Personality and Individual Differences 231 (2024) 112823

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Genetically-diverse crowds are wiser

Meir Barneron [a,*], Ilan Yaniv [b], Lior Abramson [b,c], Ariel Knafo-Noam [b]

[a] *Seymour Fox School of Education, Hebrew University of Jerusalem, Israel*
[b] *Department of Psychology, Hebrew University of Jerusalem, Israel*
[c] *Department of Psychology, Columbia University, New York, United States*

ARTICLE INFO

Keywords:
Judgment accuracy
Decision-making
Combining estimates
Wisdom of crowds
Genetic diversity
Independence of opinion
Twin studies

ABSTRACT

A fundamental question in the social sciences is how collectives of individuals form intelligent judgments. This article tests the hypothesis that genetically-diverse groups make better collective judgments than genetically more homogenous groups. Two studies were conducted (a total of $N = 602$ participants) in which sets of twins (both monozygotic and dizygotic) were required to perform the task of making numerical judgments. The accuracy of the judgments made by pairs of participants—who were either co-twins (i.e., genetically-related) or were not related—was then compared. The results indicate that the judgments made by unrelated pairs were more accurate than those of the genetically-related twins. Critically, however, this superior performance was found only among monozygotic twins, evidencing the role of genetic relatedness in collective judgment. This research provides the first empirical demonstration of the benefit of genetic diversity for collective judgments, shedding light on the origins of the 'wisdom of crowds' phenomenon.

1. Introduction

A fundamental question in the social sciences concerns the quality of judgments and decisions made by collectives versus individuals. The first to document the benefits of collective judgment was Francis Galton (1907), establishing what subsequently came to be termed the "wisdom of crowds" (Surowiecki, 2005). Since then, an impressive seam of research has shown how deliberately combining individual opinions can improve judgments and decisions (Armstrong, 2001; Clemen, 1989; Yaniv, 2004a; see Larrick & Soll, 2006, for a historical review). As a rule, the group becomes more accurate, the more it increases in size, albeit with diminishing gains. Appreciable gains are observed when combining the judgments of even a small number of individuals. For example, in one study, averaging one individual's estimate with the estimates of two other individuals decreased error by about 30 % (Yaniv & Milyavsky, 2007).

Importantly, collective wisdom is observed when the combined judgments are arrived-at independently of one another (Frey & Van de Rijt, 2021; Hogarth, 1978; Yaniv et al., 2009). A standard method for securing independent opinions is to obtain them from diverse sources, from individuals who vary in background, education, knowledge, and demography. In the present study, we examined the proposition that genetic diversity might also contribute to the accuracy derived from combining judgments. Using a twin-study design, we tested the hypothesis that genetically-diverse pairs of judges perform better than genetically-related pairs.

2. Theoretical background

We focused on situations in which *quantitative* estimates, judgments, opinions, or forecasts are elicited from a group of judges under conditions of uncertainty. In this kind of study, judgments are elicited individually, with no interaction among the judges, thereby avoiding the possibility of direct social influence (Becker et al., 2017; Klein & Epley, 2015). The collective judgment of the group is then obtained by averaging the individual judgments.

Under such conditions, the collective has been shown to outperform the average individual in terms of judgment accuracy. That is, the absolute error of the collective judgment is smaller than the average of the absolute individual errors. Such outperformance has been documented in diverse domains, including: economic and business forecasting (Kelley & Tetlock, 2013; Wolfers & Zitzewitz, 2004), crowdsourcing (Sunstein, 2006), medical diagnoses (Goldberg, 1965; Kurvers et al., 2016; Kurvers et al., 2021), meteorological forecasting (Staël von Holstein, 1971), hedonic forecasting (Müller-Trede et al., 2017), geopolitical forecasting (Ferreiro et al., 2023), performance evaluation (Barneron

---

* Corresponding author.
E-mail address: meir.barneron@mail.huji.ac.il (M. Barneron).

https://doi.org/10.1016/j.paid.2024.112823
Received 5 May 2024; Received in revised form 20 July 2024; Accepted 22 July 2024

et al., 2019), and general knowledge (Yaniv & Choshen-Hillel, 2012).

Theoretical studies have identified the conditions under which combining individual judgments improves accuracy (Davis-Stober et al., 2014; Einhorn et al., 1977; Müller-Trede et al., 2017; Wallsten & Diederich, 2001). Generally speaking, a judgmental estimate can be expressed as the sum of three components: the truth (the true value of the estimated quantity), a common bias (a systematic tendency to underestimate or overestimate the true value), and a random error (random fluctuations in the judge's estimation; see Yaniv, 2004a). Combining estimates increases accuracy by reducing both the random and systematic error components. To the extent that different individuals make independent errors (Soll, 1999), then these errors cancel each other out when the judgments are combined, thereby decreasing the error of the combined judgment compared with the average of the individual errors (Larrick & Soll, 2006).

Why do people make different kinds of judgment errors? For one thing, they might draw on different sources of information, knowledge, and experience. Moreover, individuals could rely on different assumptions, models, and inference methods. This diversity of approaches increases independence among the judgments and, as a result, the accuracy of collective judgment (Aminpour et al., 2020; Clemen et al., 2000; Makridakis & Winkler, 1983; Winkler & Clemen, 2004).

