

Group discussion improves lie detection

Nadav Klein¹ and Nicholas Epley¹

Booth School of Business, University of Chicago, Chicago, IL 60637

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved May 6, 2015 (received for review March 2, 2015)

Groups of individuals can sometimes make more accurate judgments than the average individual could make alone. We tested whether this group advantage extends to lie detection, an exceptionally challenging judgment with accuracy rates rarely exceeding chance. In four experiments, we find that groups are consistently more accurate than individuals in distinguishing truths from lies, an effect that comes primarily from an increased ability to correctly identify when a person is lying. These experiments demonstrate that the group advantage in lie detection comes through the process of group discussion, and is not a product of aggregating individual opinions (a “wisdom-of-crowds” effect) or of altering response biases (such as reducing the “truth bias”). Interventions to improve lie detection typically focus on improving individual judgment, a costly and generally ineffective endeavor. Our findings suggest a cheap and simple synergistic approach of enabling group discussion before rendering a judgment.

lie detection | group decision-making | social cognition | wisdom of crowds | mind reading

Detecting deception is difficult. Accuracy rates in experiments are only slightly greater than chance, even among trained professionals (1–4). This meager accuracy rate appears driven by a modest ability to detect truths rather than lies. In one meta-analysis, individuals accurately identified 61% of truths, but only 47% of lies (5). These results have led researchers to develop costly training programs targeting individual lie detectors to increase accuracy (6–10). We test a different strategy: asking individuals to detect lies as a group.

There are three reasons that groups might detect deception better than individuals. First, because individuals have some skill in distinguishing truths from lies, statistically aggregating individual judgments could increase accuracy (a “wisdom-of-crowds” effect) (11, 12). If individuals detect truths better than lies, aggregating individual judgments would increase truth detection more than lie detection.

Second, individuals show a reliable “truth bias,” assuming others are truthful unless given cause for suspicion (5, 13). If groups are less trusting than individuals (14–15), then they could detect lies more accurately because they guess someone is lying more often.

Finally, group deliberation could increase accuracy by providing useful information that individuals lack otherwise (16–18). This predicts that group discussion alters how individuals evaluate a given statement to increase accuracy. Because individuals already possess some accuracy in detecting truths, unique improvement from group discussion would increase accuracy in detecting lies.

We know of only two inconclusive experiments that test a group advantage in lie detection. In one experiment, participants first made an individual judgment before group discussion, making the independent influence of the subsequent group discussion unclear (17). Although groups were no more accurate than individuals overall, they were marginally better ($0.05 < P < 0.10$) detecting lies. In the other experiment, groups were no more accurate than individuals (19), but this experiment sampled only two targets, leaving open the possibility of stimulus-specific confounds.

We therefore designed four experiments to directly test whether groups could detect liars better than individuals, and, if so, why. Existing research demonstrates that increasing incentives for accuracy among lie detectors does not increase accuracy, but that

increasing incentives for effective deception among lie tellers can make lies easier to detect (5). We therefore did not manipulate lie detectors’ incentives to detect truths vs. lies accurately, but instead asked participants to detect truths vs. lies in low-stakes (experiments 1, 2, and 4) and high-stakes contexts (experiment 3) for the lie tellers.

Experiments 1 and 2

Throughout this article, we refer to “real groups” when describing lie detection results obtained by group discussion and “nominal groups” when describing results obtained by statistically aggregating individual guesses.

In experiment 1, real groups were more accurate [mean (M) = 61.7%, $SD = 18.2\%$] than individuals [$M = 53.6\%$, $SD = 16.0\%$; $t(118) = 2.32$, $P = 0.02$, $d = 0.52$, 95% $CI_{\text{difference}}$: 1.2%, 15.0%]. As Table 1 shows, the group advantage came primarily from detecting lies more accurately than individuals [$t(118) = 2.80$, $P < 0.01$, $d = 0.57$, 95% $CI_{\text{difference}}$: 3.7%, 21.6%]. There was no group advantage when detecting truths [$t(118) = 0.73$, $P = 0.47$, $d = 0.13$, 95% $CI_{\text{difference}}$: -6.1%, 13.1%].

