Structured Groups Make More Accurate Veracity Judgments Than Individuals

Iain Hamlin[1]*, Fergus Bolger[1], Alexandrina Vasilichi[1], Ian Belton[1], Megan M Crawford[1], Aileen Sissons[1], Courtney Taylor Browne Lūka[1], and George Wright[1]

[1] Management Science, University of Strathclyde, Glasgow, UK

*Corresponding author address: Department of Psychology, University of Derby, Kedleston Road, Derby, DE22 1GB

Email address: i.hamlin@derby.ac.uk

The authors declare that there is no conflict of interest in respect of this research.

The data that support the findings of this study are openly available in OSF at http://doi.org/10.17605/OSF.IO/NA3CY

Acknowledgement: This research is based on work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), under Contract [2017-16122000003]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

Structured Groups Make More Accurate Veracity Judgments Than Individuals

Abstract

Groups often make better judgments than individuals, and recent research suggests that this phenomenon extends to the deception detection domain. The present research investigated whether the influence of groups enhances the accuracy of judgments, and whether group size influences deception detection accuracy. 250 participants evaluated written statements with a pre-established detection accuracy rate of 60% in terms of veracity before viewing either the judgments and rationales of several other group members or a short summary of the written statement and revising or restating their own judgments accordingly. Participants' second responses were significantly more accurate than their first, suggesting a small positive effect of structured groups on deception detection accuracy. Group size did not have a significant effect on detection accuracy. The present work extends our understanding of the utility of group deception detection, suggesting that asynchronous, structured groups outperform individuals at detecting deception.

Keywords: deception detection; group decision making; written cues; virtual groups; group size

The accuracy of participants in deception detection experiments is typically only a few percentage points above chance (Bond & DePaulo, 2006). However, it is likely that this poor level of performance can be improved upon. Statistical models based on objective measurements of cues to deception outperform humans at predicting veracity (Hartwig & Bond, 2014; Vrij et al., 2000; Vrij, Akehurst, et al., 2004), suggesting that it is possible for humans to make significant gains in deception detection performance. Furthermore, a recent meta-analysis demonstrated that the accuracy of veracity judgments can be enhanced through training (Hauch et al., 2014). Further, various other techniques are currently being investigated as a means to improving deception detection, including harnessing unconscious deception judgments (Moi & Shanks, 2015; ten Brinke et al., 2014) and using interview techniques to elicit stronger cues to deception (Vrij & Granhag, 2012).

Recently, researchers have investigated whether harnessing the decision-making power of groups is an additional means of improving deception detection accuracy. Group judgments typically outperform those of individuals on estimation tasks (e.g., Bonner et al., 2007; Laughlin et al., 2003), almanac questions (e.g., Henry, 1995) and forecasting questions (e.g., Flores & White, 1989). This group advantage has also been found in the deception detection domain. In a series of experiments examining individual and group deception detection across both high-stakes and low-stakes lies, Klein and Epley (2015) found that groups were consistently better than individuals at detecting lies. However, groups and individuals did not differ in their ability to detect truths. The same pattern of results was found by Frank et al. (2004). In both of these studies, the superior performance of groups was not simply a result of a reduced truth bias (i.e. a reduced tendency to evaluate messages as being truthful) relative to individuals.

In a study investigating the relative performance of individuals, ad-hoc groups and established groups, McHaney et al. (2015) found that, across both truths and lies, established groups had significantly higher levels of detection accuracy than individuals. However, two studies failed to find this group effect. Culhane et al. (2013) found no evidence that dyads outperform individuals in terms of detection accuracy. Similarly, Park et al. (2000) found that small groups of three to six were not significantly better at detecting deception than individuals. However, this experiment employed only two deception targets, raising the possibility that stimulus-specific factors account for the findings.

There are several possible explanations of a group advantage in deception detection. Groups are composed of members with different demographics and expertise, each of whom can potentially provide complimentary information, the additive effect of which is an information resource significantly greater than that available to an independent individual (Bonner et al., 2002). In the context of deception detection, by pooling information on cues to deception that is dispersed amongst individual group members, groups could outperform individuals on deception detection tasks. Such unshared information could include insights into lying styles specific group members share with the deception detection targets, or insights into the validity of the semantic content of a target’s messages.