Diversity of opinion can arise from a wide range of individual differences. For example, people vary in their cognitive styles in a manner that could influence their reasoning about uncertainties (Nisbett et al., 2001). Differences in personality traits and abilities have also been related to cognitive systematic differences (Baron & Ritov, 2004; Boyce et al., 2016; Eickhoff, 2018; Furnham et al., 2012; Levin et al., 2002; Liberali et al., 2012; Shiloh et al., 2002; Stanovich & West, 1998).

In this study, we focused on a source of diversity that, to the best of our knowledge, has not been studied in this context: *genetic* diversity. Our working assumption was that, because genetic diversity is reflected in individual cognitive and behavioral differences (Plomin et al., 2016; Polderman et al., 2015), genetically-diverse individuals would produce more independent opinions. If this were the case, then the combined opinions of a group of genetically-diverse individuals would contribute to judgment accuracy, when accuracy is compared to an appropriately chosen benchmark.

Genetic variance is reflected in individual differences across a wide variety of phenotypes (Johnson et al., 2009) and most, if not all, psychological traits show significant and substantial genetic influence (Plomin et al., 2016). This generalization is so robust that it has been suggested as the first law of behavioral genetics (Turkheimer, 2000). A meta-analysis based on 50 years of research findings on the heritability of human traits estimates that about half of the variance in complex traits could be attributed to genetic variance (Polderman et al., 2015). Of special relevance here is the notion that genetic variance accounts for individual differences in personality traits and cognitive abilities, which influence judgment and decision-making (Cesarini et al., 2012; Cronqvist & Siegel, 2014; Ebstein et al., 2010; Polderman et al., 2015; Simonson & Sela, 2010; Zakharin & Bates, 2023).

Taken as a whole, the foregoing findings from the literature set the stage for our investigation of whether and how genetic diversity might contribute to the accuracy of collective judgments. Specifically, we tested the hypothesis that groups (in fact, pairs) make more accurate collective judgments if they are genetically-diverse than if they are genetically-related.

3. Method

3.1. Research transparency statement

We conducted two separate studies. Study 1 and Study 2 employed identical materials, procedure, design, data analysis, and statistical inference, and differed only in terms of sample characteristics (see below). Study 2 was pre-registered at: https://aspredicted.org/1RS_DZV. The data, the code, and the materials for the study can be provided by the corresponding author on request. The two studies received the ethical approvement of the Faculty of Social Sciences at the Hebrew University of Jerusalem. We reported all measures, manipulations, and participant exclusions.

3.2. Rationale for the present methodology

We tested the effect of genetic diversity on judgment accuracy using a sample of twins, controlling for their genetic relatedness. Two twins are more similar to one another compared to two individuals who are not twins, due to either their shared genes, shared environment, or both (Martin et al., 1997). In our studies, we asked a sample of twins to make (individually and separately) numerical estimates of quantities. From this pool of estimates, we then created artificial pairs of twins, by pairing the estimates made by two individual twins. Next, we averaged the estimates of the two individuals in the pairs and assessed the accuracy of the combined estimates. This represents the accuracy of the collective judgment of a pair.

Importantly, the pairs we created consisted of either (i) natural co-twins (hereafter: “related” pairs) or (ii) twins who were not co-twins (hereafter: “unrelated” pairs; see example below). We then compared the accuracy of collective judgments made by related versus unrelated pairs.

Note that both types of pairs involved the very same participants. The only difference between related and unrelated resided in the way the pairs were (re)arranged by us for analytical purposes, once we had collected their individual estimates. The related pairs comprised two individuals who were each other's siblings, whereas the unrelated pairs involved two individuals organized in such a way that they were *not* co-twins of the other person in the pairing. As the unrelated pairs consisted of two individuals who were not co-twins, they were more diverse than the related pairs. Thus, if it was found that the former made more accurate collective judgments than the latter, this would corroborate the hypothesis that diversity (genetic, environmental, or both) contributes to the accuracy of collective judgments.

To test the role of genetic diversity, we made a distinction between monozygotic and dizygotic twins (so-called identical and non-identical twins, respectively). One of the basic working assumptions in twin research is the following: the shared environmental factors that lead to the recognized greater similarities between genetically-related twins (compared to pairs of non-twins) exert a marked effect of similar extent on both (a) monozygotic twins and (b) same-sex dizygotic twins (Martin et al., 1997; Willoughby et al., 2023). In contrast, genetic influences make monozygotic twins more similar to one another compared to dizygotic twins because the former share virtually 100 % of their genetic variance, whereas dizygotic twins share, on average, 50 % of the genetic variance. Therefore, we compared the accuracy of pairs of related and unrelated twins, among monozygotic and same-sex, dizygotic twins. In statistical terms, we explored the interaction between relatedness and zygosity. To the extent that genetic diversity contributes to collective accuracy, we expected the effect of relatedness (unrelated versus related pairs) to be higher for monozygotic twins than for same-sex dizygotic twins. Finally, our sample included different-sex dizygotic twins as a third group, as gender is also a source of diversity.

