Does consulting with others affect answerability judgments of difficult questions?

Social Influence (ISSN: 1553-4510)

Citation for the published paper:
Allwood, C. ; Karlsson, B. ; Buratti, S. (2016) "Does consulting with others affect answerability judgments of difficult questions?". Social Influence, vol. 11(1), pp. 40-53.

http://dx.doi.org/10.1080/15534510.2016.1140675

Does consulting with others affect answerability judgments of difficultquestions?

Carl Martin Allwood, Bodil S. A. Karlsson and Sandra Buratti
Department of Psychology, University of Gothenburg, Göteborg, Sweden

ABSTRACT

People's judgments of the answerability of questions relating to how things are in the world can have important consequences for society and people's lives. Thirty-one individuals and 30 pairs made answerability judgments of 20 general knowledge questions, many with less known, or unknown, answers. Four questions had high expected consensus regarding their answerability (consensus questions) and the rest had less expected consensus with respect to their answerability (non-consensus questions). The pairs showed two polarization effects: pairs gave higher answerability ratings for questions with answerability ratings over 80% and lower ratings than individuals for questions with the lower answerability ratings. Stronger consensus-seeking tendencies and a more active memory environment in the pairs may have contributed to these results.

KEYWORDS

Question answerability; judgments; pairs; individuals; polarization effect

Misjudgments of the extent to which questions have been answered have contributed to some of mankind's worst catastrophes and can contribute to unfortunate events and developments in everyday life. An example is the Thalidomide catastrophe in the 1950s, where questions regarding side effects of a pregnancy-nausea medication were prematurely regarded as answered, resulting in severely handicapped babies. In everyday contexts, poor judgments of the extent to which questions have been answered may occur for example concerning the effects of new building materials and chemicals and in social contexts such as health assessments, welfare, and civic safety.

By a question being *answerable*, we mean that a correct, well-argued, answer at a relevant granularity level can be provided to the question, by at least one now living person. In this research, the *answerability* of questions is seen as a matter of degree. That is, answerability judgments are seen as judgments that may involve considerations about the amount of work or time needed by me, or others, to answer the question, if possible at all. For example, how answerable today is the question: Can waste from nuclear power plants be safely long-time stored? Misjudgments about answerability may occur because relevant factors have not been attended to, or wrong conclusions have been drawn about a relevant factor and because ignorance has not been considered. In spite of the importance of judgments of questions' answerability, very little research has investigated how people evaluate the answerability of questions relating to facts in the world. Therefore, the present research investigated this issue.

In general, cognition is distributed in a society and globally. Moreover, no one person

can do everything and mankind's understanding develops by collaboration in research and other contexts (Atran, Medin, & Ross, 2005; Perkins, 1993). Therefore, it often makes good cognitive economic sense to rely on, or to consider, what other people and groups, including researchers, have concluded. Furthermore, plenty of research, reviewed by Surowiecki (2004), has shown that taking other people's knowledge into consideration often leads to better results. People's tendency to rely on socially prevalent understanding is evidenced on the individual level by phenomena such as social learning, imitation, conformation seeking, submission to authorities, etc. Thus, and relevant for this study, people's responses often, implicitly or explicitly, to some extent rely on socially available understanding. For example, when considering answerability, socially prevalent knowledge may refer to the specific answer to the question judged (answer consensus) or to conceptions for or against question answerability (answerability consensus), for example that certain types of questions are easier to answer than other types. Socially prevalent knowledge is henceforth called *consensus knowledge,* thus total consensus is not assumed.

In everyday life, people either judge the answerability of questions alone or with others. For example, people may deliberate together about how much time and resources should be given to an investigation or a research project, or make such decisions alone. Both individuals and groups have drawbacks with respect to how they process information (Davis, 1992; Lu, Yuan, & McLeod, 2012; Paulus & Yang, 2000). Given the potential importance of answerability judgments, the present study investigated differences between individuals' and groups' judgment of questions' answerability.

Research on answerability

In general, a reasonable way for people to make answerability judgments is to get an initial understanding of the question and then to consider if they know the answer (if so, the question may be judged to be answerable), and if not, if other persons or categories of people (such as researchers) know the answer or not. Finally, they may consider if the question is answerable at all. Kruger (1999) and Nickerson (1999) reported evidence for models where people use their own knowledge as a starting point in order to infer what others know. Similarly, people, when forming answerability judgments, may first consider if they know the answer themselves.