An alternative explanation of the success of groups draws on the majority effect. The majority effect is well documented in the empirical literature, with several decades of research reporting the persuasive effect of majority opinions on individual judgments in many domains, such as preferences (White & Dahl, 2006), attitudes (Wood, 2000) and opinions (Glynn, 1997). If baseline deception detection performance is above chance, as it is in most deception detection studies then, more often than not, the majority judgment of a deception detection group will be correct. Individual members of such groups, if experiencing the pull of the majority effect, would be expected to often adopt the correct answer when exposed to the judgments of such a group. By this process groups will serve to amplify weak signals of deception relative to individuals.

Extending previous research on group deception detection, the present research investigates the deception detection accuracy of asynchronous, structured groups relative to that of individuals. In particular, we investigate groups structured according to the judge-advisor system. A judge-advisor system is a type of advice structure in which an individual (judge) receives and evaluates advice on a particular issue before making their final decision (Sniezek & Buckley, 1995). Structured-group techniques such as this bring several pragmatic benefits over unstructured groups. For example, such groups can be asynchronous in nature, with members contributing at different times and from different locations, therefore reducing the cost and pragmatic challenges of organising physical meetings, and potentially increasing the number of people who can inform decision making. Moreover, structured processes such as the judge-advisor system also help minimise many of the problems associated with unstructured groups, such as social loafing and unnecessary repetition of information (Savadori et al., 2001). Consequently, structured groups often produce more accurate judgments than unstructured groups (Rowe & Wright, 1999). If deception detection is enhanced by using asynchronous groups, then this represents a quick and effective method for boosting the deception of detection across a wide variety of domains, from business and politics to law enforcement and intelligence services. We hypothesise that structured groups will outperform individuals in terms of deception detection accuracy.

The present research also investigates the effect of group size on accuracy. Group size is one of the most fundamental dimensions along which groups vary, and one of the easiest variables to change in most groups. It would therefore be of great pragmatic value to establish whether larger groups confer a deception detection advantage. Relatively little research has investigated the effect of interacting group size on decision making quality. However, the extant literature suggests that larger groups generally make more accurate decisions. For example, Yetton and Bottger (1983) found that performance on a judgment task increased as group sizes rose from 3 to 5 members. Similarly, studying group sizes of 1, 2, 5 and 10, Littlepage (1991) found a linear relationship between group size and performance on both additive and disjunctive problem-solving tasks. In the area of computer-mediated decision making, Hwang & Guynes (1994) reported that groups of nine made higher quality decisions than groups of three. However, findings in the area are inconsistent and it appears that group size effects are sensitive to the type of task being undertaken by the group (Laughlin et al. 2006). For example, Bray et al. (1978) reported that task difficulty interacts with group size to influence group performance, with group size effects only being found for problems of moderate difficulty.

Research on this issue in the deception domain is very limited. The only previous empirical investigation into real groups used a very restricted range of group sizes (groups of 6 or 8) and investigated the detection of deception within interacting groups rather than using groups to detect deception in external targets (Zhou, 2013). Because increasing asynchronous group sizes is often relatively easy both in the lab and in applied settings, it is important to establish whether doing so represents a simple and reliable means of enhancing deception detection accuracy. In line with previous research demonstrating a significant effect of group size on performance in other domains, we hypothesise that deception detection accuracy will increase with group size.

Method

Participants

250 participants (170 females, 79 males, and 1 non-binary) took part in the main study for a small monetary reward. A power analysis conducted in G*Power 3.1 suggested that for a 2x4 mixed ANOVA, a total sample size of 136 would be required to detect our hypothesised, medium sized effects. Participants were recruited over Prolific Academic, an online participant recruitment platform. Stimuli were presented and data recorded using Qualtrics. The mean age of participants was 36.3 years (SD = 13.3). Participants were given £4.00 for completing the study.

