Reference:

Ackerman, R., Leiser, D., & Shpigelman, M. (2013). Is comprehension of problem solutions resistant against misleading heuristic cues? *Acta Psychologica, 143*(1), 105-112.

Is Comprehension of Problem Solutions Resistant to Misleading Heuristic Cues?

Rakefet Ackerman[1], David Leiser[2], and Maya Shpigelman[1]

[1] Technion – Israel Institute of Technology

[2] Ben-Gurion University

Running Head: Comprehension of Problem Solutions

Corresponding author:

Rakefet Ackerman

Faculty of Industrial Engineering and Management

Technion – Israel Institute of Technology

Technion City, Haifa, Israel, 32000

Tel. +972 4 8294438, Fax +972 4 8295688

E-mail address: ackerman@ie.technion.ac.il

Abstract

Previous studies in the domain of metacomprehension judgments have primarily used expository texts. When these texts include illustrations, even uninformative ones, people were found to judge that they understand their content better. The present study aimed to delineate the metacognitive processes involved in understanding problem solutions—a text type often perceived as allowing reliable judgments regarding understanding, and was not previously considered from a metacognitive perspective. Undergraduate students faced difficult problems. They then studied solution explanations with or without uninformative illustrations and provided judgments of comprehension (JCOMP). Learning was assessed by application to near-transfer problems in an open-book test format. As expected, JCOMPs were polarized—they tended to reflect good or poor understanding. Yet, JCOMPs were higher for the illustrated solutions and even high certainty did not ensure resistance to this effect. Moreover, success in the transfer problems was lower in the presence of illustrations, demonstrating a bias stronger than that found with expository texts. Previous studies have suggested that weak learners are especially prone to being misled by superficial cues. In the present study, matching the difficulty of the task to the ability of the target population revealed that even highly able participants were not immune to misleading cues. The study extends previous findings regarding potential detrimental effects of illustrations and highlights aspects of the metacomprehension process that have not been considered before.

Keywords: Problem solving; metacognitive monitoring; judgment of comprehension; illustrations

1. Introduction

Traditionally, research into metacognitive monitoring and control processes has taken place within the domain of memory (e.g., Koriat & Goldsmith, 1996). Accordingly, studies have used tasks requiring participants to memorize itemized stimuli, such as lists of paired associates. These studies have undoubtedly enhanced our understanding of the metacognitive processes involved in learning (e.g., Metcalfe & Finn, 2008; Nelson & Narens, 1990). While this paradigm has ecological validity for some tasks, such as learning definitions of concepts (Wahlheim, Finn, & Jacoby, 2012), it clearly does not encompass the entire scope of metacognitive processes. Indeed, studies involving text learning have exposed factors that affect metacognitive processes but are not relevant when memorizing itemized stimuli. For example, text coherence was found to enhance predictions of performance in tests (Rawson & Dunlosky, 2002), and writing delayed summaries was found to improve the relative accuracy of metacomprehension judgments (Anderson & Thiede, 2008).

Another factor of unique relevance to texts is the inclusion of concrete supplements such as graphical illustrations and specific examples (Clark & Paivio, 1991; Sadoski, Goetz, & Fritz, 1993). Research into the design of study materials has repeatedly found that well-constructed supplements, judiciously applied, improve learning outcomes, while other supplements do not benefit and may even hinder learning (Ainsworth, 2006; Carney & Levin, 2002; Mayer, 1999; Prangsma, Boxtel, Kanselaar, & Kirschner, 2009). As for the subjective experience, concrete supplements have been found to enhance the subjective comprehensibility of texts, in addition to their contribution to recall (Sadoski et al., 1993). To examine the effects of such supplements on the subjective experience when they are dissociated from their actual contribution to the acquired knowledge, previous studies have used uninformative supplements, which did not contribute new information. For example, uninformative photos attached to headlines of false news items increased the likelihood that people would “identify” the items as known beforehand (Strange, Garry, Bernstein, & Lindsay, 2011). The only study focusing on the effects of illustrations on metacomprehension judgments was conducted by Serra and Dunlosky (2010). They used a text explaining how lightning storms develop. In their Experiment 2, one group studied the text in its base, textual only, version, the second group studied the same text accompanied by informative diagrams, and a third group received the text accompanied by uninformative photos. Immediately after learning each paragraph of the text, the participants rated the likelihood (0-100%) that they would be able to answer questions based on the studied information. It was found that while the informative diagrams enhanced both judgments and performance at test relative to the base text, the uninformative photos led to significantly higher judgments while performance rose only slightly. No differences were found between the groups in the invested study time. Together, these studies suggest that concreteness, like coherence, is a heuristic cue for metacognitive judgments regarding one’s knowledge that might mislead the learner in cases of uninformative supplements.

The existing literature in the context of metacomprehension has used expository texts almost exclusively, while a few studies used narratives (see Lin & Zabrucky, 1998 for a review). In the present study, we used a task novel to metacognitive theorizing but common in educational settings; namely, studying solution explanations for difficult logic problems. Overall, problem solving tasks are rarely studied in the metacognitive literature, although there is an increasing interest in recent years (Ackerman & Zalmanov, 2012; Thompson, Prowse Turner, & Pennycook, 2011; Thompson et al., in press; Topolinski & Reber, 2010). Still, none of the existing studies have used solution explanations.

Texts presenting problem solutions are of major educational importance, because students are often challenged to solve a problem and are then presented with an explanation of the solution (Atkinson, Derry, Renkl, & Wortham, 2000). These texts are often accompanied by illustrations, that are sometimes uninformative and merely decorative (see Carney & Levin, 2002). Mayer (1999) provided a thorough review of encouraging evidence that multimedia learning can promote constructivist learning, enabling problem-solving transfer. However, the reviewed studies focused on beneficial supplements and the metacognitive aspects were not considered. It is therefore unknown how people monitor their knowledge and regulate their efforts in the case of uninformative illustrations incorporated in solution explanations.