Previous research has mostly studied one or two of these aspects in the same study (own knowledge, others' knowledge, and if the question is answerable at all) and has used questions for which people tend to agree on the answer. For example, "don't know" judgments have been studied in different contexts, e.g., for learnt sentences and facts (e.g., Glucksberg & McCloskey, 1981; Kolers & Palef, 1976), reasoning in a developmental context, (e.g., Scholnick & Wing, 1988), and in research on confidence judgment accuracy (Griffin & Brenner, 2004). Another context is eyewitness research in psychology on suggestibility (e.g., Buratti, MacLeod, & Allwood, 2014; Candel, Memon, & Al-Harazi, 2007; Frey & Scoboria, 2012; Roebers & Fernandez, 2002; Roebers, Linden, & Howie, 2007; Scoboria, Mazzoni, & Kirsch, 2008).

Judgments of others' knowledge (including researchers' knowledge), have been studied in open contexts in sociology (e.g., Gross, 2007a), risk perception research (e.g., Sjoberg, 2001), science communication studies, educational psychology (e.g., Hofer & Pintrich, 2002; Scharrer, Stadtler, Bromme, & Bromme, 2014; Shtulman, 2013), and in other parts of psychology (e.g., Johansson & Allwood, 2007).

Finally, philosophers, risk decision researchers and others have considered how a general lack of knowledge (ignorance) can be conceptualized. Ignorance is relevant for answerability judgments since the extent of the person's own, and others' ignorance may affect how persons perceive answerability. Various taxonomies of ignorance (in a broad sense) have been presented (e.g., Armour, 2000; Croissant, 2014; Faber, Manstetten, & Proops, 1992; Gross, 2007b; Rescher, 2009; Smithson, 1993). The present research contributes by studying answerability judgments that may include evaluations of both one's own and others' knowledge and of whether the question is answerable at all for questions where the answer may not be generally agreed upon or is currently unknown.

Processing differences between individuals and pairs

Individuals' and pairs' information processing differ in two important respects. Firstly, pairs on average are likely to have more knowledge than the average individual. Secondly, pairs are likely to have a more lively memory activation situation than individuals since the pair members, in addition to their own associations, also provide memory cues for one another. Thus, pairs may, especially in uncertain contexts, consider more arguments than individuals.

However, groups seem to be more prone to discuss their common knowledge than information unique to each member (Lu et al., 2012). Therefore, questions with a high expected consensus about the answer may be judged as more answerable by pairs than by individuals. This, since groups are more prone to focus on, and discuss, common information and thus information with generally high expected consensus.

More in general, research on judgments and decisions by groups have reported mixed effects of dyadic (or group) information processing. Some studies indicate a positive effect, e.g., research investigating the wisdom of the crowd effect (e.g., Surowiecki, 2004). Similarly, given suitable conditions, pairs can show more accurate confidence judgments than individuals (e.g., Allwood & Granhag, 1996; Allwood, Granhag, & Johansson, 2003). Other studies, for example involving social loafing effects or inappropriate memory cueing between members, have indicated possible negative effects of group collaboration (Andersson, Helstrup, & Rönnberg, 2007; Dewett, 2003; Geen & Shea, 1997; Karau & Williams, 1993; Paulus & Yang, 2000; Rajaram & Pereira-Pasarin, 2010).

Furthermore, pairs may show a polarization effect (the effect that groups tend to choose more extreme alternatives, e.g., Kaplan & Miller, 1983; Kugler, Kausel, & Kocher, 2012). Polarization effects can be important. In contrast to the wisdom of the crowd effect (Surowiecki, 2004) and bootstrapping (Herzog & Hertwig, 2009) that tend to build on averaging group members opinions, polarization effects may serve to distance groups from accurate judgments. Two main competitive explanations are recognized to explain polarization effects, informational influence, and social comparison (Isenberg, 1986). Informational influence explanations, in the form of persuasive argument theory, propose that groups will move in the direction of the arguments they find most persuasive. A central thought in social comparison explanations is that people compare themselves to others and adjust

their self-presentation according to how they perceive the other, often with the (maybe implicit) aim to appear "in a socially favorable light" (Isenberg, 1986, p. 1142). The result may be a shift towards greater perceived social value.

Both explanations may be needed to understand polarization, but to different extent for different types of task (Isenberg, 1986). In "comparison-poor" tasks (involving judgments about matters of fact) arguments may be more important and in "argument-poor" tasks (such as face comparison tasks) social comparison may be more important. In the present research, pair members were fairly close friends and the answerability task involved judgments about facts. Therefore, arguments may have played a central role for the outcome and social comparisons (as described in social comparison explanations) somewhat less so.

However, we suggest that another form of social comparison may influence answerability judgments as well. Arguments may be perceived more valid when evaluated against a socially derived comparison norm of what is prevalent opinions (common knowledge), for example in different more or less well-respected groups. Specifically, we suggest pairs may be more influenced by such norms than individuals, even if individuals are affected too (e.g., Koriat, 2008, 2012; Koriat & Adiv, 2012). This is elaborated next.