Design

The study design was unbalanced – see Table 1. There were 4 experimental groups and one control, with 50 participants in each group. In Table 1, ‘Response Frequency’ refers to the number of responses fed back to each participant (each response corresponds to a virtual group member) on each of 16 trials. In each group, the ‘Response Type’ consisted of two judgments – Judgment 1 and Judgment 2 – by each participant. Between the two judgments, participants in all experimental groups received feedback from their virtual group members consisting of their judgment and their rationale for their judgment – this was again for each of 16 trials. In the control group, participants simply received summaries of the initial holiday stories as feedback on each trial.

The effects of feedback were investigated as a within-subjects factor and the effects of group size as a between-subjects factor (for the experimental groups only, thus 4 levels). The dependent variable was deception detection accuracy, which was measured using sensitivity (*d'*), a signal detection statistic. To investigate whether participants who get feedback improve between rounds more than participants who don't get feedback (but simply see a summary of the initial written statement) we compared the two groups shown in italics (i.e. Control and Experimental Group 1).

[Table 1 about here]

Materials

Overview of materials and Simulated Group Response Paradigm (SGRP). SGRP is a method for studying group decision making (Bolger et al., 2020). Virtual groups are created by presenting individual participants with information, advice or other responses collected from previous participants during an earlier response collection exercise. In the context of the current study, the SGRP consisted of four steps. First, an initial set of 200 participants produced a pool of written lies and truths. Second, a new set of 200 participants judged these written statements for veracity so that the 16 statements judged correctly approximately 60% of the time could be selected for use the study. Third, a new set of 50 participants provided explicit veracity judgements for these 16 statements and provided rationales for each of their judgments. These constituted the ‘virtual group member responses’. Finally, 200 participants in the experimental (i.e. virtual group) conditions in the main study provided explicit veracity judgements and rationales for these 16 statements. After making a judgment and providing a rationale, participants were presented with a random selection of judgments and attendant rationales drawn from the pool of judgments and rationales produced by the 50 previous participants. Participants in the main study were thus in a virtual group that gave them advice on whether each of the 16 written statements were truths or lies.

Written statements. 16 written statements were used as stimuli. Half of the statements were truths, half were lies. Each of the statements, which were between 42 and 424 words in length (M = 148.6, SD = 112.7), described a holiday in response to one of the two following requests:

'Please describe, in as much detail as possible, a fictitious account of the most recent holiday you went on - i.e. lie and describe a holiday you never actually went on. This account should be written in such a way that those reading it will believe that it is the truth. Your account should be entirely fictitious: it should not simply be based on another holiday that you went on, or a holiday taken and described to you by someone you know.'

or

'Please describe truthfully, in as much detail as possible, the most recent holiday you went on. This account should be written in such a way that those reading it will believe that it is the truth.'

The 16 statements used in the study were selected from a pre-existing pool of 100 lies and 100 truths created during a stimulus creation exercise. The stimulus selection exercise consisted of a unique sample of 200 participants drawn from the same population as the main study (i.e.

Prolific Academic). The sample was split into 4 groups of 50 participants, with each group judging a different set of 25 lies and 25 truths. The 8 truthful statements judged with an overall detection accuracy rate closest to 60% were used as stimuli in the main study, as were the 8 deceptive statement judged with an overall detection accuracy rate closest to 60%. The truthful statements on average had an accuracy rate of 60% (SD = 1.9%), whilst the deceptive statements on average had an accuracy rate of 59.8% (SD = 3.5%). Statements judged with 60% accuracy were used because previous research has found that the veracity of *written* statements is judged at levels no better than chance (Masip et al., 2012) and therefore interventions to boost deception detection performance, such as training or measures to elicit cues, are required before group effects would be expected to emerge. Baseline detection rates were artificially increased to 60% so as to mimic the increase such successful interventions would have on performance (e.g. Colwell et al., 2012; deTurck et al., 1997; Porter et al., 2000; Stanley & Webster, 2019; Vrij, Evans, et al., 2004).

Virtual group member responses. 50 deception judgments and attendant rationales were elicited in respect of each of the 16 written statements selected for use in the main study. These judgments and rationales were obtained from 50 unique participants, drawn from the same population as those in the main study (i.e. Prolific Academic), who completed a short deception detection task created using the 16 written stimuli described above. The computer-based task, run using Qualtrics, displayed a written statement followed by an on-screen request to indicate whether the author was lying or telling the truth. After answering the question, participants provided a written rationale for their judgment. This process continued until all 16 statements had been viewed and judged. The order in which statements were displayed was randomised across participants.