As a first step in the metacognitive analysis of learning solution explanations, the aim of the present study was to examine whether feelings of understanding are also prone to superficial heuristic cues and the consequences of their bias. By comparing explanations with and without uninformative illustrations, we aimed to consider several features of the metacognitive processes that were concealed when considering expository texts.

First, expository texts include a collection of facts, with only weak implicational connections. Part of the text may therefore be understood, while others are not, or to a lesser extent. This manifests itself in a normal distribution of the metacomprehension ratings, that exhibit medium or medium-high judgments on average (e.g., Serra & Dunlosky, 2010; Thiede, Wiley, & Griffin, 2011). In contrast, solution explanations require the participant to develop an integrative, holistic grasp of a complex system of relationships (Luo & Knoblich, 2007). The struggle to solve a problem is presumably accompanied by a sense of cognitive tension that is resolved when the solution becomes clear. A feeling of insight is often experienced at this point, marking the closure of an integrative structure, in the terminology of Piaget (1985). The accompanying sense of certainty can arise both when one finds the solution on one's own and when one has it explained (Metcalfe & Wiebe, 1987; Trout, 2002). Subjective feelings regarding understanding solutions are therefore expected to be polarized, reflecting either good or poor understanding of the cohesive explanation.

Second, as explained above, most metacognitive studies have focused on memorization and recall of details. The challenge of understanding explanations, in contrast, directs learning to high-order understanding, which is the ultimate educational purpose (Kintsch, 1998). Instead of testing participants’ knowledge of the studied materials per se, we tested their understanding by using near-transfer problems. This approach requires knowledge implementation of the types required in educational and real-life settings (Atkinson et al., 2000; Sasson & Dori, 2012).

Third, scientists and laypeople alike use the sense of understanding that an explanation produces as a cue to a good or correct explanation, and it hard for them to doubt the reliability of this feeling (Trout, 2002). This is even indicated by the Oxford Dictionary in its definition “insight: the capacity to gain an accurate and intuitive understanding" (cf. Soanes & Hawker, 2005, p. 1264). Considering other metacognitive biases when studying from texts (e.g., Rawson & Dunlosky, 2002) raises the possibility that the strength of the sense of understanding does not ensure its reliability or its resistance to misleading cues. Such biases are not epiphenomena. Learners, who unjustifiably sense that understanding has taken place, are expected to be misguided in their learning regulation (de Regt, 2004; Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003). As mentioned above, Serra and Dunlosky (2010) did not find effects of the uninformative illustrations on performance at test. Because we anticipated polarized judgments, and a biasing effect of the uninformative illustrations on them, we expected the illustrations to shift feelings of not understanding to unjustified feelings of understanding. We also hypothesized that the strength of this bias would be more pronounced than that found with expository texts and would harm the effectiveness of studying the solution explanation of a similar problem. Thus, we predicted lower success rates after studying from illustrated explanations than after studying from plain explanations.

In addition to the examination of the metacognitive aspects of learning problem solutions, we address a broader question, regarding individual differences. Both the initial and the transfer problems had to be challenging for the explanations to afford new insights for the participants to acquire and potentially make a difference in their chance for success (Merrill, 2002). The use of challenging tasks allowed us to consider the role of general ability in the susceptibility to misleading heuristic cues. Previous studies have suggested that relatively weak participants are more vulnerable to being swayed by superficial cues (Evans, Handley, Neilens, & Over, 2010; Novick & Sherman, 2008;

Thiede, Griffin, Wiley, & Anderson, 2010). However, it is not clear whether it is a lower general cognitive ability that underlies this vulnerability or the extent of the challenge imposed by the task. A well-established idea regarding numerous judgments (Tversky & Kahneman, 1974) is that a subjective state of uncertainty increases the susceptibility of judgments to misleading heuristic cues. This suggests that uncertainty, rather than general ability, may underlie the vulnerability to the misleading cues. With this in mind, we examined whether the biasing force of the illustrations is associated with general ability (as indicated by SAT scores) or pose a hurdle for all when facing challenging tasks.

2. Overview of the experiments

In the two experiments reported below, participants attempted to solve challenging problems. Immediately after trying to solve each problem, they read an explanation of its solution. They then faced a near-transfer problem that dealt with the same topic (see Appendix). In a pretest, a sample of the same population solved the initial and the transfer problems without guidance. This way, we could verify that all the problems are indeed challenging, and use the pretest success rates in the transfer problem as a baseline for examining the benefit gained by learning the explanations for the first problems before solving them.

The explanations, which served as the target texts, appeared either in base (unillustrated) versions, or with concrete illustrations of objects mentioned in the text (see example in the Appendix). By presenting each participant with several problems and explanations, half in their base versions and the others illustrated, we focused on the effects of illustrations on the metacognitive processes, and ruled out individual differences besides those in the study's focus.

Past metacognitive studies using two ratings, one aimed at memory for details and another at comprehension, have found that participants differentiate between the two in their metacognitive judgments (Ackerman & Goldsmith, 2011; Zaromb, Karpicke, & Roediger, 2010). We therefore specifically collected a judgment of comprehension (JCOMP; cf. Zaromb et al., 2010) on a 0-100% scale, immediately after participants had a chance to study each solution explanation, and before they faced the transfer problem.