Groups, including pairs, may be affected by assumed consensus knowledge in two ways: *present consensus* and *referred consensus*. Present consensus is the effect of (assumed) consensus between group members. For example, when pair members agree they may take this as a sign that they are correct in their judgment. Referred consensus may influence both individuals and pairs, and relates to beliefs about common knowledge or consensus among not present others (e.g., people in general, or experts such as researchers).

Koriat (2008) researched the referred consensus effect for individuals. Individuals showed high confidence both when they in a two-alternative general knowledge task selected the commonly believed answer alternative that was also the correct answer, and when they selected the commonly believed, but *incorrect*, answer. When present and referred consensus point in the same direction, the effect of consensus may be amplified for pairs.

The present study

This study compared answerability judgments of individuals and collaborating pairs. We used general knowledge questions with different levels of assumed answerability (henceforth called consensus and non-consensus questions). *Hypothesis 1* expected a polarization effect for questions with high answerability ratings (including our consensus questions). Thus, we expected pairs to rate these questions more answerable than individuals. One reason is that pairs may show a stronger *present consensus* effect.

Hypothesis 2 expected a different polarization effect for questions rated low in answerability: here, pairs were expected to rate answerability lower than individuals. One reason is that the pair members may, together, identify more difficulties to answer the non-consensus questions than the individuals. Furthermore, in the perspective of Koriat's consensuality principle (2008), we speculated that less, and more fragmented, social consensus knowledge cues might be available for the non-consensus questions than for the consensus questions. Moreover, for the non-consensus questions, consensus knowledge about answerability might to greater extent include reasons both for and against the question being answerable, than for the consensus questions. Accordingly, both answer, and answerability, consensus knowledge might be less influential for the answerability judgments of non-consensus questions,

or work against high answerability ratings. Finally, given these observations and the pairs' more active memory environment, pairs might, more than singles, become aware of that they lack knowledge about the issues considered for the non-consensus questions which may contribute to lower ratings in the pairs.

In order to get more information about the answerability judgments, the individuals were asked to verbalize their thoughts aloud while arriving at the answerability judgments. The results from the content analysis of these verbal data will be presented in a forthcoming paper.

Method

Participants and design

Ninety-one persons from the participant pool at a university psychology department in Sweden participated (38 men and 53 women, mean age 27 years; range 18–62 years). Each person received approximately 15 USD for participating.

Sixty-one participants were first randomized to two conditions – the *Individual condition* and the *Pair condition*. The 30 participants in the pair condition then recruited a friend as the second pair member. Thereby we created pairs where the members knew each other and had a fairly equal status without any obvious dominance of one pair member. The pairs were created according to the criterion that the maximum age difference should be 7 years and the maximum age 65 years.

Fifteen men and 16 women participated in the individual condition (mean age 28 years). There were 23 men and 37 women (14 same-sexed dyads) in the pair condition (mean age 26 years). Half of the individuals and half of the pairs received the answerability questions in order 1–20 and the other half in order 11–20, 1–10.

Materials

Answerability questionnaire

A questionnaire with 20 answerability judgment tasks was prepared (see Appendix 1). Each task consisted of a stated question for which the participants were to judge how *answerable* it was (defined below). The questions were general knowledge questions, but the answers to many of the questions were controversial or unknown. They concerned for example, genetically modified food, the next prime minister of Sweden, and the location of NASDAQ's head office. We attempted to select questions so that they, on average, would be judged as having from high to low answerability. Four of the 20 questions were selected to have a high level of expected consensus about their answerability (*consensus* questions). An example is: "Where is the main office of NASDAQ situated?" We expected the remaining 16 questions to have a lower expected answerability consensus (*non-consensus* questions). An example is: "How many galaxies are there in the Universe?"

In a pilot study, 100 independent participants from a student pool at the University of Gothenburg rated the proportion of Swedes that would consider each of 16 of the 20 questions to be answerable (on a scale from 0 to 100%) The tested questions included the four consensus questions and 12 of the non-consensus questions used in the present study. On average participants judged that 64% of the Swedes would consider the consensus questions

answerable, but only 54% of the non-consensus questions. This supports our separation of the 20 questions into two sets of questions. However, the division into consensus and non-consensus questions is best considered a matter of degree. We do not argue that there is a qualitative difference.

The answerability of each of the 20 questions was judged on a scale indicating how probable the participants judged that the question could be answered by a now living human being. The scale ranged from 0% (*Cannot be answered*) to 100% (*Can be answered*) in increments of 10%.

Procedure

Participants were first told that the study involved judging the answerability of a number of questions and agreed to their reports being tape recorded. The participants in the individual condition were notified that they would be asked to think aloud before making numerical answerability judgments. The pairs were instructed to discuss the answerability of each question with each other and that it was important that both members should take part in the discussion, before they gave a joint, agreed upon, answerability judgment of the question.