The group advantage in lie detection, but not truth detection, could come from a response bias if groups are more likely to guess that someone is lying. However, there was no significant difference in the frequency of guessing “truth” between groups ($M = 50.3\%$, $SD = 11.0\%$) and individuals [$M = 53.6\%$, $SD = 14.4\%$; $t(118) = 1.12$, $P = 0.26$, $d = 0.26$]. In addition, a linear regression predicting overall accuracy from condition (individual vs. group), controlling for the propensity to guess truth, still yielded a significant effect of condition ($\beta = 0.08$, $t = 2.32$, $P = 0.02$). Groups were not better lie detectors because they were more likely to guess that someone was lying.

A signal detection analysis confirms these results. This analysis calculates two parameters, d' (which represents participants’ skill

Significance

Detecting lies is difficult. Accuracy rates in experiments are only slightly greater than chance, even among trained professionals. Costly programs aimed at training individual lie detectors have mostly been ineffective. Here we test a different strategy: asking individuals to detect lies as a group. We find a consistent group advantage for detecting small “white” lies as well as intentional high-stakes lies. This group advantage does not come through the statistical aggregation of individual opinions (a “wisdom-of-crowds” effect), but instead through the process of group discussion. Groups were not simply maximizing the small amounts of accuracy contained among individual members but were instead creating a unique type of accuracy altogether.

Author contributions: N.K. and N.E. designed research; N.K. performed research; N.K. analyzed data; and N.K. and N.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. Email: nklein@chicagobooth.edu or epley@chicagobooth.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1504048112/-DCSupplemental.

Table 1. Lie detection accuracy of individual judgments, group discussion, nominal groups, and simulated nominal groups in experiment 1

Condition	Accuracy, %			"Truth" guesses
	Overall	Truths	Lies	
Real groups	61.7 _a *	60.6 _a *	63.9 _a *	50.3
Individuals	53.6 _b *	57.0 _{a,b} *	51.2 _b	53.6*
Nominal groups				
Simultaneous	54.0 _{a,b} *	58.0 _{a,b} *	51.7 _b	54.0*
Simulated group of 3	54.5 _b *	58.4 _{a,b} *	52.7 _b *	54.3*
Simulated group of 7	56.9 _a *	62.1 _a *	53.4 _b *	55.9*
Simulated group of 11	57.7 _a *	64.3 _a *	53.2 _b *	57.2*
Simulated group of 15	57.1 _a *	65.1 _a *	50.9 _b	58.7*
Aggregated real groups	70.0	74.7	68.7	55.3
Best individuals	65.3 _a *	66.1 _a *	64.4 _a *	50.7

All means are reported as percentages. In the accuracy columns, means that do not share the same subscript within columns differ at $P < 0.05$. Simulated nominal groups are compared with real groups and individuals on one-sample t tests. The aggregated real groups figures are provided for descriptive purposes only as a result of lack of sufficient degrees of freedom to perform statistical tests.

*Means differ from 50% (chance responding) at $P < 0.05$.

distinguishing lies from truths) and C (which measures response bias). Consistent with the overall accuracy rates, groups ($M = 1.65$, $SD = 3.49$) achieved higher d' scores than individuals [$M = 0.45$, $SD = 2.35$; $t(118) = 2.13$, $P = 0.04$, $d = 0.40$], but we observed a nonsignificant difference in C scores between groups ($M = -0.22$, $SD = 1.37$) and individuals [$M = 0.02$, $SD = 1.13$; $t(118) = 0.95$, ns , $d = 0.19$].

To assess whether the statistical aggregation of individual judgments can increase accuracy, we created two different types of nominal groups (see Table 1). The first type of nominal group was of individuals who participated simultaneously in the same session. Recall that participants in the individual condition were recruited in groups of three, but provided their judgments in isolation. We calculated a majority vote among these "simultaneous" individuals to calculate the nominal group accuracy (identical to averaging in this context). As shown in Table 1, these simultaneous nominal groups performed no better than individuals.

The second type of nominal groups came from a simulation method that sampled 1,000 randomly selected groups of sizes 3, 7, 11, and 15 from within each sequence of video clips, calculating accuracy based on the majority rule for each video. As Fig. 1 shows, increasing the nominal group size increased overall accuracy significantly compared with individuals, although these gains came from detecting truths more accurately rather than from detecting lies more accurately. Aggregating individual judgments into nominal groups appears to increase the amount of accuracy already contained in the individual judgments, namely the ability to detect truths, but it does not mirror the increased accuracy in detecting lies that we observe with real groups (of only three members) following actual discussion.