Further, previous studies have suggested that open-book testing encourages higher-level processing during study, as compared to closed-book testing (see Heijne-Penninga, Kuks, Hofman, & Cohen-Schotanus, 2011, for a review), and allows the learning process to continue during test taking (Agarwal, Karpicke, Kang, Roediger Iii, & McDermott, 2008). This focus presumably also directs metacognitive monitoring towards assessing high-order understanding (Thiede et al., 2011). Accordingly, we utilized the opportunity provided by using a transfer task, rather than recall of information that can be directly found in the text, and employed an open-book test format.

The following hypotheses guided the data analyses. First, building upon the above mentioned previous findings, we hypothesized that the presence of uninformative illustrations would heighten JCOMP ratings for solution explanations. However, as mentioned above, JCOMPS were expected to deviate from the normal distribution and tend to extremes. When people experience poor understanding (PU) there is room for JCOMP to rise, but how can the presence of illustrations raise high JCOMPS that reflect certainty and are already close to the ceiling? We expected that because of the tendency to extremes, illustrations would transfer PU states to good understanding (GU) states; that is, increase the GU frequency. Because we could not know how to relate continuous JCOMPs (e.g., 80%) to one's subjective experience of PU or GU, to help the analyses we asked participants to also provide a PU or GU indication. We then looked into the effects of illustrations on JCOMP in cases of PU and GU separately, in addition to their overall effect.

Second, inflated JCOMPs were expected to lead participants to wrongly assess their level of knowledge as satisfactory (Nelson & Narens, 1990; see Winne, 2004 for a review). As mentioned above, Serra and Dunlosky (2010) did not find consequences of the inflated monitoring. In our case, we predicted a detrimental effect on performance, due to the anticipated transition from PU to GU. Thus, a finding of performance reduction when uninformative illustrations are incorporated in solution explanations would indicate that their misleading effect is even stronger in this case. The hypothesized opposite effects on JCOMP and success rates should increase the gap between them, yielding a larger bias[1], which may misguide future regulatory decisions, such as whether restudying the materials is needed (Metcalfe & Finn, 2008).

Finally, we collected participants' SAT scores for analyzing the interplay between metacognitive monitoring and performance, while taking into account the predictive value of cognitive ability. SAT scores were used in addition to the overall level of

[1] It should be noted that the mere gap between JCOMP and performance on the transfer problems (i.e., the extent of overconfidence) is not of interest in the present study, because the participants were not required to predict their performance. The important question is whether the gap is larger for the illustrated versions than for the base versions.

performance in the transfer problems, for examining both general and task-specific abilities.

3. Experiment 1

Experiment 1 examined the hypotheses detailed above by the array of measures collected: JCOMP ratings, GU or PU indications, success in the transfer problems, time devoted to each phase, and SAT scores.

3.1 Method

3.1.1 Participants

Forty-one undergraduate students in the social sciences from a university and a college in Israel participated in this experiment. Their reported Israeli SAT scores averaged 625.2 (*SD* = 71.5).[2]

3.1.2 Materials

Four logic problems were used. For each initial problem, there was a base version of the solution and an illustrated version with the same wording; the latter was accompanied by a drawing of objects mentioned in the text but provided no new information for solving the problem (see example in the Appendix). Each solution was followed by a near-transfer problem dealing with the same topic and designed to be easier after understanding the solution to the first problem at the pair. Pretesting ($N = 20$) verified that the initial and transfer problems were difficult for the target population (less than 20% correct) with no significant difference between them, $t < 1$. As mentioned above, this pretest also provided the base-line success rate for assessing learning from the explanations.

[2] Israeli SAT scores range from 200-800; the national mean is 540; 15% of test-takers nation-wide score above 650; 5% score above 700.

3.1.3 Procedure

The participants were invited in small groups and seated in a classroom. Each participant faced all four problems successively, with two explanations in their base versions and two in their illustrated versions, counterbalanced and randomly ordered for each participant. The experiment was conducted on paper, and the entire procedure was explained in advance.

The procedure consisted of three phases for each problem. First, participants were instructed to do their best to solve the initial problem on their own. Then they read the solution explanation. Immediately afterward, participants rated their JCOMP by responding to the question, “To what extent do you understand the solution?” Participants marked their judgment by drawing a vertical line on a 0-100% horizontal scale. They were also asked to indicate either GU or PU, and to indicate whether they had known the problem and its solution beforehand. Finally, the relevant transfer problem was on the following page. Participants worked at their own pace, moving from one problem to the next, and documented the starting time for each step and the final ending time in dedicated fields on the forms. The text for each phase of the procedure appeared on a separate sheet of paper, and the pages were prearranged in a pile for each participant, upside down. The participants picked up one sheet of paper at a time.

3.2 Results

Overall, 3% of the problems were reported to be known in advance, and answers for initial and transfer problems were indeed correct. These problems, one for each participant, were removed from the analyses.

3.2.1 Time Regulation

The time invested in studying the illustrated solutions (1.9 min.) was similar to the time invested in the base versions (2.3 min.), $t(40) = 1.53$, $p = .13$, $d = 0.24$. Participants took 3.6 minutes on average to solve each transfer problem, with no difference between the solution versions, $t < 1$.

3.2.2 Success Rate

A judge blind to the experimental manipulation graded the solutions for the transfer problems. The solutions were scored at either 100 (for a full and correct solution) or zero (for any incorrect solution).[3] Fig. 1A shows the mean for each of the two solution versions. The aggregated success rates for the two transfer problems each participant faced after studying each version could be 0, 50, or 100, which required using a-parametric tests. Scores on the transfer problems were elevated relative to the pretest ($M = 12.1\%$, $SD = 13.5$), meaning that participants benefitted from working on the initial problems and learning the solution explanations in both versions (both *p*s < .01 by Wilcoxon signed-rank test). Despite the expected tendency, the success rate difference between the versions was not significant, Wilcoxon's $Z = 1.36$, $p = .18$. Thus, the versions did not differ by either of the two objective measures of time and performance.