Next, all participants read a definition of the concept "answerability" so that they would use this concept similarly when giving the answerability judgments. The instructions stressed that the judgments concerned the participants' rated probability that the questions could be answered by *a now living human being*, not necessarily themselves. Moreover, the instructions emphasized that this task was *not* the same as rating how many percent of Sweden's or the world's population could answer the question. That is, if *one* person could answer the question it was answerable. Furthermore, they did not need to know the answer in order for the question to be answerable. Next, two examples were given, first of a question that might be judged as answerable although they did not know the answer themselves ("What is the Icelandic word for '(to) sing'?") and then of a question that might seem less evident that it could be answered by a now living human being ("How many centimeters is the earth's largest leaf on a now living tree?"). The instructions also noted that people can think differently about whether questions are answerable. Next the instructions explained that by "can be answered" was meant that the question can be answered in a correct way, that one can provide good arguments for the answer and that the answer should be interesting and not too vague. As an example was mentioned that to answer "less than a meter" to the question "How tall was the world's ever shortest creature?" was probably correct but was such a vague and inexact answer that it was uninteresting. Next, participants in the individual condition were given think-aloud instructions (rendered in Appendix 2). When, on a few occasions, the participants gave answers that showed that they had misunderstood the task, the experimenter reminded them of the definition of the task.

When starting to answer the 20 questions, the individuals were instructed to "Think aloud and then mark a judgment by circling that number on the scale that corresponds to your judgment". Pairs were reminded that both members should be active in the discussion for every question and that they should try to agree on the answer before marking their judgment on the scale. Finally, participants reported their gender, age, and education level.

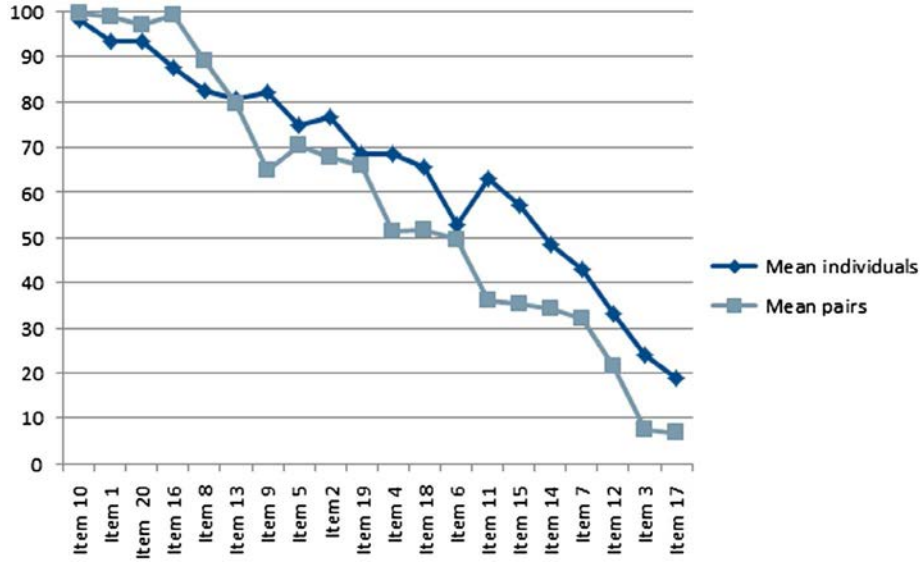

Figure 1. Means for the individuals' and the pairs' answerability ratings for the 20 questions.

Table 1. Means (SDs) for the answerability judgments of the consensus questions and of the non-consensus questions for the individual and pair conditions. number of words for the corresponding questions in the individual and pair conditions.

	Answerability judgments		number of words	
Condition	Consensus	non-consensus	Consensus	non-consensus
individuals	93.15 (10.31)	58.78 (15.20)	47.77 (29.21)	78.91 (40.94)
Pairs	98.75 (2.77)	47.76 (14.09)	92.51 (62.12)	209.21 (157.94)

notes: consensus = consensus questions, non-consensus = non-consensus questions.

Results

Figure 1 shows the answerability judgments for each of the 20 questions for individuals and pairs. As seen, the answerability ratings decrease gradually from the highest to the lowest rated question for both pairs and individuals. The ratings for the consensus questions (questions 1, 10, 16, and 20) were higher than for the non-consensus questions, for both pair and individuals. Furthermore, for the questions with the highest answerability ratings (including the consensus questions), the level of answerability tended to be *higher* for the pairs, compared to the individuals. For the questions with lower answerability ratings, the ratings tended to be *lower* for the pairs, compared to the individuals. Table 1 shows the means and standard deviations of the answerability judgments for consensus and non-consensus questions and the two conditions.