Procedure

Participants in the main study completed a short deception detection task created using the 16 written stimuli. The procedure was the same as for the virtual group member response elicitation exercise described above, with one exception: on clicking ‘Next’ after having provided a judgment and rationale, participants in the experimental conditions were presented onscreen with the judgments and rationales of 1, 5, 9, or 13 virtual group members (that is, participants were in virtual groups of 2, 6, 10 or 14 respectively), randomly selected from the pool of 50 virtual group member responses. The selection of judgments and rationales displayed to participants was randomised across participants. Participants were asked to read all of these responses before restating or revising their own, initial judgment and rationale. This process continued until all 16 statements had been viewed and judged.

A control condition was run to investigate the possibility that improvements in accuracy between initial and revised judgments are not simply a result of having the opportunity to revise one’s guess after further consideration, but the result of group influence. The procedure for the control condition was identical to that for the experimental condition with a nominal group size of 2, except that, rather than being presented with a virtual group members’ judgment of the message’s veracity and a rationale for that judgment, participants instead viewed a short summary of the message. For example, ‘The paragraph describes a holiday in Dublin, which included exploring botanical gardens, visiting several bars, and meeting some relatives for the first time’ was the short summary presented in respect of one of the statements. Whereas participants in the experimental condition were presented with judgments and rationales that were selected from a large pool of responses and varied randomly across participants,

participants in the control condition were each presented with the same set of 16 summaries, one per trial. The summaries were written by the research team and were approximately matched in terms of word count with the rationales presented to participants in the experimental condition with a group size of two (mean word count = 23.63 and 22.6 respectively). The control condition was designed to match the experimental condition with a nominal group size of 2, because matching with a group size larger than 2 would necessitate the production of multiple, unique summaries for each of the written statements, something that, because of the short length of many of the statements, would be very difficult to do.

Results

Taking all conditions together, participants had an overall baseline deception detection accuracy rate of 64.63% at round 1 and 66.58% at round 2. Truths were more successfully detected than lies (71.45% at round 1 for truths; 57.8% at round 1 for lies). Table 1 presents accuracy rates by group size, veracity condition and round.

[Table 2 about here]

Accuracy of judgments was measured by calculating sensitivity (*d'*), a signal detection statistic. Sensitivity is a measure of perceivers' ability to discriminate between truths and lies. Accurately judging a lie produces a 'hit', whereas inaccurately judging a truth as a lie produces a 'false alarm'. Sensitivity is calculated by subtracting the z-transformed false alarm rate from the z-transformed hit rate (see Green & Swets, 1966). Higher *d’* scores represent greater accuracy rates. Before calculating the hit and false alarm rates, a loglinear transformation was applied to the data so as to address the statistically problematic issue of having hit rates or false alarm rates of 1 or 0. This was done by adding 0.5 to both the number of hits and the number of false alarms and adding 1 to both the number of signal (i.e. lie) trials and the number of noise (i.e. truth) trials. The loglinear approach works well (Brown & White, 2005) and is widely used in signal detection research.

We began our analyses with an initial test of whether increases in accuracy between rounds might at least in part be the result of something other than group influence, such as the benefit arising from having the opportunity to revise one’s initial judgment. To investigate this, we performed a 2 (judgment round: round 1, round 2) x 2 (feedback type: other participant’s judgment of message, message summary) mixed ANOVA with detection accuracy scores as the dependent variable. To ensure maximum comparability between control and experimental conditions, the ANOVA was performed on the control condition data and the data from the experimental condition with a group size of 2. The interaction effect was not significant, $F(1,98) = 3.91$, $p=.051$, partial $\eta^2 = .038$, suggesting that the increase in detection accuracy between rounds in the group conditions may have been caused at least in part by having the opportunity to revise one’s original guess.