[3] Using a three-tiered scoring system (correct, partially correct, and wrong) did not affect the pattern of results but inserted ambiguity into the scoring. To confirm the reliability of the grades, another judge re-graded 20% of the problems. The grades of the two judges matched perfectly.

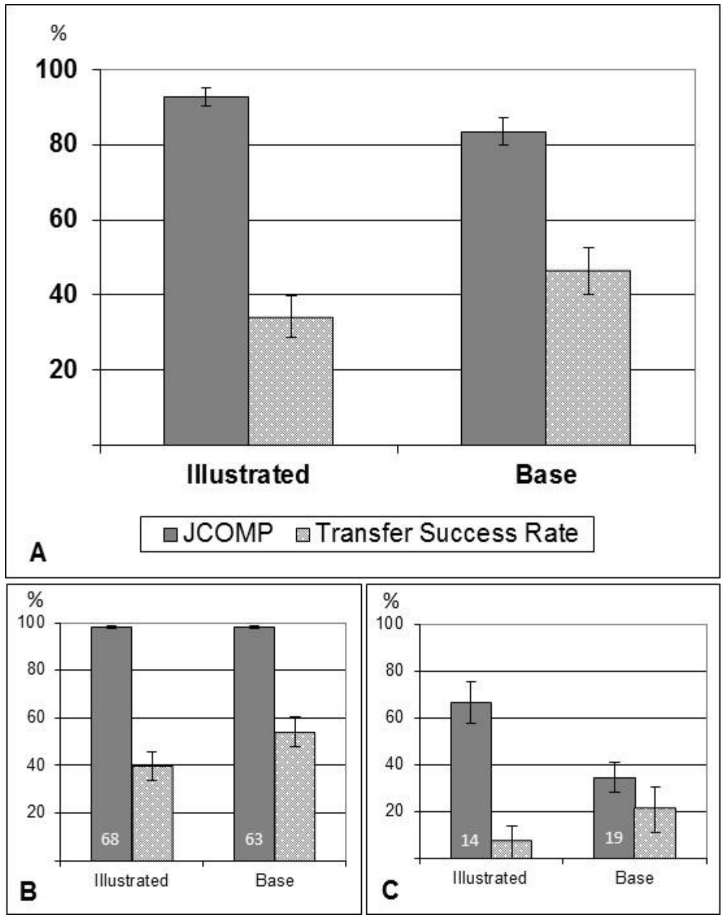

Figure 1. Experiment 1: Mean judgment of comprehension (JCOMP) ratings for the illustrated and base solution versions and success rates on the transfer problems. (A) Results for all the problems; (B) Results for solutions for which the participants indicated good understanding; (C) Results for solutions for which the participants indicated poor understanding. The white numbers on the bars represent the number of problems (N) represented by the respective JCOMP and success rate bars. Error bars represent standard error of the mean.

3.2.3 Judgment of Comprehension

As expected, the JCOMP distribution deviated significantly from the normal distribution (one-sample Kolmogorov-Smirnov $Z = 1.93$, $p = .001$ and $Z = 1.654$, $p < .01$ for the illustrated and base versions, respectively). The distribution was skewed to the right, with 6% of the JCOMP ratings below 10% and 78% above 90%, and the rest spread along the scale. As predicted, JCOMP ratings were higher for the illustrated than the base versions, Wilcoxon's $Z = 2.63$, $p < .01$ by a Wilcoxon signed-rank test (see Fig.

1A). The distribution of the gap between JCOMP ratings and success rates did not deviate from normality for both versions. A larger gap was found for the illustrated than for the base versions, $t(40) = 2.76$, $p < .01$, $d = 0.43$.

The frequency of GU and PU can be seen in Fig. 1B and Fig. 1C. Overall, there was a tendency to report GU, but, in contrast to our prediction, the distribution did not differ between the versions, $\chi^2(1) = 0.95$, $p = .33$. Participants' JCOMP ratings for GU solutions were near the ceiling for both versions (see Fig. 1B). Regarding the PU solutions, JCOMP ratings were higher for the illustrated than for the base versions (see Fig. 1C)[4], suggesting that this set of problems underlie the significant difference in JCOMP. An interesting observation is that for the poorly understood illustrated solutions, the success rate was particularly low—close to zero—and this was not reflected in the JCOMP ratings. The poorly understood base solutions allowed the most judicious JCOMP ratings.

To test whether the effect of illustrations on the gap between JCOMP ratings and success rates was associated with the overall success rates of the participants, we examined the effects of Achievement (continuous; mean success rate in all the transfer problems) × Text Version (illustrated vs. base) on this gap using a General Linear Model (GLM or ANCOVA). The interaction between these factors was not significant ($F < 1$), suggesting no change in the JCOMP-success rate gap in association with achievement level. The correlation between SAT scores and success rates was significant ($r = .59$, $p < .0001$). A similar analysis based on SAT scores revealed highly similar results, with $F < 1$ for the interactive effect. These findings suggest that in contrast to previous findings

[4] A within-participant t-test could not be performed for this data because most participants did not experience either GU or PU for both versions. Therefore the report of comparisons between the versions within either GU or PU is descriptive.

with other tasks, including uninformative illustrations in problem solutions affected JCOMP deviation from applicable knowledge across the entire sample, regardless of participants' achievement level or cognitive ability.