3.3. Comparing related and unrelated pairs

Let us now explain how we compared the accuracy of related versus unrelated pairs. Consider a sample comprising a number (T) of twins. Therefore, the sample involved $T/2$ pairs of related twins. For example, consider the following $T = 6$ twins: $1_a, 1_b, 2_a, 2_b, 3_a, 3_b$. The numbers (1, 2, … $T/2$) are indices for the pairs; and the letters (a and b) are indices for the co-twins within a pair.

From those six twins, we can create a total of 15 different artificial pairs, three of which involve related individuals (co-twins); and the remaining twelve pairs involve unrelated individuals (see Table 1). Generally speaking, the number of all different pairs that could be generated from T twins is $T^*(T-1)/2$. This set of all possible pairs includes $T/2$ pairs of related individuals (the actual twins) and $T^*(T-1)/2-T/2$ pairs of unrelated individuals (i.e., not co-twins). In our studies, for each pair, having computed the accuracy of the combined judgment, we then compared the average accuracies of the related ($T/2$) and unrelated ($T^*(T-1)/2-T/2$) pairs.

3.4. Participants

Our participants were part of the larger Longitudinal Israeli Study of Twins (LIST, Vertsberger et al., 2019), a longitudinal research project on genetic and environmental contributions to social development among twins. In each study, participants were invited via an electronic link to take part in what we termed a "guessing contest", and were promised an individual prize of $55 for the best performance. As our analyses focused on pairs, we included only pairs in which both twins completed the procedure. Zygosity for same-sex twins was assessed in previous waves of the twin project, via a DNA test or parental questionnaires on twin similarity (Goldsmith, 1991).

In total, across Studies 1 and 2, we recruited 602 participants. The sample in Study 1 comprised 316 participants (i.e., 158 pairs of twins): 72 monozygotic twins, 140 same-sex dizygotic twins, and 104 different-sex dizygotic twins (mean age = 11.65 years, 50.6 % females). The sample in Study 2 comprised 286 participants (i.e., 143 pairs of twins): 66 monozygotic twins, 112 same-sex dizygotic twins, and 108 different-sex dizygotic twins (mean age = 16.5 years; 53.1 % females). Twenty-seven pairs of twins who took part in Study 2 had previously taken part in Study 1. Results were stable even after excluding those participants from Study 2.

3.5. Materials and procedure

Each study consisted of nine trials, in which participants responded to the prescribed task individually and remotely, via computer. The task consisted of estimating the number of candies in a jar, shown in a photograph on-screen. To familiarize the participants with the research materials, they were first shown photographs of two sample jars, along with the correct number of candies in each (see Yaniv & Choshen-Hillel, 2012, for a similar procedure). All participants were shown the same two jars, which were not re-shown subsequently during the real task. The number of candies in the jars ranged between 120 and 607 and the order of the photographs was randomized for each participant. To ensure that estimates would be reasonable, participants were told to keep within a range of 90–750 (for similar practice, see, for example, Yaniv & Choshen-Hillel, 2012, or Yaniv, 2004b). Anyone providing estimates outside this range was asked to estimate afresh. Other than that, no feedback was given to the participants on their performance or on the correct number of candies. Once the participants' individual estimates were collected, the trial was complete.

Table 1
All possible pair combinations of six twins, $1_a, 1_b, 2_a, 2_b, 3_a, 3_b$.

Pairs	Type	Pairs	Type	Pairs	Type
$\{1_a; 1_b\}$	**Related**	$\{1_b; 2_a\}$	Unrelated	$\{2_a; 3_a\}$	Unrelated
$\{1_a; 2_a\}$	Unrelated	$\{1_b; 2_b\}$	Unrelated	$\{2_a; 3_b\}$	Unrelated
$\{1_a; 2_b\}$	Unrelated	$\{1_b; 3_a\}$	Unrelated	$\{2_b; 3_a\}$	Unrelated
$\{1_a; 3_a\}$	Unrelated	$\{1_b; 3_b\}$	Unrelated	$\{2_b; 3_b\}$	Unrelated
$\{1_a; 3_b\}$	Unrelated	**$\{2_a; 2_b\}$**	**Related**	**$\{3_a; 3_b\}$**	**Related**

4. Analyses

4.1. Dependent variables

Our main dependent variable was the Mean of Standardized Absolute Error (MSAE), which was computed as follows. First, for each pair (related or unrelated), we averaged the two estimates of the judges, by jar. The absolute error of the collective judgment was defined as its distance from the true number of candies in the jar. Next, the absolute errors were standardized, by jar. This enabled us to combine the MSAEs for each pair, by averaging the standardized absolute errors across the jars. Finally, the MSAEs of the pairs were standardized again, across the pairs, so that we could interpret the results in a standardized manner. Note that lower MSAEs reflect higher accuracy.

The Supplementary Materials report the results in terms of (non-standardized) Mean Absolute Errors (MAE). Although MAEs have an intuitive meaning (they can be interpreted in terms of numbers of candies in jars), in this case, they could have been biased by the variance between the jars as they contained different numbers of candies. Therefore, all analyses reported here were performed on MSAEs.

Our analyses compared the accuracy of related versus unrelated pairs. We performed three separate analyses: (i) among the monozygotic twins; (ii) among the same-sex dizygotic twins; and (iii) among the different-sex dizygotic twins.