Experiment 2 is a replication of experiment 1 using different statements and nearly doubling the sample size. Experiment 2 replicated all major results from experiment 1. Real groups were again more accurate ($M = 60.3\%$, $SD = 16.1\%$) than were individuals [$M = 53.6\%$, $SD = 15.7\%$; $t(233) = 2.83$, $P = 0.005$, $d = 0.42$, 95% $CI_{\text{difference}}$, 2.1%, 11.5%], with real groups detecting lies ($M = 56.6\%$, $SD = 21.9\%$) more accurately than individuals [$M = 46.4\%$, $SD = 21.6\%$; $t(233) = 3.09$, $P = 0.002$, $d = 0.47$, 95% $CI_{\text{difference}}$, 3.7%, 16.6%], but not detecting truths more accurately ($M = 64.1\%$, $SD = 21.8\%$) than individuals [$M = 60.7\%$, $SD = 21.9\%$; $t(233) = 1.04$, $P = 0.30$, $d = 0.16$, 95% $CI_{\text{difference}}$, -3.1%, 10.0%].

Real groups were again no more likely to guess that someone was lying ($M = 46.2\%$, $SD = 14.7\%$) than individuals [$M = 43.2\%$, $SD = 15.2\%$; $t(233) = 1.31$, $P = 0.19$, $d = 0.17$]. A linear regression predicting overall accuracy from condition (individual vs. group) and controlling for the propensity to guess truth still yielded a significant effect of condition ($\beta = 0.07$, $t = 2.85$, $P < 0.005$), suggesting that groups outperformed individuals above and beyond their nonsignificantly higher propensity to guess "lie."

In a signal detection analysis, groups again achieved a higher d' score ($M = 1.52$, $SD = 2.95$) than individuals [$M = 0.46$, $SD = 2.38$; $t(233) = 2.77$, $P < 0.01$, $d = 0.40$], but no significant difference emerged in C scores between groups ($M = 0.11$, $SD = 1.37$) and individuals [$M = 0.40$, $SD = 1.22$; $t(233) = 1.57$, $P = 0.12$, $d = 0.22$].

We again tested whether statistical aggregation of individual judgments increases accuracy. To do so, we created the same two types of nominal groups from experiment 1: (i) the "simultaneous individuals" who made judgments in the same experimental session and (ii) the simulated groups that sampled 1,000 randomly selected groups of sizes 3, 7, 11, and 15 from within each sequence of video clips, calculating accuracy based on the majority rule for each video.

As Table 2 shows, neither the simultaneous individuals nor the simulated groups performed significantly better than individuals (and sometimes performed significantly worse) in overall accuracy and in detecting lies accurately. For accuracy in detecting truths, however, the simulated groups performed significantly better than individuals starting at a group size of seven. These results again suggest that aggregating individual judgments increases accuracy when individuals already possess some skill, namely amplifying modest abilities in detecting truths. It does not mimic the real group advantage observed in detecting lies.

These results suggest that the group advantage in lie detection comes from increased skill in lie detection. Additional suggestive evidence for increased skill comes from a third variety of nominal groups: statistically aggregating across real groups' judgments. We did this by using the majority rule to determine the real groups' collective judgment about each video for each of the four video sequences in both experiments 1 and 2 (see Tables 1 and 2). Although this leaves too few observations for statistical comparisons, aggregated real groups descriptively improved overall accuracy, lie detection, and truth detection compared with real groups and individuals. This suggests that real groups create skill

Table 2. Lie detection accuracy of individual judgments, group discussion, and nominal groups, and simulated nominal groups in experiment 2

Condition	Accuracy, %			"Truth" guesses
	Overall	Truths	Lies	
Real groups	60.3 _a *	64.1 _{a,c} *	56.6 _a *	53.8*
Individuals	53.6 _b *	60.7 _a *	46.4 _b *	56.8*
Nominal groups				
Simultaneous	54.4 _b *	61.7 _a *	47.0 _b *	57.4*
Simulated group of 3	55.8 _b *	65.3 _{a,c} *	46.2 _b *	60.5*
Simulated group of 7	52.6 _b *	70.5 _c *	46.4 _b *	65.0*
Simulated group of 11	47.3 _c *	73.4 _d *	46.7 _b *	68.2*
Simulated group of 15	49.0 _c *	74.8 _d *	47.1 _b *	71.0*
Aggregated real groups	70.3	74.1	66.6	53.8
Best individuals	61.8 _a *	67.0 _c *	56.6 _a *	55.1*

All means are reported as percentages. In the accuracy columns, means that do not share the same subscript within columns differ at $P \leq 0.053$. Simulated nominal groups are compared with real groups and individuals on one-sample t tests. The aggregated real groups figures are provided for descriptive purposes only as a result of lack of sufficient degrees of freedom for performing statistical tests.

