition and thinking.

Finally, people are obviously not solely guided by their assessments of the difficulty of the events. Other factors such as person’s interest and motivation may play part in study time allocation, as much as metacognitive judgments. Yet, one thing is undeniable, every single day most humans think about their own thoughts and make real decisions on how to act, based on these metacognitions. Becoming more aware of their existence and of the way they are responsible for memory monitoring should be of great benefit for everyone, not only in an educational context, but in everyday situations as well.

References