4.2. Statistical inference

Our computation of combined judgment involved pairing each participant with every other participant, calculating the absolute error of the average estimate of each pair. This means that our observations (MSAEs) were not independent from one another. Therefore, we turned to non-parametric permutation tests to assess the statistical significance of our results. This was done by creating an empirical reference distribution, simulating what the difference between related and unrelated pairs would look like under the null hypothesis that they would not perform differently. We then compared the actual, observed, difference between related and unrelated pairs to this distribution.

The construction of the distribution involved 1,000,000 cycles of computation on our data, resulting in a distribution of 1,000,000 random differences. First, all possible pairs of judges were created. Then, in each cycle, each pair was randomly assigned to one of two types: assigned-related or assigned-unrelated. This process mimicked the real data: $T/2$ pairs were assigned to the assigned-related type, and the other ($T^*(T-1)/2-T/2$) pairs to the assigned-unrelated type. This randomized assignment process was subject to only one restriction: that, within a specific simulation, once a pair was assigned to the related type, the two judges could not be assigned again to the related type. This reflects the fact that, in the real observed distribution, the related types involved each judge just once. Next, we computed the MSAE across all pairs in each type, in the same way as the observed MSAEs were computed, and obtained the difference between the mean of the MSAE of the related-assigned and unrelated-assigned pairs. At the end of the process, we obtained 1,000,000 differences between the MSAE of the related-assigned and unrelated-assigned pairs.

In the final step, we compared the real observed difference between the mean error of the related and unrelated pairs, on the one hand, and the distribution of the simulated differences created at *random*, on the other. Our p-value was defined as the percentage of the simulated differences that were above the observed difference. The interpretation of the p-value is analogous to that based on a parametric test (e.g., t-test), except that our p-value was based on an empirical distribution, not a theoretical one (e.g., t-distribution). This entire procedure was pre-registered for Study 2.

Note that, by construction, the number of related pairs is lower than the number of unrelated pairs. In the example shown in Table 1, given six twins, there are three related pairs and 12 unrelated pairs. This difference in the number of pairs is reflected also in the permutation test: $T/2$ pairs were assigned to the related type. The remaining pairs ($T^*(T-1)/2 - T/2$) were assigned to the unrelated type.

5. Results

5.1. Comparing the accuracy of related versus unrelated pairs

Since the two studies yielded similar results, we report them together. Fig. 1 presents the difference between the related and unrelated pairs, for each analysis separately. Positive numbers mean that unrelated pairs performed better than related pairs. (Table 2 of Supplementary Materials reports the results separately for the related and unrelated pairs, by zygosity, in terms of both MSAE and MAE.)

Overall, Studies 1 and 2 presented a similar pattern whereby unrelated pairs performed better than related pairs. However, this difference was observed only among the monozygotic twins. Consider, first, the analysis involving such twins. The differences between the MSAEs were 0.11 in Study 1 ($p = .049$) and 0.11 in Study 2 ($p = .021$). Therefore, in both studies, unrelated pairs performed significantly better than related pairs.

For the dizygotic twins, however, the obtained pattern was different. Here, there was no significant difference between related and unrelated pairs. Among same-sex dizygotic twins, the MSAE difference was 0.06 in Study 1 ($p = .092$) and $-$ 0.09 in Study 2 ($p = .958$). As for the dizygotic different-sex twins, the MSAE difference was -0.02 in Study 1 ($p = .676$) and $-$ 0.02 in Study 2 ($p = .671$).

In the above analyses, the related pairs naturally involved judges of an identical sex for both the monozygotic twins and the dizygotic same-sex twins. The unrelated pairs, however, comprised both same-sex and different-sex judges. This difference could have confounded the previous results. Therefore, we repeated the above analyses, this time involving the related and unrelated pairs who were identical in their gender composition. Among the monozygotic twins and the same-sex dizygotic twins, we included only pairs of judges who were of the same sex. Among the different-sex dizygotic twins, we included only pairs of judges who were of different sexes. Here, too, the statistical inference was based on the permutation test. Fig. 2 presents the difference between the related and unrelated pairs for each analysis, separately. (Table 3 of Supplementary Materials reports the results separately for the related and unrelated pairs, by zygosity, in terms of both MSAE and MAE.)

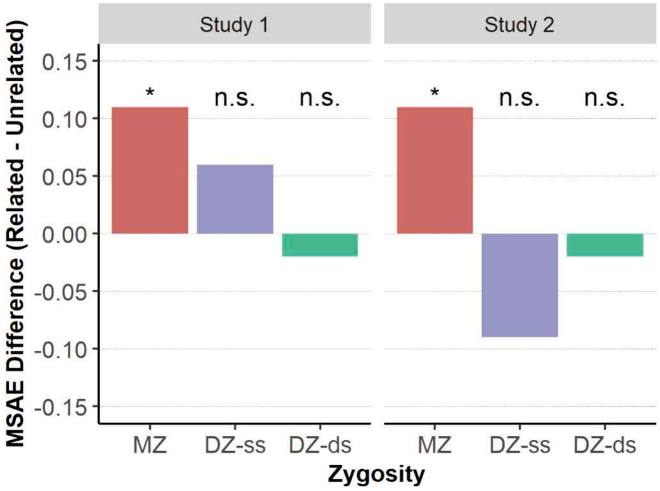

Fig. 1. Differences between related and unrelated pairs, by zygosity.
Key: MZ = monozygotic; DZ-ss = dizygotic same-sex; DZ-ds = dizygotic different-sex; * indicates that the MSAE difference between related and unrelated twins was statistically significant at $p < .05$.