*Means differ from 50% (chance responding) at $P < 0.05$.

Table 3. Lie detection accuracy of individual judgments, group discussion, and nominal groups, and simulated nominal groups in experiment 3

Condition	Accuracy, %			"Truth" guesses
	Overall	Truths	Lies	
Real groups	53.2 _a *	52.2 _{a,b}	54.4 _a *	49.1
Individuals	48.7 _b	52.9 _a	43.8 _b *	54.4*
Nominal groups				
Simultaneous	46.3 _b *	53.2 _{a,b}	38.3 _c *	57.2*
Simulated group of 3	48.2 _a *	54.5 _{a,b} *	37.4 _c *	56.4*
Simulated group of 7	39.7 _c *	56.5 _{a,c} *	38.1 _c *	59.0*
Simulated group of 11	32.8 _c *	57.6 _c *	35.9 _c *	60.6*
Simulated group of 15	29.8 _c *	58.3 _c *	33.5 _c *	62.2*
Aggregated real groups	67.2	61.4	77.0	41.7
Best Individuals	57.5 _d *	71.6 _d *	46.6 _b	54.6*

All means are reported as percentages. In the accuracy columns, means that do not share the same subscript within columns differ at $P < 0.05$. Simulated nominal groups are compared only to real groups and individuals on one-sample t tests. The aggregated real groups figures are provided for descriptive purposes only as a result of lack of sufficient degrees of freedom for performing statistical tests.

*Means differ from 50% (chance responding) at $P < 0.052$.

in detecting lies, which is then amplified when real groups' judgments are statistically aggregated.

Experiment 3

Experiment 3 tested whether the group advantage in lie detection extends to large-stakes and intentional lies. Groups were again more accurate ($M = 53.2\%$, $SD = 15.0\%$) than were individuals [$M = 48.7\%$, $SD = 11.9\%$; $t(176) = 2.24$, $P = 0.0265$, $d = 0.34$, 95% $CI_{\text{difference}}$, 0.54%, 8.56%], with groups detecting lies ($M = 54.4\%$, $SD = 19.1\%$) more accurately than individuals [$M = 43.8\%$, $SD = 18.0\%$; $t(176) = 3.82$, $P = 0.0002$, $d = 0.58$, 95% $CI_{\text{difference}}$, 5.2%, 16.1%], but not detecting truths more accurately ($M = 52.2\%$, $SD = 20.1\%$) than individuals [$M = 52.9\%$, $SD = 19.0\%$; $t(176) = -0.27$, $P = 0.79$].

Groups were significantly more likely to guess that someone was lying ($M = 50.9\%$, $SD = 12.6\%$) than individuals were [$M = 45.6\%$, $SD = 14.5\%$; $t(176) = 2.62$, $P = 0.001$, $d = 0.39$]. However, a linear regression predicting overall accuracy from condition (individual vs. group) and controlling for the propensity to guess "lie" still yielded a significant effect of condition ($\beta = 0.055$, $t = 2.68$, $P = 0.008$). The effect of the propensity to guess lie was also significant but tended to reduce accuracy ($\beta = -0.17$, $t = -2.35$, $P = 0.02$), suggesting that groups outperformed individuals above and beyond—and in this case, despite—their higher propensity to guess that a contestant was lying. A response bias cannot explain groups' improved accuracy.

In a signal detection analysis, groups again achieved a higher d' score ($M = 0.33$, $SD = 1.36$) than individuals ($M = 0.02$, $SD = 1.34$), although only directionally so [$t(176) = 1.53$, $P = 0.13$]. Groups' d' score was nevertheless significantly greater than 0 [$t(90) = 2.31$, $P = 0.023$], indicating that real groups were more accurate than chance guessing. Individuals' d' score was not different from 0 [$t(86) = 0.13$, $P = 0.90$]. No difference emerged in C scores between groups ($M = 0.04$, $SD = 0.62$) and individuals [$M = 0.13$, $SD = 0.78$; $t(176) = 0.79$, $P = 0.43$], suggesting that groups' directional outperformance on their d' score was not a result of groups' greater tendency to guess that a contestant was lying. A linear regression predicting d' from condition (individual vs. group) and controlling for C scores still yielded a significant effect of condition ($\beta = 0.39$, $t = 2.24$, $P = 0.027$). The effect of C scores was also significant ($\beta = 0.98$, $t = 7.82$, $P < 0.001$). Overall, these results suggest that real groups

outperformed individuals above and beyond groups' higher propensity to guess "lie."