3.3 Discussion

The findings of Experiment 1 make it clear that subjective feelings of understanding regarding problem solutions are not resistant to the misleading force of superficial heuristic cues, as some might have expected. The results suggest that the inclusion of uninformative concrete illustrations affects JCOMP and its bias from actual success rates, especially when understanding is poor.

However, our prediction of performance reduction was not supported by a significant difference, nor did the frequency of GU increase. Experiment 2 extended the investigation in an attempt to further examine these aspects and to delve into the question of the effects of students' ability on their metacognitive processes.

4. Experiment 2

As explained above, previous studies have found weaker students to be more affected by superficial cues (Novick & Sherman, 2008; Thiede et al., 2010), but Experiment 1 did not replicate this finding, as participants along the entire range of ability were misled similarly by the inclusion of the illustrations. These seemingly conflicting findings can be reconciled by the uncertainty explanation: perhaps the stronger students were found to be less susceptible to superficial cues in previous studies because the tasks were relatively easy for them. According to this explanation, for people who experience high certainty in understanding all explanations, there is little room for the effects of the illustrations.

An alternative interpretation would assert that we did not sample the whole range of ability in Experiment 1, that indeed the entire sample of Experiment 1 should count as weak, and that for sufficiently strong students no or little bias would be caused by the illustrations. To examine this possibility, we drew the sample for Experiment 2 from selective engineering and science programs and matched the task to their abilities. Using the same pretesting procedure as for Experiment 1, we constructed a problem set designed to challenge these students according to the same difficulty criterion as for Experiment 1 (success rate < 20%). Eight problems passed this criterion, but the problems used in Experiment 1 were not among them. This preliminary finding supports our intention that the present population is stronger at solving this type of problem. As in Experiment 1, we used the success rates of the pretest participants as a baseline for examining learning effectiveness.

Recently, Ackerman and Lauterman (2012) found, with a similarly strong sample, that mild time pressure exposed monitoring biases that were otherwise concealed. Thus, we added time pressure manipulation (between participants) to examine the possibility that the stronger students would not be affected by the illustrations when given ample time, while time pressure would expose their vulnerability.

4.1 Method

4.1.1 Participants

Fifty-one undergraduate students, all majoring in engineering or science, were randomly assigned to ample (N = 21) and limited time (N = 30) conditions. Their reported SAT scores (M = 686.5, SD = 36.3) were significantly higher than in

Experiment 1, $t(90) = 5.34$, $p < .0001$, $d = 1.13$. It is worth mentioning that this sample is in the top 10% of the entire population.

4.1.2 Materials

Eight problem pairs, comprising an initial problem and a near-transfer problem, similar in type to those used in Experiment 1, were chosen by pretesting ($N = 32$) to challenge the target population (success rate < 20%).

4.1.3 Procedure

The procedure for each problem was the same as for Experiment 1. Each participant faced all eight problems, four in illustrated versions and four in base solution versions, counterbalanced and randomly ordered for each participant. For the time pressure group, the experimenter announced the time every ten minutes and reminded participants that they should finish all eight problem pairs within the allotted time.

4.2 Results

One problem was known in advance to one participant; this data was removed from the analysis. Four participants in the time pressure group managed fewer than seven problems. This result suggests that these participants did not comply with the time pressure instruction when solving each problem, and thus these participants were replaced. Overall, the participants solved 96% of the problems (7.7 problems) on average within the time frame. In contrast to our expectations, no effect was found for the time condition on any of the measures, except that participants in the ample time condition worked longer on each problem ($p < .0001$). Thus, this factor is not included in the results report.

A preliminary analysis revealed that for two of the eight problems, success at the transfer problem was improved, relative to the pretest, by studying the explanations. Although the overall pattern of results was not affected by their inclusion, these problems were removed from the analysis, as they did not reflect the intended learning process.

4.2.1 Time Regulation

Solution study time for the illustrated versions (1.38 min.) was not significantly shorter than for the base versions (1.52 min.), $t(50) = 1.42$, $p = .16$, $d = 0.20$. No difference was found in the time devoted to solving the transfer problems (3.7 min.) between the solution versions, $t < 1$. Thus, again, there was no effect of the solution version on time regulation.

4.2.2 Success Rate

As intended, success rates were low (see Fig. 2A). In the present experiment, the success rate for the illustrated versions was lower than for the base versions, $Z = 2.0$, $p < .05$. Combining the data from the two experiments supported our hypothesis of general performance reduction, $Z = 2.1$, $p < .05$. Interestingly, a comparison to the results of the pretest ($M = 9.7\%$, $SD = 14.3$) revealed a difference between the versions. Success rates after studying the illustrated solutions were equivalent to the success rates without studying them ($M = 16.0\%$, $SD = 23.7$), $Z = 0.85$, $p = .40$, while exposure to the base solutions improved the success rates ($M = 26.5\%$, $SD = 28.1$), $Z = 2.84$, $p < .01$. Thus, exposure to the illustrated solutions for the initial problems did not promote performance on the transfer problems, although the explanations had beneficial potential, as shown with the base versions.