A three-way mixed ANOVA with the repeated measure question type (consensus or non-consensus) and the independent variables individual/pair condition and question order was performed. No significant effect of question order on the answerability judgments was found. Therefore, the two question orders were collapsed in the following analyses. Next, a two-way mixed ANOVA was performed with the repeated measure question type (consensus or non-consensus) and the independent variable individual/pair

condition. The consensus questions were judged more answerable than the non-consensus, $F(1, 59) = 416.12$, MSE = 133.44, $p < .001$. However, there was an interaction between question type and the individual/pair condition $F(1, 59) = 15.79$, MSE = 133.44, $p < .001$.

As the two-way mixed ANOVA showed, consensus questions were rated as more answerable than non-consensus questions. This was also true for individuals $F(1, 59) = 137.15$, MSE = 133.44, $p < .001$ and pairs $F(1, 59) = 292.22$, MSE = 133.44, $p < .001$ separately (simple effects). As indicated by the *F*-values, the pairs showed a bigger difference between the answerability of the two question-categories than the individuals. A comparison of simple effects between the conditions confirmed that the pairs gave *higher* answerability judgments for the consensus questions than the individuals, $F(1, 59) = 8.29$, MSE = 57.78, $p = .006$. In contrast, the pairs gave *lower* answerability judgments for the non-consensus questions $F(1, 59) = 8.61$, MSE = 215.11, $p = .005$.

Furthermore, in order to investigate the effect of answerability level as such, the questions were divided into two categories based on the answerability ratings made in the individual condition which, in this context, can be seen as a control group. The 10 items with the highest answerability ratings constituted one category and the other 10 items another category.

Since a three-way mixed ANOVA showed no effect of question order, a two-way mixed ANOVA was performed with the independent variables individual/pair condition and question Category (top 10 vs bottom 10) and the dependent variable answerability. The analysis confirmed that the top 10 questions were rated more answerable than the bottom 10, $F(1, 59) = 469.38$, $p < .001$. Also, there was an effect of condition so that individuals gave significantly higher answerability ratings than pairs, $F(1, 59) = 6.41$, MSE = 281.88, $p = .014$. In addition, there was an interaction effect, $F(1, 59) = 8.79$, MSE = 114.48, $p = .004$. Simple effects analyses showed that pairs rated the answerability for the bottom 10 questions lower than individuals, $F(1, 59) = 9.20$, MSE = 299.24, $p = .004$. However, no such difference was found for the top 10 questions, $F(1, 59) = .60$, MSE = 97.12, $p = .443$.

An analysis of the number of words for the individuals and the pairs per question category was carried out in order to get an indication of the activity in the two conditions, and thereby also about amount of arguments and considerations (see Table 1). A two-way mixed ANOVA with the repeated measure question type (consensus or non-consensus) and the independent variable individual/pair condition showed that the pairs used significantly more words than the individuals, $F(1, 59) = 20.78$, MSE = 11238.38, $p < .001$. Furthermore, more words were used for non-consensus questions compared to consensus questions, $F(1, 59) = 39.62$, MSE = 4205.27, $p < .001$. There was also an interaction effect between condition and question type. Simple effects analyses showed that the individuals used less words than pairs both for consensus, $F(1, 59) = 13.09$, MSE = 2330.46, $p = .001$ and non-consensus questions, $F(1, 59) = 19.17$, MSE = 13113.19, $p < .001$. There was no difference between consensus and non-consensus questions with respect to number of words used for individuals, $F(1, 59) = 3.57$, MSE = 4205.27 $p = .064$, but this was the case for the pairs, $F(1, 59) = 48.58$, MSE = 4205.27, $p < .001$. The same ANOVA computed for the top 10 and bottom 10 answerability-rated questions showed similar results, except that no interaction effect was found. Finally, pairs took more time on the task, compared to individuals, $t(35.1) = -4.15$, $p < .001$, $d = 1.40$.

Discussion

Individuals and pairs were compared with respect to their answerability ratings of questions with a broad range of expected level of judged answerability. Both individuals and pairs rated consensus questions (i.e., questions with a high level of expected consensus regarding their answerability) higher in answerability than the non-consensus questions. This result replicates results in a study where participants rated questions' answerability on a scale ranging from 0% meaning "cannot be answered" to 100% meaning "can be answered" (Karlsson, Allwood, & Buratti, 2016). This result is in line with a general *referred consensus* effect, that is, both individuals and pairs may have been more affected by consensus considerations when judging the consensus questions compared to the non-consensus questions. Furthermore, judgments of the non-consensus questions may have been associated with more possibilities to detect complexities, uncertainties and difficulties with respect to providing an answer to them.

No time restrictions were set for the participants. Pairs used more words than individuals, for both types of questions, indicating higher activity in the pairs. However, per se, the absolute number of words may not have a simple relation to the level of the answerability judgments since pairs showed *higher* answerability judgments for the consensus questions and *lower* for the non-consensus questions, than individuals.