We again tested whether statistical aggregation of individual judgments increases accuracy. To do so, we created the same two types of nominal groups from experiment 1 and 2: (i) the "simultaneous individuals" who made judgments in the same experimental session and (ii) the simulated groups that sampled 1,000 randomly selected groups of sizes 3, 7, 11, and 15 from within each sequence of video clips, calculating accuracy based on the majority rule for each video.

Replicating experiments 1 and 2, neither the simultaneous individuals nor the simulated groups performed significantly better than individuals (and sometimes performed significantly worse) in overall accuracy, detecting lies, and detecting truths (see Table 3). These results again demonstrate that aggregating individual judgments does not lead to meaningful increases in accuracy, and does not mimic the advantage of group discussion.

Similar to experiments 1 and 2, statistically aggregating across real groups' judgments led to descriptively higher overall accuracy, higher accuracy in detecting truths, and higher accuracy in detecting lies. Although we again caution that aggregating real groups' judgments leaves too few observations for statistical comparisons, these results nevertheless bolster the conclusion that real groups create skill in detecting lies, which is then amplified when real groups' judgments are statistically aggregated. Indeed, the difference between aggregated groups and aggregated individuals is stark. Whereas aggregated groups of 15 individuals are accurate, overall, only 29.6% of the time, aggregated real groups are accurate 67.2% of the time (Fig. 2). In sum, experiment 3 tested lies that were consequential, intentional, and done for personal gain. Real groups again outperformed individual and nominal groups on overall accuracy. This group advantage was again driven by a marked improvement in detecting lies compared with individuals. Experiments 1–3 suggest that groups can improve lie detection accuracy across very different types of lies.

Experiment 4

Group discussion could increase lie detection skill for two reasons. First, group discussion could identify the most accurate individual within a group, increasing accuracy through a sorting mechanism. Second, group discussion may elicit conjectures and

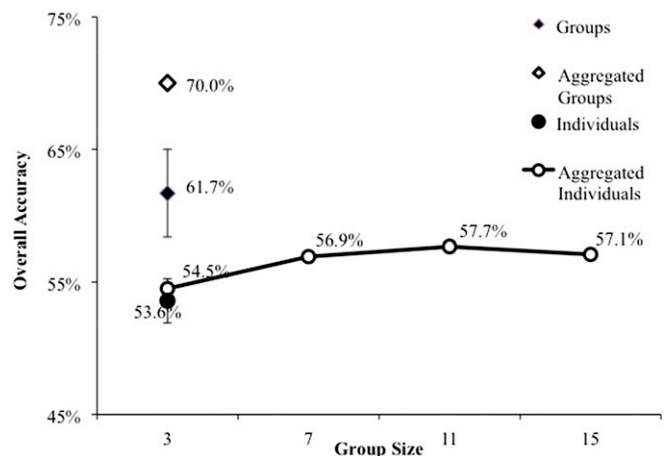


Fig. 1. Overall accuracy of simulated nominal groups relative to group discussion and individual judgments (experiment 1). SEMs are presented for groups and for individuals. Aggregated individuals were created from computer simulations using the data for individuals. Aggregated groups contain only four observations and are displayed for descriptive purposes.

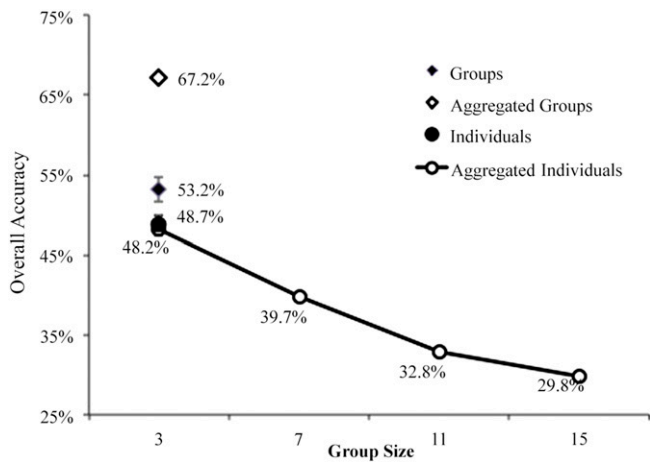


Fig. 2. Overall accuracy of simulated nominal groups relative to group discussion and individual judgments (experiment 3). SEMs are presented for groups and for individuals. Aggregated Individuals were created from computer simulations using the data for individuals. Aggregated groups contain only three observations and are displayed for descriptive purposes.

observations about the target that provide the information needed to make an accurate assessment, increasing accuracy through synergistic mechanism by exposing individuals to each other's preliminary points of view.