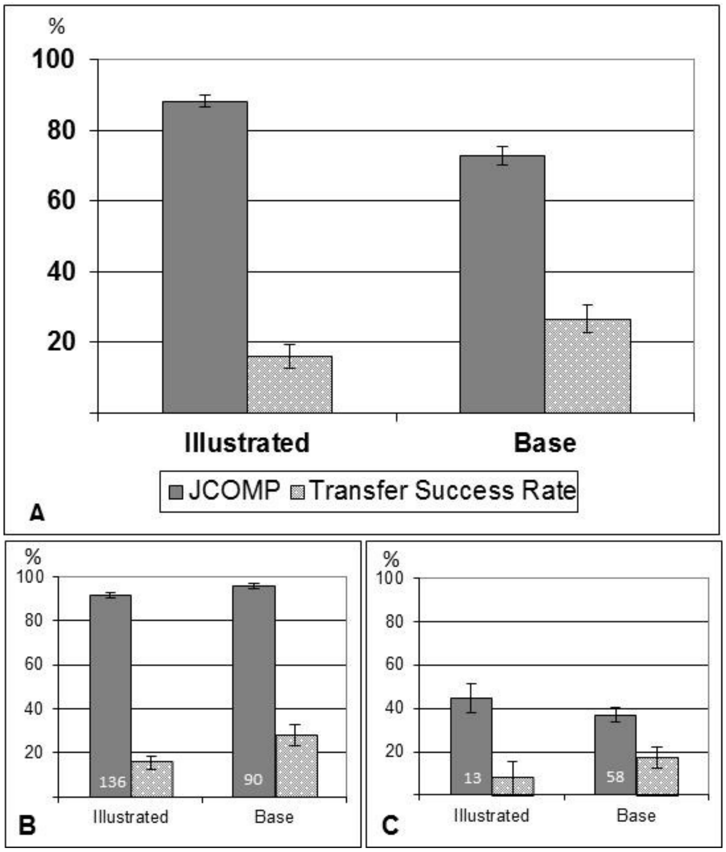

Fig. 2. Experiment 2: Mean judgment of comprehension (JCOMP) ratings for the illustrated and base solution versions and success rates on the transfer problems. (A) Results for all the problems; (B) Results for solutions for which the participants indicated good understanding; (C) Results for solutions for which the participants indicated poor understanding. The white numbers on the bars represent the number of problems (N) represented by the respective JCOMP and success rate bars. Error bars represent standard error of the mean.

4.2.3 Judgment of Comprehension

As in Experiment 1, the JCOMP distribution was again skewed to the right, *p*s < .0001, with 7% of the JCOMPs below 10%, 61% above 90%, and the rest spread along the scale. Again, JCOMP ratings were higher for the illustrated than for the base versions, $Z = 4.18$, $p < .0001$. This, on top of the lower performance for the illustrated versions, resulted in a larger gap between JCOMP ratings and success rates for the illustrated versions, $t(50) = 4.82$, $p < .0001$, $d = 0.67$.

Importantly, unlike in Experiment 1, division of the problems into GU and PU showed that GU was more frequent and PU less frequent for the illustrated than the base versions (see Fig. 2B and Fig. 2C), $\chi^2(1) = 37.88$, $p < .0001$. This finding supports our hypothesis regarding the second source of JCOMP increase with the illustrations. Although, in this case as well, the most judicious JCOMP was found in the case of poorly understood base solutions.

In examination of the effect of participants' overall ability, a GLM as above again revealed no interactive effect with the solution version, $F < 1$. The correlation between SAT scores and success rates was again significant, although somewhat weaker than in Experiment 1 ($r = .34$, $p = .01$). With SAT scores as the continuous factor in the GLM, the interactive effect was also not significant, $F(1, 49) = 2.92$, $MSE = 714.92$, $p = .14$, $\eta_p^2 = .04$. Thus, even among the strongest students, the biasing effect of the illustrations was found across the entire sample.

4.3 Discussion

The results of Experiment 2 support all our hypotheses and allow to decisively reject the notion that strong students are less prone to the misleading effects of illustrations when learning problem solutions.

5. General Discussion

Whereas most previous metacomprehension studies have examined learning of expository texts, the present study involved understanding solution explanations for hard problems. By choosing a text type novel to the metacognitive literature, we aimed both to illuminate aspects of metacognitive processes not previously considered, and to determine whether known effects are generalizable. The high-level question that drove the present study was whether the strong understanding experiences attendant the grappling with solution explanations are impervious to superficial heuristic cues. Our findings establish firmly that JCOMPs for solution explanations are not impervious to such misleading cues.

The distinction between GU and PU situations revealed that the higher JCOMPs for illustrated solutions stemmed from two sources: a) a continuous increase in JCOMP when people experience that they do not completely understand the explanations (both experiments), and b) a qualitative switch from PU to GU states that raises the JCOMP by increasing the frequency of top JCOMPs (Experiment 2). This upward bias was accompanied by lower performance, a finding not found in the only previous study with this manipulation (Serra & Dunlosky, 2010), and suggestive of a stronger misleading force of the illustrations in the present case.

Moreover, the effects of the illustrations were particularly salient among the most able students. Evidently, and in contrast to previous conclusions (Novick & Sherman, 2008; Thiede et al., 2010), individuals with very high cognitive ability are not immune to misleading heuristic cues. In addition, unlike the findings of Ackerman and Lauterman (2012), in this case, the pronounced metacognitive bias took place in the mostly able population even without the additional challenge of time pressure.

A larger JCOMP-success rate gap in the presence of illustrations is important when considering further decisions of the participants, such as whether to restudy an explanation or seek for help (Aleven & Koedinger, 2000; Metcalfe & Finn, 2008). This gap could have stemmed from the hard-easy effect—an underestimation of conditions that lead to low performance (e.g., Lichtenstein, Fischhoff, & Phillips, 1982; Suantak, Bolger, & Ferrell, 1996). This explanation would hold if JCOMPs had been lower when performance was lower, but not adequately low. This was not the case. The result of lower performance with significantly higher JCOMPs highlights the metacognitive illusion generated by the illustrations and expected to misguide future decisions. Notably, the smallest JCOMP-success rate gap was found in PU of solutions without illustrations.