To check that the four group size levels did not differ in terms of the percentage of correct advice given to them, a one-way between-subjects ANOVA was performed with percentage of correct statements shown to participants as the dependent variable. There was no significant difference between groups in terms of percentage of correct statements displayed to participants, *F*(3,196) = .06, *p* = .98, partial η^2 = 0. The average percentage of statements shown to participants that were correct was 69.67% (SD = 30.42).

Table 2 presents the means and standard deviations for each group size's detection accuracy. To further investigate whether group-based judgments (i.e. round two judgments) outperformed individual judgments (i.e. round one judgments), and whether any such improvements varied as a function of group size, a 2 (judgment round: round 1, round 2) x 4 (group size: 2, 6, 10, 14) mixed ANOVA with detection accuracy scores as the dependent variable was performed on the data. In line with our expectation that structured groups would have higher deception detection accuracy rates than individuals, there was a significant main effect of judgment round on accuracy, with second round (i.e. group-informed judgments) outperforming initial, individual judgments, *F*(1,196) = 26.92, *p* <.001, partial η^2 = .121. Contrary to our expectations, and despite the consistent increase in accuracy between rounds as group size increases, there was no significant interaction effect between judgment round and group size on deception detection accuracy, *F*(3,196) = 1.64, *p*=.18. To examine changes in accuracy between judgments within each group size level, a series of paired t-tests were performed comparing detection accuracy scores on round 1 to those on round 2 for each of the four group sizes. There was no significant difference in performance between first and second judgments for group sizes of 1 (i.e. the control condition), *t*(49) = 1.76, *p* = .08, and 2, *t*(49) = .93, *p* = .36, but there were significant improvements in accuracy between first and second judgments for group sizes of 6, *t*(49) = 2.78, *p* = .01, 10, *t*(49) = .2.47, *p* = .02, and 14, *t*(49) = 4.34, *p* <.001.

[Table 3 about here]

In order to investigate whether the likelihood of opinion change was influenced by the percentage of contradictory opinions displayed after the first judgment was made, we created a generalised linear mixed model. Whether the participant changed their guess on the trial was the binary dependent variable. The independent variable was the percentage of opinions shown to participants on that trial that contradicted the participant's own initial judgment. Because statement number might conceivably influence both percentage of contradictory opinions and willingness to change one's mind, it was added to the model as a level two variable. We allowed intercepts and slopes to vary across the level two variable. Percentage of contradictory advice did not significantly predict whether participants changed their mind on their second guess, $\beta = 0$, $z = -.18$, $p = .86$.

Discussion

The present study sought to uncover whether use of a structured group decision making technique boosted deception detection performance over that of individuals. The results of previous research had provided some evidence that unstructured groups outperform individuals, but whether structured groups also outperform individuals was an open question. We found that structured groups do indeed have higher accuracy rates than individuals. The size of the structured groups had no significant effect on performance: larger group sizes were not better at detecting deception than smaller groups.

Taken together, the results suggest that increases in accuracy between rounds were caused at least in part by group influence. An initial analysis comparing dyads to individuals in terms of improvements in accuracy between rounds did not find a significant difference between these two groups, and so failed to find support for the hypothesis that it was group influence rather than some other, confounding variable that was responsible for improvements in performance between rounds. However, subsequent analyses revealed that whereas performance in the control condition, where individuals are not exposed to group feedback, did not significantly improve between rounds, it did in three of the four group conditions, suggesting positive effects of group influence on deception detection accuracy. The finding that groups benefit deception detection supports our first hypothesis and is in line with the results of several previous studies (Frank, 2004; Klein & Epley, 2015; McHaney et al., 2015). These results provide further evidence for the superior performance of small groups at detecting deception. They also provide the first evidence that asynchronous, structured groups outperform individuals on deception judgments tasks.