Our main focus was answerability judgments when consulting others. Pairs, compared with individuals, evidenced two types of polarization effects. In line with our first hypothesis, in the first polarization effect, pairs gave higher answerability ratings for consensus questions than individuals. This polarization effect held for answerability ratings above 80% but was not significant when the analysis was made by separating the 10 questions with the highest answerability ratings from the remaining 10 questions. The effect may be at least partly explained by stronger consensus effects in the pairs. A present consensus effect in the pairs may be amplified by the tendency for groups to share common information compared to unique information (e.g., Lu et al., 2012). Moreover, considerations concerning the answerability of questions with higher consensus with respect to their answerability may be less controversial, and the pairs' present consensus effects may have amplified this effect, compared to the individuals' judgments.

In line with our second hypothesis, the second polarization effect showed that pairs gave lower answerability ratings for non-consensus questions (irrespective of if 16 or 10 questions were included in the analysis) than individuals. Higher uncertainty may, in general, be expected to be associated with the possible answers to the non-consensus questions (including the 10 questions with the lowest answerability ratings), compared with the consensus questions. The more active processing environment and the expected greater effect of *present consensus* in the pairs may have facilitated detecting complexities and uncertainties about the possibility to answer the non-consensus questions and this may have contributed to their lower answerability ratings, as may the presence of disagreements within pairs[1]. The result that pairs used substantially more words when judging the non-consensus questions than individuals is in line with this suggestion. Further research should investigate these effects.

Our results show similarities to those of Koriat (2015) testing the consensual amplification hypothesis, although Koriat (2015) investigated accuracy and confidence and we did not. Koriat (2015) found that the tendencies showed by individuals were amplified in the pairs and explained the effect of group deliberation as follows "[it] amplifies the contribution

of the shared clues, whether these clues favor the correct decision or the wrong decision." (p. 947). He noted that the observed pair amplification is consistent with polarization and that other factors, apart from consensuality effects, also contribute to explain the results. Furthermore, Koriat (2008) found that the level of confidence judgments is more clearly related to consensuality than to accuracy. In addition, research reviewed by Koriat (2015), shows that pair members with the highest confidence tend to influence pairs' answer more. These observations support that socially prevalent knowledge may help explain our polarization results. However, we suggest that the effects of socially prevalent knowledge may be especially clear for the polarization effect pertaining to the consensus questions used in this study since here both answer and answerability consensus are likely to clearly point in the same direction and the effect of these cues are likely to be amplified in the pairs. The consensus questions for the pairs reached almost 100% answerability, indicating nearly absolute certainty. For the non-consensus questions, the cues from social knowledge, as discussed above, may be more fragmented and less consistent.

Like most studies, the present research has limitations. For example, only two question types were investigated. In the future, the answerability of other question types such as yes/no questions compared to open questions, questions about present factual states of the world compared to future states, and questions concerning beliefs about the existence of unknown risks, should be studied. Moreover, the number of questions, especially the consensus questions, was limited. The reason was that we speculated that the effect of even a limited number of consensus questions would be fairly clear. However, the sample of consensus questions was small and lack in representativeness. For example, we did not specifically aim to include questions for which there may be a consensus opinion that they are unanswerable. Although we, in line with our reasoning above, speculate that socially prevalent beliefs about high *unanswerability* may tend to be more diffuse for the types of questions included in the present research, it may still be that, as suggested by one of the reviewers, socially prevalent beliefs about unanswerability could have contributed to the answerability assessments for some of the questions with the lowest answerability ratings. In brief, the issue of questions with high consensus about their unanswerability is of interest and should be investigated in future research. It is still noteworthy that all the included consensus questions were rated as the most answerable of the 20 questions and especially that a polarization effect was found for questions with high answerability ratings. However, future research should attempt to select consensus and non-consensus in a more systematic way than in this study. Also, only one group size was investigated, namely dyads, which is a special type of group. Future studies should investigate a broader range of group sizes.

Moreover, the effects of differences between group members should be explored, for example, status differences. Allwood and Granhag (1996) found that when one member of the pair dominated the interaction the level of accuracy in the pairs' joint confidence judgments was especially low. Furthermore, in our study, present and referred consensus tend to be consistent with each other. Future research may investigate what happens when they contradict one another, e.g., when two people with the same minority opinion meet. This study used no time constraints for the participants' judgments, but the effect of time for the deliberations should be explored. When deliberation time is limited judgments may be more in line with social consensus, especially for pairs. Finally, we asked the individuals to think aloud. Some researchers have suggested that thinking aloud may influence thought processes. However, Ericsson and Simon (1980) and Fox, Ericsson and Best (2011) argued

convincingly that think aloud is not the same as introspection (as claimed by e.g., Schooler, 2011), but simply the reading out of verbal contents from STM and such reading out in most cases does not affect the participants' thought processes.