The critical difference between sorting and synergy is that individual judgments are formed before discussion on the sorting account and discussion then identifies the best judge, whereas individual judgments are formed during discussion on the synergy account and discussion itself creates a more accurate judgment.

We tested between these mechanisms in experiment 4 by having participants make judgments as a group and individually for each target, manipulating the order in which they did so. The sorting mechanism predicts that making individual judgments first will not affect the subsequent group advantage in lie detection, whereas the synergy mechanism predicts that forming an individual judgment first will disable the group advantage because accuracy comes from the additional information acquired while forming a group opinion.

A 2(judgment: group vs. individual) \times 2(order: individual first vs. group first) mixed-model ANOVA on overall accuracy revealed only a marginally significant main effect for judgment [$F(1, 130) = 2.90, P = 0.09$], with groups being marginally more accurate ($M = 59.5\%$, $SD = 15.1\%$) than individual judgments [$M = 57.7\%$, $SD = 15.0\%$; paired $t(131) = 1.67, P = 0.10$]. This design weakens the distinction between groups and individuals from experiments 1–3 because participants cycle between individual and group judgments continuously. Nevertheless, within the group-first condition, groups ($M = 60.0\%$, $SD = 13.9\%$) were still significantly more accurate overall than individuals [$M = 57.1\%$, $SD = 15.4\%$; paired $t(62) = 2.01, P = 0.048$].

Significant order effects emerged when examining accuracy in detecting truths vs. lies. When detecting lies, a 2(judgment: group vs. individual) \times 2(order: individual first vs. group first) mixed-model ANOVA yielded only a significant main effect for order [$F(1, 130) = 4.89, P = 0.03, \eta_p^2 = 0.036$]. Groups were more accurate detecting lies when they made group judgments first ($M = 59.7\%$, $SD = 18.4\%$) than when they made individual judgments first [$M = 51.2\%$, $SD = 21.8\%$; $t(130) = 2.40, P = 0.02, d = 0.42$]. This suggests that the group advantage in lie detection was disabled when individual judgments were made first, consistent with the synergy account and inconsistent with the sorting account.

When detecting truths, a 2(judgment: group vs. individual) \times 2(order: individual first vs. group first) mixed-model ANOVA yielded a significant main effect for order [$F(1, 130) = 4.72, P =$

$0.03, \eta_p^2 = 0.035$]. Groups were less accurate detecting truths when group judgments were first ($M = 62.9\%$, $SD = 22.3\%$) than second [$M = 71.2\%$, $SD = 25.9\%$; $t(130) = 1.94, P = 0.05, d = 0.34$].

This effect of order suggests that whichever judgment participants made first affected the judgment participants made second, consistent with the synergy account of the group advantage in lie detection and inconsistent with the sorting account. Groups do not seem better able to identify the best lie detector in their midst. Instead, they seem to create better lie detectors through their discussion. Presumably, this occurs because new and useful information is presented in discussion that helps to form a more accurate judgment.

Additional evidence consistent with the synergy mechanism comes from “individual-only” control condition, whose accuracy rates (Table 4) closely mirror those of participants who made individual judgments first and are clearly discrepant from the accuracy rates of participants who made group judgments first. Group discussion is not altering accuracy rates when an individual judgment has been rendered. Instead, it is altering the individual judgments group members are making.

Experiment 4 provides two additional pieces of evidence consistent with the synergy mechanism and inconsistent with the sorting mechanism. First, the group and individual judgments are highly correlated, both for participants who made their individual judgment first ($r = 0.60, P < 0.05$) and for those who made their group judgment first ($r = 0.74, P < 0.05$). Participants' second judgments aligned with first judgments, and their patterns of accuracy also reveal this influence. Individuals were better detecting lies after they formed their opinion in a group, but were no better in a group after they had formed their opinion as an individual. This suggests that whatever participants did first solidified their opinion and it carried over to their second evaluations.