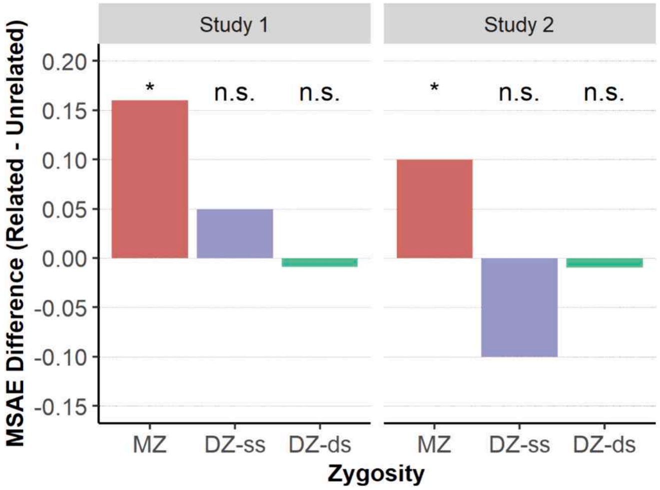

Fig. 2. Differences between unrelated and related pairs, by zygosity, controlling for gender composition.
Key: MZ = monozygotic; DZ-ss = dizygotic same-sex; DZ-ds = dizygotic different-sex; * indicates that the MSAE difference between related and unrelated twins was statistically significant at $p < .05$.

Essentially, the analyses revealed the same pattern as before: it was only among monozygotic twins that unrelated pairs performed better than related pairs. Specifically, for the monozygotic twins, the differences between the MSAE were 0.16 in Study 1 ($p = .023$) and 0.10 in Study 2 ($p = .024$). In the case of the dizygotic twins, among the dizygotic same-sex twins, the MSAE differences were 0.05 in Study 1 ($p = .115$) and $-$ 0.10 in Study 2 ($p = .971$). Among the dizygotic different-sex pairings, the MSAE differences were $-$ 0.009 in Study 1 ($p = .646$) and $-$ 0.01 in Study 2 ($p = .705$).

5.2. The interaction between relatedness and zygosity

The above results revealed a difference between related and unrelated pairs but only for the monozygotic twins. Our final analysis compared the effects between the monozygotic versus the same-sex dizygotic twins. In other words, this analysis assessed the statistical significance of the interaction between relatedness and zygosity; and it was performed while controlling for gender composition.

Here, too, the statistical significance of these results was assessed using permutation tests. We randomly assigned the pairs of monozygotic and dizygotic same-sex twins into two sets: “assigned monozygotic” and “assigned dizygotic same-sex”. For each of these sets, we computed the difference, in terms of MSAE, between the (real) related and unrelated pairs. Therefore, the procedure randomly changed the identity of the pair's zygosity but retained the identity of the type of pairs (related or unrelated). Next, we computed the difference between the effects obtained in the assigned monozygotic and assigned dizygotic same-sex sets. We repeated this procedure 1,000,000 times, creating a distribution of effects obtained under the null hypothesis that the effect of relatedness would not differ between monozygotic and dizygotic twins. In the final step, the p-value was obtained by computing the percentage of random differences that exceeded the observed one.

In Study 1, the difference between related and unrelated pairs was 0.16 for the monozygotic twins and 0.05 for the dizygotic same-sex twins (see Fig. 2). In Study 2, the difference between related and unrelated pairs was 0.10 for the monozygotic twins and $-$ 0.10 for the dizygotic same-sex twins. Therefore, in both studies, the effect of relatedness was larger for the monozygotic than for the dizygotic same-sex twins. The permutations tests revealed a p-value of 0.141 for Study 1, and 0.009 in the pre-registered Study 2.

What might account for our findings? To answer this question, we devlopped three indices tracking the underlying mechanism at play.[1] The first index captured the diversity in the pair estimates. We computed the absolute distance between the estimates made by the twins in each pair. We hypothesized that, among the monozygotic twins, where the unrelated pairs performed better than the related pairs, the former would also exhibit higher absolute differences between the twins' estimates. The second index captured the dependency between the twins' estimates. We computed the Pearson's correlation between the estimates made by the two twins (all estimates first being standardized by jar). We hypothesized that, among the monozygotic twins, the correlations between the estimates of the unrelated pairs would be lower than the correlations between the estimates made by related pairs. Our third index was based on the idea that one condition for obtaining improved collective judgments is that the estimates should "bracket" the truth (Larrick et al., 2012). Following this idea, we computed, for each pair, the percentage of times that the truth fell between the estimates of the two twins. We hypothesized that the estimates made by unrelated twins would bracket the truth more often.