Turning to the regulation of study time, in neither experiment was the time devoted to studying or solving the transfer problems affected by the presence of concrete supplements. Two observations should be made about this finding. First, according to the discrepancy reduction models for the regulation of study time (Nelson & Narens, 1990), upward-biased JCOMPs should have shortened the time invested in studying the solutions (see illustration in Ackerman & Goldsmith, 2011). That this was not found could result from an imperfect measuring methodology, or from a bias too weak to significantly affect the regulation of time, that is known to depend on additional factors (see Son & Sethi, 2010).

The second observation derived from analyzing the time regulation is that the presence of illustrations seems to increase cognitive load (Chandler & Sweller, 1991). The outcome was that success rates were lower for the illustrated solution versions despite the similar amounts of time invested. These findings suggest that learners who faced these versions did not act accordingly to improve their knowledge during learning and even later on, during the open-book test phase (Agarwal et al., 2008). Interestingly, the manipulation of time pressure in Experiment 2—ample versus limited—did not increase performance differences. Future studies could focus more on the ways people regulate their effort when they learn and apply solution explanations and other texts of similar complexity.

Overall, the findings support our contention that people are particularly prone to the misleading effects of illustrations when challenged by a task they experience as hard. However, two caveats are called for. First, for the sake of generality, future studies should examine directly the correlation between task difficulty and the susceptibility of JCOMPs to misleading cues. For example, in cases in which previous studied have found the stronger students to be more sensitive to reliable cues, a direct examination of the effects found with the weaker students should be examined with comparably hard tasks chosen to challenge the stronger students. Second, the results suggest that certainty does not ensure immunity to superficial heuristic cues. We found that the presence of illustrations increased the frequency of the combination of GU and top levels of JCOMP. The bias is therefore not limited to cases of subjective uncertainty (Tversky & Kahneman, 1974), as these cases were accompanied by certainty and the bias still occurred.

Our findings raise the broader question of how concrete supplements may harm performance in general. The study suggests that metacognitive processes play a mediating role: the supplements increase the concreteness of the text, which in turn leads to focusing on the surface level rather than on the deep structure of the subject matter (Smith, diSessa, & Roschelle, 1993). It is widely accepted that concrete representation is easier to understand and remember, because it can be more readily associated with prior knowledge (Kieras, 1978; Sadoski et al., 1993). In addition, graphical representation often does help to deal with understanding challenges by providing information that is hard to express verbally (Chen & Daehler, 2000). For example, when trying to follow instructions for an assembling task, graphical illustration may convey information regarding spatial relationships among the components that is expressed less effectively by verbal descriptions (Morrell & Park, 1993; Stone & Glock, 1981). Indeed, as mentioned above, a positive correlation was found between concreteness and comprehensibility (Sadoski et al., 1993). If people overgeneralize from informative concrete supplements to cases of uninformative ones, they may be led astray and think that they understand the information better than they do without these supplements. This, in turn, may prevent them from investing the extra effort required in the presence of the supplements (Chandler & Sweller, 1991) and lead to lower performance, as found here. Future studies could examine the mediating role of metacognitive processes in the cases of lower achievement in the presence of other concrete supplements to texts (e.g., Mayer, Heiser, & Lonn, 2001). For example, the study offers a possible explanation for the performance reduction found with seductive details—concrete details that make the text more interesting without contributing to the understanding of the main ideas—in addition to previously considered explanations (Garner, Gillingham, & White, 1989; Harp & Mayer, 1997; Sanchez & Wiley, 2006) .

In summary, the use of problem solutions as the target study materials, the focus on high order understanding, and the vulnerability of even the strongest students to misleading superficial heuristic cues are instructive in general, shed new light on metacognitive processes involved when facing daily tasks. The results reveal that illustrations can have detrimental effects on both learning and metacognitive estimations of learning, leading to the conclusion that the use of illustrations in problem-related study materials requires scrutiny.

The study may also provide directions for metacognitive instruction. For example, the IMPROVE methodology (Mevarech & Kramarski, 1997) includes questions that students use while learning (e.g., solving practice mathematical problems). The questions direct the students to make sure they understand the task and its association to the studied topic, to consider possible strategies for progress and the appropriateness of the answer they come up with. The effects of IMPROVE on problem solving achievements have been reported in a large number of studies, even among elementary school children (e.g., Mevarech, Terkieltaub, Vinberger, & Nevet, 2010). Another approach makes use of the advantages of computerized learning environments. For example, appropriate software can facilitate learning by presenting learning activities adapted to student's knowledge and by practicing self-assessment with feedback regarding actual performance (see Winne & Nesbit, 2009; Wong & Looi, 2012). It is interesting to examine whether these methods attenuate metacognitive biases of the type found in the present study and whether general experience and directions are enough, or more focused awareness to particular biases is required for overcoming them (e.g., Koriat & Bjork, 2006).

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied, 17*(1), 18-32.

Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in human behavior, 28*, 1816-1828.

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychonomic Bulletin & Review, 19*, 1187-1192.

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger Iii, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology, 22*(7), 861-876.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183-198.

Aleven, V., & Koedinger, K. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292-303). Berlin: Springer-Verlag.

Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica, 128*, 110-118.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181-214.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14*(1), 5-26.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chen, Z., & Daehler, M. W. (2000). External and internal instantiation of abstract information facilitates transfer in insight problem solving. *Contemporary Educational Psychology, 25*(4), 423-449.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

de Regt, H. W. (2004). Discussion Note: Making Sense of Understanding. *Philosophy of Science, 71*(1), 98-109.

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *The Quarterly Journal of Experimental Psychology, 63*(5), 892-909.