Second, the synergy account suggests that group discussion creates more accurate judgments, rather than simply selecting the best judgment within the group. Therefore, individuals should be more likely to report a judgment consistent with the group opinion when group discussion comes first than when it comes second. Indeed, a higher percentage of individual judgments were consistent with the group judgment when group discussion came first ($M = 86.7\%$, $SD = 13.9\%$) than when it came second [$M = 80.7\%$, $SD = 12.0\%$; $t(130) = 2.63, P < 0.01, d = 0.46$]. Consistent with existing research on social influence (20), group discussion created opinions, in this case creating opinions that were better able to detect a liar.

Finally, we examine the potential role of response bias in these results. Because participants continuously cycled between making individual judgments and group judgments, we again caution that differences in response biases are not likely to be the function of differences between group discussion and individual

Table 4. Group and individual accuracy in lie detection under changing judgment order (experiment 4)

Condition	Accuracy, %			
	Overall	Truths	Lies	“Truth” guesses
Group first condition				
Group	60.0 _a	62.9 _a	59.7 _a	50.5 _d
Individual	57.1 _a	60.3 _a	57.3 _a	50.5 _d
Individual first condition				
Individual	58.1 _a	71.2 _b	51.2 _c	57.2 _e
Group	59.1 _a	69.0 _b	51.4 _c	58.3 _e
Individual only	57.3 _a	68.5 _b	49.3 _c	58.2 _e

All means are reported as percentages. Means that do not share the same subscript within columns and rows differ at $P < 0.05$. Judgment order is a within-subjects manipulation.

judgments per se, but the function of the order in which participants made these judgments.

As shown in Table 4, participants who rendered their individual judgments first were more likely to guess “truth” in their individual and group judgments ($M = 57.2\%$, $SD = 14.3\%$) than were participants who rendered their group judgment first [$M = 51.3\%$, $SD = 15.4\%$; $t(130) = 2.30$, $P = 0.02$]. To test whether this difference in response bias was related to differences in accuracy, we conducted a 2(judgment: group vs. individual) \times 2(order: individual first vs. group first) mixed-model ANOVA on overall accuracy with individuals’ frequency of “truth” guesses as a covariate. The results revealed that the frequency of “truth” guesses was not statistically significant [$F(1, 129) = 0.06$, $P = 0.81$]. In addition, the frequency of “truth” guesses was not correlated with accuracy rates in individuals ($r = 0.02$, $P = 0.78$) or groups ($r = -0.01$, $P = 0.90$). This suggests that differences in response bias did not contribute to the accuracy advantage obtained when group judgment was first.

Discussion

Detecting deception in everyday life is so difficult that extraordinary measures seem required, from lengthy training of human judges to using ever more sophisticated technology in place of human judges. We examined the potential efficacy of a different approach: allowing untrained individuals to detect liars as a group. Our experiments showed a consistent group advantage for detecting small “white” lies in the laboratory as well as high-stakes lies told intentionally for personal gain.

This group advantage in lie detection did not come through the statistical aggregation of individual opinions as often shown in existing research (a wisdom-of-crowds effect), but instead through the process of group discussion. Groups were not simply maximizing the small amounts of accuracy contained among individual members but were instead creating a unique type of accuracy altogether. To see the magnitude of this accuracy gain, we identified the most accurate individuals in experiments 1–3 from each session of three participants in the individual condition. Real groups did not perform significantly worse than the best individual in experiments 1–3 ($ts < 1.46$; Tables 1–3, bottom rows). Of course, these “best individuals” are identified post hoc, benefitting from actual skill but also chance accuracy, therefore providing the highest accuracy rates one could hope to observe. Real groups performed well even against this extremely high bar.

Researchers have made a concerted effort to improve individual lie detection, but have not pursued how much individuals could help each other in detecting lies. Our research therefore leaves many open research questions. What about group discussion, exactly, increases accuracy, particularly in detecting lies compared with individuals? Could group discussion be guided to improve accuracy rates even further? Do larger groups perform even better than smaller groups? Do trained individuals perform even better in a group than untrained individuals? Given the concerted efforts put into training individuals to detect lies more accurately, the productive path for further studying the effectiveness of group lie detection is clear.