Table 4 of the Supplementary Materials presents the mean of the absolute distance, the mean of correlations, and the percentage of estimates bracketing the truth, for related versus unrelated pairs, separately for monozygotic and same-sex dizygotic twins. The results revealed a coherent pattern: among the monozygotic twins, where unrelated pairs performed significantly better than the related pairs, the former exhibited higher mean absolute distances and lower mean correlations, and bracketed the truth more often.

6. General discussion

The goal of this research was to investigate the influence of genetic diversity on the accuracy of collective judgments. We tested the hypothesis that pairs of judges who differ from one another genetically should produce more accurate collective judgments than pairs who present genetic variation. In our study, diversity was captured by comparing pairs of judges who were either genetically-related or unrelated. Differences in levels of genetic diversity were captured by comparing related and unrelated pairs *separately* for monozygotic versus dizygotic twins.

We conducted two studies, the second of which was pre-registered, in which we collected data from a sample of $N = 602$ twins. We found that, on average, the collective judgments of the unrelated pairs of twins were more accurate than those of the related pairs. Critically, however, we identified the presence of an interaction, whereby the effect of relatedness was greater among the monozygotic than the dizygotic twins, which points to the (negative) influence of genetic relatedness on the accuracy of collective judgment. Further analyses revealed that, among the monozygotic twins, where the unrelated pairs performed significantly better than the related pairs, the former made more diverse and less dependent estimates, bracketing the truth more often.

It should be emphasized that, although the present research investigates genetic factors, it is not about the heritability of the wisdom-of-crowds phenomenon. Rather, it is about the critical role of *genetic* diversity in obtaining accurate collective judgments. In studies on this issue, diversity is typically obtained by focusing on factors such as background, education, knowledge, or demography. Our findings suggest that genetic diversity is also a valid factor to take into account, with relatedness negatively affecting the accuracy of combined judgments.

To our knowledge, the present research provides the first empirical demonstration of the effects of genetic diversity on accuracy gains in collective judgments. These findings shed light on the origins of the wisdom-of-crowds phenomenon, as they show how diversity in people's genetic heritage affects the quality of collective judgments.

6.1. Leveraging diversity in aggregation

Research testing the efficient use of opinions in forming collective judgments has suggested two main strategies. One strategy involves the identification of the best judges and combining their judgments (Budescu & Chen, 2014; Mannes et al., 2014); and the other relies on the idea that the accuracy gains derived from combining the estimates decrease as a function of the interdependence among the judges (Hogarth, 1978). This second strategy, therefore, involves the identification of the judges, among those available, who are least correlated. One thus needs to analyze the inter-correlations among the judges (based on past performance, for instance), and then combine the estimates of the judges who are less interdependent. Indeed, smaller panels of less interdependent judges may outperform larger panels of judges who are more interdependent. The reason for this is that the estimates of the less interdependent judges provide more information (Larrick & Soll, 2006; Yaniv et al., 2009).

A related method is to use diversity as a proxy for independence. For example, Shi et al. (2019) found that, compared with homogeneous groups of editors, ideologically-diverse groups of editors produced higher-quality collections of articles (see also Page, 2008). The present study provides further evidence in support of the importance of diversity in making collective judgments.

To the extent that genetic diversity does contribute to the accuracy of collective judgments, we might expect unrelated dizygotic pairs of twins to produce more accurate judgments than related dizygotic pairs, as they share 50 % of their genes. This, however, was not obtained in our studies, as relatedness did not decrease accuracy among dizygotic twins. We suggest that, at least for the type of judgments we analyzed, dizygotic twins are sufficiently diverse for their relatedness not to negatively impact the quality of their collective judgments. In contrast, when pairs are identical in terms of their genetics (among monozygotic twins), the collective judgment is impaired.

6.2. Genetically-diverse crowds from an evolutionary perspective

The idea that genetic variability improves collective judgments may provide a clue as to the evolution of individual differences in cognitive processes. One intriguing question is how individual differences in psychological traits have been preserved through evolution. According to one classic theory, individual variability in traits should be transmitted to the next generation because different phenotypes present an evolutionary advantage in different periods, thereby perpetuating the genetic variability accounting for the differences (Nettle, 2006). A different approach, supported by the current findings, corroborates the notion that evolution selects people not only according to how adapted they are individually but (also) according to how adaptive they are collectively (Wilson et al., 2008). As genetic diversity plays a role in making collective judgments, it is possible that genetic diversity was favored through evolution, since it produces individual differences adaptive at the group level.

6.3. Concluding remarks

Improving judgment is fundamental to better decision-making, which is a core concern in domains such as medicine, economics, politics, and law. There are theoretical and empirical bases for the contention that combining diverse sources could improve judgment accuracy. Generally speaking, today's culture highlights diversity and pays tribute to its virtues. The present research advances beyond the previous literature by showing that diversity in people's unique signature—that is, their genetic heritage—could also enhance the quality of collective judgments.

---

[1] These analyses were not pre-registered as they were proposed by one of the reviewers of this article.

CRediT authorship contribution statement

Meir Barneron: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ilan Yaniv:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Lior Abramson:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Investigation. **Ariel Knafo-Noam:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

We have no conflicts of interest to disclose.

Data availability

Data will be made available on request.