Contrary to our second hypothesis, group size did not influence deception detection accuracy. This finding is in line with research by Zhou et al. (2013), who found that groups of 8 were no better than groups of 6 at detecting deception. However, the consistent, albeit non-significant, trend for detection accuracy scores to improve with increasing group size suggests that future research on the effect of group size on accuracy is warranted. Future research should investigate what task and group factors moderate the benefit of group deception detection. For example, if the content of the lies requires specialist knowledge to produce, then larger groups might be more efficacious, simply because increasing with number of judges is the chance that a group member will have the expertise to evaluate the validity of the semantic content of the deceptive message. Little specialist knowledge was required by participants when evaluating lies in the present research. Similarly, deception detection tasks involving more complex stimuli might benefit from larger groups. The deception detection task used in the present research utilised only written stimuli. It is possible that deception detection tasks that utilise videos, and so involve a greater quantity of information on which to base veracity judgments, would benefit even more from larger groups, because larger groups might be likely to notice and utilise more of the large number of cues to deception. More broadly, group effects, and the benefits of larger groups, would be expected to be greater when there is more useful information pertinent to veracity to be exchanged, whether that be cues or other aspects of message content.

Two limitations of the present research are worth noting. First, the group sizes investigated in the present research ranged from 2 to 14. Although it is unlikely that within this range there is an optimal group size that we did not investigate, it is not improbable that larger group sizes than those studied might perform significantly differently from the ones in the present study. In particular, it is possible that significantly larger groups might perform more poorly. When the number of group members' responses presented to participants is very large, participants might struggle to easily perceive which judgment, 'truth' or 'lie', is the majority judgment, which should in turn negatively impact their ability to be influenced by the opinion of the group.

Second, a potential criticism of the study might be that its main result, the significant effect of group on deception detection performance, was an artefact of the experimental design. That is, because the baseline deception detection accuracy rate was artificially set at 60%, and so, on average, the majority of group members were correct, and because decision makers are often swayed by the majority opinion (Glynn, 1997; White & Dahl, 2006; Wood, 2000), it was perhaps not surprising that participants in the present study often emerged from the group process with a more accurate judgment than the one they started with. However, majority influence does not appear to be the source of improvements in accuracy in the present study: the proportion of contradictory advice given to participants did not significantly influence whether they changed their minds between their first and second guesses. Also, and importantly, it should be noted that even without artificial base rate inflation, as long as baseline deception detection performance is at above chance levels, the majority effect should lead to a group advantage over individuals. This combination of inflated base rate accuracy and majority effect is not the source of an artefactual effect, but rather it is one of the mechanisms by which group decision making processes improve deception detection accuracy. Nonetheless, it would be prudent to attempt to replicate these results using written statements which have had their baseline accuracy boosted by means of more natural interventions, such as training in deception cue use. The use of such natural interventions would also help address other issues with the study, such as the effect of minimal variation across stimuli, the implications of which are currently unclear.In conclusion, the present study extended previous research on deception detection by investigating the effect of structured groups on deception detection performance. Across a variety of group sizes, group-based veracity judgments were more accurate than individual judgments. Group size, however, did not have a significant effect on performance. These results provide further evidence of the relative efficacy of group deception detection and demonstrate that the group advantage demonstrated by unstructured groups extends to asynchronous, structured groups, and point to the future potential of harnessing the power of asynchronous groups when detecting deception.

References

Bolger, F., Rowe, G., Belton, I., Crawford, M. M., Hamlin, I., Sissons, A., Taylor Browne Lūka, C., Vasilichi, A., & Wright, G. (2020, June 27). *The Simulated Group Response Paradigm: A new approach to the study of opinion change in Delphi and other structured-group techniques.* https://doi.org/10.31219/osf.io/4ufzg

Bond, C. F. & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*, 214-234. https://doi.org/10.1207/s15327957pspr1003_2

Bonner, B. L., Baumann, M. R., & Dalal, R. S. (2002). The effects of member expertise on group decision-making and performance. *Organizational Behavior and Human Decision Processes, 88,* 719-736. https://doi.org/10.1016/S0749-5978(02)00010-9

Bonner, B. L., Sillito, S. D., & Baumann, M. R. (2007). Collective estimation: Accuracy, expertise, and extroversion as sources of intra-group influence. *Organizational Behavior and Human Decision Processes, 103*, 121-133 https://doi.org/10.1016/j.obhdp.2006.05.001

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology*, *36,* 1224 -1240. https://doi.org/10.1037/0022-3514.36.11.1224

Brown, G. S. & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods*, *37*, 436–449. https://doi.org/10.3758/BF03192712