Materials and Methods

Experiments 1 and 2. In the first two experiments, participants watched a series of different statements from different speakers and guessed whether each statement was a truth or a lie, either individually or in three-person groups. These two experiments had identical designs. Participants in the individual condition guessed whether the person in the video was telling the truth or lying and then reported a level of confidence in this guess on a scale ranging from 1 (not at all) to 9 (very confident). Participants in the group condition received only one survey form and were asked to arrive at a joint decision through discussion about whether the person in the video was telling the truth or lying, and then to arrive at a joint estimate of confidence (on the same 1–9 scale).

We created the stimuli by video-recording 18 speakers who provided truthful and deceitful answers to 10 questions (*Supporting Information*). These participants were recruited during regular business hours at a research laboratory in downtown Chicago, signed an appropriate waiver, and were compensated \$2 for their time. An experimenter stood behind the video camera, read each question aloud, and then held up a card (visible only to the participant) that read either “tell the truth” or “lie.” Half of the participants were allowed to write down brief notes for themselves before answering each question (“scripted statements”), whereas the other half answered each question immediately (“spontaneous statements”). We thus generated a bank of 180 truthful and deceitful statements.

Experiment 2 ($n = 351$) replicates experiment 1 ($n = 180$) using different statements and nearly doubling the sample size. Participants were all visitors to the Museum of Science and Industry in Chicago who agreed to participate in a quiet laboratory setting in exchange for candy and a small gift from the Museum’s gift store. For experiments 1 and 2, we randomly selected 40 videos from our video bank and created four different sequences of 10 videos each. Each participant, whether in the group or individual condition, watched and judged one sequence—amounting to 10 guesses for each individual and group.

Experiment 3. Experiment 3 ($n = 360$) tested whether the group advantage in lie detection extends to large-stakes and intentional lies. As stimuli, we used video clips containing footage from the last segment the British television game show “Golden Balls,” in which two contestants attempt to deceive each other, and, if successful, win amounts ranging from £6,550 to £66,885. We collected all available video clips (19 in total) showing this critical segment posted on YouTube.com on April 2, 2014 (cf. ref. 21).

In this part of the show, each contestant must choose whether to “split” or to “steal.” If both contestants decide to “split,” each keeps half of the pot. If one chooses to “steal” and one chooses to “split,” the stealer takes the whole pot and leaves the splitter with nothing. If both choose to “steal,” both walk away with nothing. In our video clips (as in nearly all of the actual episodes), all contestants claim that they would “split” and exhort one another to do the same. In reality, some lie and choose to “steal.” This context is therefore suitable for testing detection of high-stakes and intentional deception. Contestants who claim to “split” but choose to “steal” are engaging in intentional deception in attempt to win the entire pot, whereas those who choose to “split” are truth-tellers. We edited the videos to include only contestants’ conversations and exclude the revelation of their “split” or “steal” choices.

We obtained all available videos we could find at the time showing the desired segment from the show, and excluded one video that was dramatically more popular than the others and we believed could therefore have been known to our participants. We randomly divided the resulting 18 videos into three sequences of six videos each. Because each video contained two contestants, participants made two truth/lie guesses per video. This created 12 guesses for each individual and group in experiment 3.

As in previous experiments, participants in the group condition made a collective judgment after discussion whereas participants in the individual condition made judgments alone without discussion. Participants first watched one video clip as a demonstration (the same clip for all three sequences). Then, each participant saw and evaluated only one of the three six-video sequences. Before watching each clip, participants were informed of the amount contestants were playing for. At the end of the experiment, we allowed participants to watch the contestants’ actual choices to satisfy their curiosity.

Experiment 4. Participants were visitors to the Chicago Museum of Science and Industry ($n = 183$; 54.6% women) who participated in exchange for candy and a small gift. Participants again completed the experiment in groups of three in a procedure similar to experiments 1–3. We again selected 40 videos from our larger bank of stimuli and created four sequences of 10 videos, with approximately an equal number of truths and lies within each sequence. Each group was randomly assigned to watch only one of these sequences.

Participants in the individual-first condition watched one target video, decided individually whether the person was telling the truth or lying, and then discussed it as a group and made a group decision. These participants repeated this process for each of the 10 videos. Participants in the group-first condition followed the same procedure, except they discussed the video and made a group decision first, and then made their individual decision second. Participants in the control condition (individual-only condition) simply made individual judgments about the videos, to serve as a baseline for comparison. When finished evaluating all 10 videos, all participants were debriefed and dismissed.