Acknowledgments

We would like to thank the Israeli National Institute for Testing and Evaluation for its support during the project. We express our gratitude to the Living Lab in Memory of Noam Knafo at the Jerusalem Bloomfield Science Museum, for allowing us to use their platforms for piloting our measures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.paid.2024.112823.

References

Aminpour, P., Gray, S. A., Jetter, A. J., Introne, J. E., Singer, A., & Arlinghaus, R. (2020). Wisdom of stakeholder crowds in complex social–ecological systems. *Nature Sustainability, 3*(3), 191–199.

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. New York: Springer.

Barneron, M., Allalouf, A., & Yaniv, I. (2019). Rate it again: Using the wisdom of many to improve performance evaluations. *Journal of Behavioral Decision Making, 32*(4), 485–492.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74–85.

Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences, 114*(26), E5070–E5076.

Boyce, C. J., Wood, A. M., & Ferguson, E. (2016). Individual differences in loss aversion: Conscientiousness predicts how life satisfaction responds to losses versus gains in income. *Personality and Social Psychology Bulletin, 42*(4), 471–484.

Budescu, D. V., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. *Management Science, 61*(2), 267–280.

Cesarini, D., Johannesson, M., Magnusson, P. K., & Wallace, B. (2012). The behavioral genetics of behavioral anomalies. *Management Science, 58*(1), 21–34.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559–583.

Clemen, R. T., Fischer, G. W., & Winkler, R. L. (2000). Assessing dependence: Some experimental results. *Management Science, 46*(8), 1100–1115.

Cronqvist, H., & Siegel, S. (2014). The genetics of investment biases. *Journal of Financial Economics, 113*(2), 215–234.

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision, 1*(2), 79–101.

Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S., & Knafo, A. (2010). Genetics of human social behavior. *Neuron, 65*(6), 831–844.

Eickhoff, C. (2018). Cognitive biases in crowdsourcing. In *WSDM '18: Proceedings of the eleventh ACM international conference on web search and data mining* (pp. 162–170). Association for Computing Machinery.

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin, 84*(1), 158–172.

Ferreiro, D. N., Deroy, O., & Bahrami, B. (2023). Compromising improves forecasting. *Royal Society Open Science, 10*(5), 216–221.

Frey, V., & Van de Rijt, A. (2021). Social influence undermines the wisdom of the crowd in sequential decision making. *Management Science, 67*(7), 4273–4286.

Furnham, A., Boo, H. C., & McClelland, A. (2012). Individual differences and the susceptibility to the influence of anchoring cues. *Journal of Individual Differences, 33*(2), 89–93.

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature, 75*(7), 450–451.

Goldberg, L. R. (1965). Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs: General and Applied, 79*(9), 1–28.

Goldsmith, H. (1991). A zygosity questionnaire for young twins: A research note. *Behavior Genetics, 21*, 257–269.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance, 21*(1), 40–46.

Johnson, W., Turkheimer, E., Gottesman, I. I., & Bouchard, T. J., Jr. (2009). Beyond heritability: Twin studies in behavioral research. *Current Directions in Psychological Science, 18*(4), 217–220.

Kelley, E. K., & Tetlock, P. C. (2013). How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance, 68*(3), 1229–1265.

Klein, N., & Epley, N. (2015). Group discussion improves lie detection. *Proceedings of the National Academy of Sciences, 112*(24), 7460–7465. https://doi.org/10.1073/pnas.1504048112

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., … Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences, 113*(31), 8777–8782. https://doi.org/10.1073/pnas.1601827113

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., & Wolf, M. (2021). Pooling decisions decreases variation in response bias and accuracy. *iScience, 24*(7), 1–19.

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In *Social judgment and decision making* (pp. 227–242). Psychology Press.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*(1), 111–127.

Levin, I. P., Gaeth, G. J., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes, 88*(1), 411–429.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making, 25*(4), 361–381.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science, 29*(9), 987–996.

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology, 107*(2), 276–299.

Martin, N., Boomsma, D., & Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics, 17*(4), 387–392.

Müller-Trede, J., Choshen-Hillel, S., Barneron, M., & Yaniv, I. (2017). The wisdom of crowds in matters of taste. *Management Science, 64*(4), 1779–1803.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist, 61*(6), 622–631.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291–310.

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11*(1), 3–23.

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Fifty years of twin studies: A meta-analysis of the heritability of human traits. *Nature Genetics, 47*(7), 702–709.

Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature Human Behaviour, 3*(4), 329–336.

Shiloh, S., Salton, E., & Sharabi, D. (2002). Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. *Personality and Individual Differences, 32*(3), 415–429.

Simonson, I., & Sela, A. (2010). On the heritability of consumer decision making: An exploratory approach for studying genetic effects on judgment and choice. *Journal of Consumer Research, 37*(6), 951–966.

Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology, 38*(2), 317–346.

Staël von Holstein, C. A. S. (1971). An experiment in probabilistic weather forecasting. *Journal of Applied Meteorology, 10*(4), 635–645.

Stanovich, K. E., & West, R. F. (1998). Individual differences in framing and conjunction effects. *Thinking & Reasoning, 4*(4), 289–317.

Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. Oxford: Oxford University Press.