Colwell, L. H., Colwell, K. W., Hiscock-Anisman, C. K., Hartwig, M., Cole, L., Werdin, K., & Youschak, K. (2012). Teaching Professionals to Detect Deception: The Efficacy of a Brief Training Workshop. *Journal of Forensic Psychology Practice*, *12*, 68-80. https://doi.org/10.1080/15228932.2012.629592

Culhane, S. E., Kehn, A., Hatz, J., & Hildebrand, M. (2015). Are two heads better than one? Assessing the influence of collaborative judgements and presentation mode on deception detection for real and mock transgressions. *Journal of Investigative Psychology and Offender Profiling, 12*(2), 158-170. https://doi.org/10.1002/jip.1424

deTurck, M. A., Feeley, T. H., & Roman, L. A. (1997). Vocal and visual cue training in behavioral lie detection. *Communication Research Reports, 14*, 249-259. https://doi.org/10.1080/08824099709388668

Flores, B. E., & White, E. M. (1989). Subjective vs objective combining of forecasts: An experiment. *Journal of Forecasting, 8*, 331-341. https://doi.org/10.1002/for.3980080314

Frank, M. G., Paolantonio, N., Feeley, T. H., & Servoss, T. J. (2004). Individual and Small Group Accuracy in Judging Truthful and Deceptive Communication. *Group Decision and Negotiation, 13*, 45-59. https://doi.org/10.1023/B:GRUP.0000011945.85141.af

Glynn, C. J., Hayes, A. F., & Shanahan, J. (1997). Perceived support for one's opinions and willingness to speak out a metaanalysis of survey studies on the "spiral of silence". *Public Opinion Quarterly, 61*, 452-463. https://doi.org/10.1086/297808

Green, D.M. & Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. Wiley.

Hartwig, M., & Bond, C. F., Jr. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology, 28*, 661-676. https://doi.org/10.1002/acp.3052

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve detection of deception? A meta-analysis. *Communication Research, 43*, 283-343. https://doi.org/10.1177/0093650214534974

Henry, R. A. (1995). Improving group judgment accuracy: Information sharing and determining the best member. *Organizational Behavior and Human Decision Processes, 62*, 190-197. https://doi.org/10.1006/obhd.1995.1042

Hwang, H. & Guynes, J. (1994). The effect of group size on group performance in computer-supported decision making. *Information & Management, 26*, 189-198 https://doi.org/10.1016/0378-7206(94)90092-2

Klein, N., & Epley, N. (2015). Group discussion improves lie detection. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 112*, 7460-7465. https://doi.org/10.1073/pnas.1504048112

Laughlin, P. R., Gonzalez, C. M., & Sommer, D. (2003). Quantity estimations by groups and individuals: Effects of known domain boundaries. *Group Dynamics: Theory, Research, and Practice, 7*, 55-63. https://doi.org/10.1037/1089-2699.7.1.55

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: effects of groups size. *Journal of Personality and Social Psychology, 90*, 644-651. https://doi.org/10.1037/0022-3514.90.4.644

Littlepage, G. E. (1991). Effects of group size and task characteristics on group performance: A test of Steiner's model. *Personality and Social Psychology Bulletin, 17*, 449-456. https://doi.org/10.1177/0146167291174014

Masip, J., Bethencourt, M., Lucas, G., Segundo, M. S., & Herrero, C. (2012). *Deception detection from written accounts. Scandinavian Journal of Psychology*, *53*, 103-111. https://doi.org/10.1111/j.1467-9450.2011.00931.x

McHaney, R. W., George, J. F., & Gupta, M. (2015). An Exploration of Deception Detection Are Groups More Effective Than Individuals? *Communication Research*, *45*, 1103-1121. https://doi.org/10.1177/0093650215607627

Moi, W. Y., & Shanks, D. R. (2015). Can lies be detected unconsciously? *Frontiers in Psychology, 6*, 1221. https://doi.org/10.3389/fpsyg.2015.01221

Park, E. S., Levine, T. R., Harms, C. M., & Ferrara, M. H. (2000). Group and individual accuracy in deception detection. *Communication Research Reports, 19*, 99-106. https://doi.org/10.1080/08824090209384837

Porter, S., Woodwirth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, *24*, 643-658. https://doi.org/10.1023/A:1005500219657

Rowe, G. & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15,* 353-375. https://doi.org/10.1016/S0169-2070(99)00018-7

Savadori, L., Van Swol, L. M., & Sniezek, J. A. (2001). Information sampling and confidence within groups and judge advisor systems. *Communication Research, 28*, 737-771.