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6*(1), 41-57.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*(1), 92-102.

Heijne-Penninga, M., Kuks, J. B. M., Hofman, W. H. A., & Cohen-Schotanus, J. (2011). Directing students to profound open-book test preparation: The relationship between deep learning and open-book test time. *Medical Teacher, 33*(1), e16-e21.

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. *Psychological Bulletin, 85*(3), 532-554.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. New York: Cambridge.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning Memory and Cognition, 32*(5), 1133-1145.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to l980. *Judgment under uncertainty: Heuristics and biases*, 306-334.

Lin, L. M., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*, 345-391.

Luo, J., & Knoblich, G. (2007). Studying insight problem solving with neuroscientific methods. *Methods, 42*(1), 77-86.

Mayer, R. E. (1999). Multimedia aids to problem-solving transfer. *International Journal of Educational Research, 31*(7), 611-623.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43-59.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*(1), 174-179.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory and Cognition, 15*(3), 238-246.

Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*(2), 365-394.

Mevarech, Z. R., Terkieltaub, S., Vinberger, T., & Nevet, V. (2010). The effects of meta-cognitive instruction on third and sixth graders solving word problems. *ZDM Mathematics Education, 42*(2), 195-203.

Morrell, R. W., & Park, D. C. (1993). The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychology and Aging, 8*(3), 389-399.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125-173).

Novick, L. R., & Sherman, S. J. (2008). The effects of superficial and structural information on online problem solving for good versus poor anagram solvers. *The Quarterly Journal of Experimental Psychology, 61*(7), 1098-1120.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development* Chicago: The University of Chicago Press.

Prangsma, M. E., Boxtel, C. A. M., Kanselaar, G., & Kirschner, P. A. (2009). Concrete and abstract visualizations in history learning tasks. *British Journal of Educational Psychology, 79*(2), 371-387.

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69-80.

Sadoski, M., Goetz, E. T., & Fritz, J. B. (1993). Impact of concreteness on comprehensibility, interest, and memory for text: Implications for dual coding theory and text design. *Journal of Educational Psychology, 85*(2), 291-304.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*(2), 344-355.

Sasson, I., & Dori, Y. J. (2012). Transfer skills and their case-based assessment *The Second International Handbook of Science Education* (pp. 219-250). Netherlands: Springer.

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory, 18*(7), 698-711.

Smith, J. P., III, diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*, 115-163.

Soanes, C., & Hawker, S. (Eds.). (2005) Insight. Oxford University Press.

Son, L. K., & Sethi, R. (2010). Adaptive learning and the allocation of time. *Adaptive Behavior, 18*(2), 132-140.

Stone, D. E., & Glock, M. D. (1981). How do young adults read directions with and without pictures? *Journal of Educational Psychology, 73*(3), 419-426.

Strange, D., Garry, M., Bernstein, D. M., & Lindsay, D. S. (2011). Photographs cause false memories for the news. *Acta Psychologica, 136*(1), 90-94.

Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The hard-easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes, 67*(2), 201-221.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66-73.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*(4), 331-362.

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology, 81*, 264-273.

Thompson, V. A., Prowse Turner, J., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107-140.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (in press). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition.*

Topolinski, S., & Reber, R. (2010). Immediate truth-Temporal contiguity between a cognitive problem and its solution determines experienced veracity of the solution. *Cognition, 114*(1), 117-122.

Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science, 69*(2), 212-233.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Wahlheim, C. N., Finn, B., & Jacoby, L. L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: evidence for variability neglect. *Memory & Cognition*, 1-14.

Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research, 41*(6), 466-488.

Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 259-277). New York: Routledge.

Wong, L. H., & Looi, C. K. (2012). Swarm intelligence: new techniques for adaptive systems to provide learning support. *Interactive Learning Environments, 20*(1), 19-40.

Zaromb, F. M., Karpicke, J. D., & Roediger, H. L. (2010). Comprehension as a basis for metacognitive judgments: Effects of effort after meaning on recall and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 552-557.

Appendix

The Appendix presents a problem example with its three components: the problem, the solution, and the near-transfer problem. The illustration was included in the illustrated solution version but not the base version.

The socks problem

A well organized and busy lady works most of the day in court. For her work she must wear only black and white clothes. Thus, she has a special drawer in which she keeps her black and white socks. She knows that there are 18 sock pairs, all identical, of which eight pairs are white and the rest are black. They are all stored in the same drawer, arranged in pairs and separated by colors.

Today, when she returned home, her young child, who had been alone at home for the last couple of hours, surprised her. He covered her eyes and told her that he had rearranged her sock drawer. He made a piece of art using her socks that are now all mixed in the drawer. The child asked his mother to take a corresponding pair of socks while her eyes were covered.

Question:

What is the minimum number of single socks that the exhausted mother can select from the drawer with her eyes covered to ensure she has a matching pair of socks in her hand? Explain your answer.

The solution

The mother stood embarrassed in front of her child and thought about the problem. She recalled that she had 18 pairs of socks, which means 36 single socks.

She started by taking two single socks. If they were the same color, then she already had a matching pair, because all her socks were identical. Of course, she may have selected two differently colored socks. Thus, it is enough to select one more sock. Taking three socks would ensure that she has a matched pair of socks in her hand.

The transfer problem

Now, the child directed his exhausted mother to another drawer. In this drawer she kept her colored socks. He had “rearranged” this drawer as well. In this drawer there were 50 pairs of socks: 5 green, 10 yellow, 15 orange, and the rest blue.

Question:

What is the minimum number of single socks that the mother should select with her eyes closed, to ensure that she has a matching pair of socks in her hand? Explain your answer.