Surowiecki, J. (2005). *The wisdom of crowds*. New York: Doubleday.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science, 9*(5), 160–164.

Vertsberger, D., Abramson, L., & Knafo-Noam, A. (2019). The Longitudinal Israeli Study of Twins (LIST) reaches adolescence: Genetic and environmental pathways to social, personality, and moral development. *Twin Research and Human Genetics, 22*(6), 567–571.

Wallsten, T. S., & Diederich, A. (2001). Understanding pooled subjective probability estimates. *Mathematical Social Sciences, 41*(1), 1–18.

Willoughby, E. A., Polderman, T. J., & Boutwell, B. B. (2023). Behavioural genetics methods. *Nature Reviews Methods Primers, 3*(1), 10.

Wilson, D. S., Van Vugt, M., & O'Gorman, R. (2008). Multilevel selection theory and major evolutionary transitions: Implications for psychological science. *Current Directions in Psychological Science, 17*(1), 6–9.

Winkler, R. L., & Clemen, R. T. (2004). Multiple experts vs. multiple methods: Combining correlation assessments. *Decision Analysis, 1*(3), 167–176.

Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives, 18*(2), 107–126.

Yaniv, I. (2004a). The benefit of additional opinions. *Current Directions in Psychological Science, 13*(2), 75–78.

Yaniv, I. (2004b). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*(1), 1–13.

Yaniv, I., & Choshen-Hillel, S. (2012). Exploiting the wisdom of others to make better decisions: Suspending judgment reduces egocentrism and increases accuracy. *Journal of Behavioral Decision Making, 25*(5), 427–434.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 558–563.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes, 103*(1), 104–120.

Zakharin, M., & Bates, T. C. (2023). Testing heritability of moral foundations: Common pathway models support strong heritability for the five moral foundations. *European Journal of Personality, 37*(4), 485–497.

Supplementary Materials

Table 2 presents the Mean of Standardized Absolute Errors (MSAE) and Mean Absolute Errors (MAE), separately for related vs. unrelated pairs, by zygosity. Table 3 presents the means of MSAEs and MAEs for the related vs. unrelated pairs, by zygosity, controlling for gender composition. The MSAE values shown in Figures 1 and 2 (main text) are based on the data in Tables 2 and 3. Table 4 presents the means of absolute distance, correlations between pair estimates, and percentage of estimates that “bracket” the truth (Larrick et al., 2012).

MSAE values were computed as detailed in the main text. MAE values were computed as follows: first, for each pair, we averaged the two estimates of the judges, by jar; next, we calculated the absolute error of the collective judgment, defined as its distance from the true number of candies in the jar; finally, the MAE was obtained for each pair by averaging the absolute errors across the nine jars. This measure reflects the overall accuracy of a specific pair of judges and the values can be interpreted in terms of candies.

Table 2

Mean Absolute Errors (MAE) and Mean of Standardized Absolute Errors (MSAE) for Related and Unrelated Pairs, by Zygosity

		Study 1			**Study 2**		
Zygosity	Type	N of pairs	MAE	MSAE	N of pairs	MAE	MSAE
Monozygotic	Related	36	48.3	0.11	33	48.8	-0.002
	Unrelated	2,520	46.3	-0.002	2,112	46.6	0.11
Dizygotic same-sex	Related	70	45.5	0.06	56	40.2	-0.09
	Unrelated	9,660	44.4	-0.0004	6,160	41.4	0.0008
Dizygotic different-sex	Related	52	43.0	-0.02	54	38.2	-0.02
	Unrelated	5,304	43.5	0.0002	5,724	38.4	0.0002

Table 3

Mean Absolute Errors (MAE) and Mean of Standardized Absolute Errors (MSAE) for Related vs. Unrelated Pairs, by Zygosity, Controlling for Gender Composition

		Study 1			**Study 2**		
Zygosity	Type	N of pairs	MAE	MSAE	N of pairs	MAE	MSAE
Monozygotic	Related	36	48.3	0.16	33	48.7	0.10
	Unrelated	1,224	46.1	-0.005	1,032	46.4	-0.007
Dizygotic same-sex	Related	70	45.4	0.05	56	39.9	-0.10
	Unrelated	4,764	44.5	-0.0008	3,168	41.5	0.002
Dizygotic different-sex	Related	52	43.0	-0.009	54	38.2	-0.01
	Unrelated	2,652	43.4	0.0002	2,862	38.5	0.0003

Table 4

Mean of Absolute Distance, % of Estimates Bracketing the Truth, and Mean of Correlations between Estimates, for Related vs. Unrelated Pairs, by Zygosity, Controlling for Gender Composition

		Study 1				**Study 2**			
Zygosity	Type	N of pairs	Distance	Bracket	Correl.	N of pairs	Distance	Bracket	Correl.
Monozygotic	Related	36	70.70	36.42	.14	33	60.44	28.27	.11
	Unrelated	1,224	75.52	38.46	-.02	1,032	68.16	35.21	-.03
Dizygotic same-sex	Related	70	61.86	35.14	.06	56	67.62	42.20	-.03
	Unrelated	4,764	70.21	37.80	-.01	3,168	65.62	39.25	-.01