Sniezek, J. A. & Buckley, T. (1995). Cueing and cognitive conflict in judge–advisor decision making. *Organizational Behavior and Human Decision Processes, 62*, 159–174. https://doi.org/10.1006/obhd.1995.1040

Stanley, J. T., & Webster, B. A. (2019). A comparison of the effectiveness of two types of deceit detection training methods in older adults. *Cognitive Research: Principles & Implications*, *4*, 26. https://doi.org/10.1186/s41235-019-0178-z ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some Evidence for Unconscious Lie Detection. *Psychological Science, 25*, 1098-1105. https://doi.org/10.1177/0956797614524421

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Detecting Deceit Via Analyses of Verbal and Nonverbal Behavior in Children and Adults. *Human Communication Research, 30,* 8-41. https://doi.org/10.1111/j.1468-2958.2004.tb00723.x

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24,* 239-263. https://doi.org/10.1023/A:1006610329284

Vrij, A., Evans, H., Akehurst, L., & Mann, S. (2004). Rapid judgments in assessing verbal and nonverbal cues: Their potential for deception researchers and lie detection. *Applied Cognitive Psychology*, *18*, 283-296. https://doi.org/10.1002/acp.964

Vrij, A., & Granhag, P. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition, 1*, 110-117. https://doi.org/10.1016/j.jarmac.2012.02.004

White, K., & Dahl, D. W. (2006). To be or not be: The influence of dissociative reference groups on consumer preferences. *Journal of Consumer Psychology*, *16*, 404-414. https://doi.org/10.1207/s15327663jcp1604_11

Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, *51*, 539-570. https://doi.org/10.1146/annurev.psych.51.1.539

Yetton P, & Bottger P. (1983). The relationships among group size, member ability, social decision schemes, and performance. *Organizational Behavior and Human Performance. 32*, 145-159. https://doi.org/10.1016/0030-5073(83)90144-7

Zhou, L., Zhang, D., and Sung, Y. (2013). The Effects of Group Factors on Deception Detection Performance. *Small Group Research*, *24*, 272-297 https://doi.org/10.1177/1046496413484178

Table 1. Summary of experimental design

Group	n	Response Frequency		Response Type n	
Control	*50*	*1*	*Judgment 1*	*Summary only*	*Judgment 2*
Experiment 1	*50*	*1*	*Judgment 1*	*Judgment + Rationale*	*Judgment 2*
Experiment 2	50	5	Judgment 1	Judgment + Rationale	Judgment 2
Experiment 3	50	9	Judgment 1	Judgment + Rationale	Judgment 2
Experiment 4	50	13	Judgment 1	Judgment + Rationale	Judgment 2

Table 2. Accuracy rate by veracity condition, group size and round (n=50 for each condition)

	Accuracy (%)					
Group size	Round 1 lies	Round 2 lies	Round 1 truths	Round 2 truths	Round 1 total	Round 2 total
1 [control condition]	52.50	45.50	69.75	71.25	61.13	58.38
2	60.50	62.25	70.00	70.25	65.25	66.25
6	57.00	61.00	73.75	75.25	65.38	68.13
10	61.25	66.50	74.25	75.25	67.75	70.88
14	57.75	64.25	69.50	71.50	63.63	67.88

Table 3. Means and standard deviations for detection accuracy (‘d) by group size and round (n=50 for each condition)

Group size	Round 1 accuracy (`d)		Round 2 accuracy (`d)	
	M	SD	M	SD
1 [control condition]	.59	.69	.45	.67
2	.81	.70	.86	.66
6	.8	.74	.96	.69
10	.94	.58	1.11	.61
14	.69	.69	.92	.71