Conclusion

The present research shows that answerability judgments are affected by whether they are made by individuals or by pairs but that this influence may differ depending on the level of the answerability judgment and the type of questions judged. Two types of polarization effects were demonstrated, one at high answerability levels (mostly our consensus questions), where pairs had higher answerability ratings than individuals, and one at low levels where pairs had lower answerability ratings than individuals. Our results can, at least partly, be interpreted in terms of social consensus seeking effects which may have been more amplified in the pairs. The more active processing environment in the pairs may have helped this effect. Future studies should investigate answerability judgments in contexts more representative of real social life, with possible effects of if the question is answered prematurely, for example on the basis of incomplete information or deficient conclusions from the evidence at hand.

Note

1. We thank one of the reviewers for the suggesting this effect of disagreements.

Funding

This work was supported by the Swedish Research Council (Vetenskapsrådet, grant number 2011-1784) by a grant given to the first author.

References

Allwood, C. M., & Granhag, P. A. (1996). Realism in confidence judgements as a function of working in dyads or alone. *Organizational Behavior and Human Decision Processes, 66*, 277–289.

Allwood, C. M., Granhag, P. A., & Johansson, M. (2003). Increased realism in eyewitness confidence judgments: The effect of dyadic collaboration. *Applied Cognitive Psychology, 17*, 545–561.

Andersson, J., Helstrup, T., & Rönnberg, J. (2007). Collaborative memory: How is our ability to remember affected by interaction with others? In S. Magnussen & T. Helstrup (Eds.), *Everyday memory* (pp. 135–156). New York, NY: Psychology Press.

Armour, P. G. (2000). The five orders of ignorance. *Communications of the ACM, 43*, 17–20.

Atran, S., Medin, D. L., & Ross, N. O. (2005). The cultural mind: Environmental decision making and cultural modelling within and across populations. *Psychological Review, 112*, 744–776.

Buratti, S., MacLeod, S., & Allwood, C. M. (2014). The effects of question format and co-witness peer discussion on the confidence accuracy of children's testimonies. *Social Influence, 9*, 189–205. doi:10.1080/15534510.2013.804434

Candel, I., Memon, A., & Al-Harazi, F. (2007). Peer discussion affects children's memory reports. *Applied Cognitive Psychology, 21*, 1191–1199. doi:10.1002/acp.1355

Croissant, J. L. (2014). Agnotology: Ignorance and absence or towards a sociology of things that are not there. *Social Epistemology, 28*, 4–25.

Davis, J. (1992). Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950–1990. *Organizational Behavior and Human Decision Processes, 52*, 3–38.

Dewett, T. (2003). Understanding the relationship between information technology and creativity in organizations. *Creativity Research Journal, 15*, 167–182.
Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.
Faber, M., Manstetten, R., & Proops, J. L. R. (1992). Humankind and the environment: An anatomy of surprise and ignorance. *Environmental Values, 1*, 217–241.
Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*, 316–344.
Frey, M. P., & Scoboria, A. (2012). Self-esteem predicts quality of responding to interview questions when skill in responding is low. *Personality and Individual Differences, 52*, 363–368.
Geen, R., & Shea, J. D. C. (1997). Social motivation and culture. In D. Munro, J. F. Schumaker, & S. C. Carr (Eds.), *Motivation and culture* (pp. 33–48). New York, NY: Routledge.
Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 311–325.
Griffin, D., & Brenner, L. (2004). Perspectives on probability judgment calibration. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 177–199). Malden, MA: Blackwell Publishing.
Gross, M. (2007a). Communicating ignorance and the development of post-mining landscapes. *Science Communication, 29*, 264–270.
Gross, M. (2007b). The Unknown in process: Dynamic connections of ignorance, non-knowledge and related concepts. *Current Sociology, 55*, 742–759.
Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*, 231–237. doi:10.1111/j.1467-9280.2009.02271.x
Hofer, B. K., & Pintrich, P. R. (Eds.). (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Lawrence Erlbaum.
Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology, 50*, 1141–1151.
Johansson, M., & Allwood, C. M. (2007). Own-other differences in the realism in some metacognitive judgments. *Scandinavian Journal of Psychology, 48*, 13–21.
Kaplan, M. F., & Miller, C. E. (1983). Group discussion and judgment. In P. B. Paulus (Ed.), *Basic group processes* (pp. 65–94). New York, NY: Springer-Verlag.
Karau, S., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681–706.
Karlsson, B. S. A., Allwood, C. M., & Buratti, S. (2016). Does anyone know the answer to that question? Individual differences in judging answerability. *Frontiers in Psychology, 6* (2060), http://journal.frontiersin.org/article/10.3389/fpsyg.2015.02060/full
Kofsky Scholnick, E., & Wing, C. S. (1988). Knowing when you don't know: Developmental and situational consideration. *Developmental Psychology, 24*, 190–196.
Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory & Cognition, 4*, 553–558.
Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 945–959.
Koriat, A. (2012). The self-consistence model of subjective confidence. *Psychological Review, 119*, 80–113. doi:10.1037/a0022171
Koriat, A. (2015). When two heads are better than one and when they can be worse: The amplification hypothesis. *Journal of Experimental Psychology: General, 144*, 934–950.
Koriat, A., & Adiv, S. (2012). Confidence in one's social beliefs: Implications for belief justification. *Consciousness and Cognition, 21*, 1599–1616.
Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221–232.
Kugler, T., Kausel, E. E., & Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups, *Wiley Interdisciplinary Reviews: Cognitive Science, 3*, 471–482.

Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review, 16*, 54–75. doi: 10.1177/1088868311417243
Nickerson, R. S. (1999). How we know – and sometimes misjudge – what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737–759.
Paulus, P. B., & Yang, H.-C. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes, 82*, 76–87.
Perkins, D. N. (1993). Person-plus: A distributed view of thinking and learning. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 88–110). Cambridge: Cambridge University Press.
Rajaram, S., & Pereira-Pasari, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science, 5*, 649–663. doi:10.1177/1745691610388763
Rescher, N. (2009). *Ignorance (on the wider implications of deficient knowledge)*. Pittsburgh, PA: University of Pittsburgh Press.
Roebers, C. M., & Fernandez, O. (2002). The effects of accuracy motivation and children's and adults' event recall, suggestibility, and their answers to unanswerable questions. *Journal of Cognition and Development, 3*, 415–443.
Roebers, C. M., von der Linden, N., & Howie, P. (2007). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology, 25*, 109–134.
Scharrer, L., Britt, M. A., Stadtler, M., & Bromme, R. (2014). You'd better ask an expert. *Applied Cognitive Psychology, 28*, 465–471.
Schooler, J. W. (2011). Introspecting in the spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin, 137*, 345–350.
Scoboria, A., Mazzoni, G., & Kirsch, I. (2008). "Don't know" responding to answerable and unanswerable questions during misleading and hypnotic interviews. *Journal of Experimental Psychology: Applied, 14*, 255–265. doi:10.1037/1076-898X.14.3.255
Shtulman, A. (2013). Epistemic similarities between students' scientific and supernatural beliefs. *Journal of Educational Psychology, 105*, 199–212. doi:10.1037/a0030282
Sjöberg, L. (2001). Limits of knowledge and the limited importance of trust. *Risk Analysis, 21*, 189–198. doi:10.1111/0272-4332.211101
Smithson, M. (1993). Ignorance and science: Dilemmas, perspectives, and prospects. *Science Communication, 15*, 133–156.
Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Anchor Books.

Appendix 1.

Questions used for answerability judgments. The 4 consensus questions are marked with a *.

(1) What flower is called Anemone Nemorosa in Latin?*
(2) How much environmentally hazardous material does the most common laptop contain?
(3) How many galaxies are there in the Universe?
(4) Are congestion charges good for the environment?
(5) Are humans causing global warming?
(6) Is there today a safe way to store nuclear waste?
(7) Did the Maya Indians have any knowledge about the end of the world?
(8) Does vaccination provide a safe protection against the disease TBE, transmitted via tics?
(9) Can you become allergic to electricity?
(10) Where is the head office for NASDAQ situated?*
(11) Does the human body have an unknown system of circulation?
(12) Who will be the next prime minister of Sweden?
(13) Can genetically modified fruits and vegetables be dangerous to nature's ecological systems?

(14) Which method is the best to heat fish?
(15) Will the polar ices melt in 500 years?
(16) How many petals does a Blue Anemone usually have?*
(17) How many varieties of the word snow did humans have during the ice age?
(18) How many different kinds of bacteria can be transmitted to humans from tics?
(19) Is radiation from cell phones dangerous?
(20) Who has written the books “Kris” and “Kallocain”?*

Appendix 2.

Think aloud instructions given to the participants in the individual condition.

> “As soon as we start with the task I would like you to start to think aloud. It is best if you can be as spontaneous as possible. Tell all you are thinking about, even if some details might appear irrelevant or embarrassing. You should not explain why you think as you do. Each time you start with a new question I want you to start with reading the number of the question and the actual question aloud, so that we later can know which question is treated. Then you start to think aloud immediately as you start to solve the task.”

Next, participants carried out an exercise task in thinking-aloud about an answerability question that was to judge the answerability of the question “How long is the world’s longest car?” on the same scale used for each of the following 20 questions from 0 to 100%. The participants were reminded to think aloud if they had been silent for more than 5 s by the use of phrases such “What are you thinking of now?”, or “Continue to talk” stated in a neutral way.