Department of Labour Economics

HdBA
Hochschule der
Bundesagentur für Arbeit
University of Applied Labour Studies

HdBA Discussion Papers in Labour Economics

No. 21-04

The Effect of Age Diversity in Groups on Peer Evaluations and Individual Performance

Katja Görlitz

Tim Sels

The Effect of Age Diversity in Groups on Peer Evaluations and Individual Performance

Katja Görlitz[a] and Tim Sels[b,*]

[a] HdBA, RWI and IZA

Hochschule der Bundesagentur für Arbeit, Seckenheimer Landstraße 16, 68163 Mannheim, Germany, katja.goerlitz@hdba.de

[b] FU Berlin

FU Berlin, Boltzmannstr. 20, 14195 Berlin, Germany, tim.sels@fu-berlin.de

Abstract: This study analyzes how individuals evaluate their peers' performance in a high stakes tournament in response to being randomly assigned to an age homogenous or heterogeneous group using data from two TV shows. The data also allows us to explore superior evaluations because it contains objective ratings from an independent expert. Additionally, this study investigates how age diverse groups affect individual performance in professional golf tournaments. The results show that peer and superior evaluations as well as individual performance are lower in age diverse groups. Further evidence suggests that these effects occur in the short run, but fade away once group members have gotten to know each other.

Keywords: Age diversity, peer evaluation, superior evaluation, performance

JEL classification: J14, M12, M54

* All correspondence to Tim Sels.

1. Introduction

The workforce is becoming increasingly diverse (Becker 2016). Population aging, longevity and extensions of the working life increases age diversity in organizations. How to compose effective work groups is an important leadership question as it helps managers, supervisors and execute staff to increase productivity and profits. The previous empirical literature provides ambiguous findings on the relationship between age diversity in groups and productivity. While some studies find that age diversity correlates positively with innovative ideas (e. g. Bantel and Jackson 1989) or problem-solving tasks (e. g. Kilduff et al. 2000), others show negative or no performance effects (e. g. Bunderson and Sutcliffe 2002, Timmerman 2000).[1] Kunze et al. (2011) suggest that this ambiguity "may be traced back to the researchers' neglect of possible mediators and moderators" (p. 265).

This study adds novel insights into these mediating factors by analyzing how group members behave towards each other when competing in a tournament against each other. We investigate data from more than 2,600 candidates from the TV cooking show "Come dine with me" where the aim is to prepare the most delicious three-course dinner and more than 1,600 candidates from the TV show "Shopping Queen" where the candidates buy a stylish outfit in a given time. In both shows, five group members "cook" and "shop" against each other to win the high stakes tournament price. Importantly, the candidate with the highest scores given by the other players wins the price. This is an interesting setting because workers face a trade-off: They could win with higher likelihood by downgrading the other players performance or behave in a more fair and pro-social behavior by reporting their correct subjective evaluations (which could be biased as well by preferences or tastes but not necessarily towards one's self-interest to win). We investigate whether group members that are randomly assigned to either age homogeneous or heterogeneous groups differ in how many points they award to their contestants.

This setting is also interesting from a management perspective because organizations have to rely on peer evaluations when executive staff cannot observe the objective performance, e. g., in the case of teamwork or multitasking. Large firms frequently use 360° feedback for this purpose where some authors estimate a utilization rate of more than 90 percent (Edwards and Ewen 1996, Baroda et al. 2012, Mone and London 2018, Rose and Biringer 2020). This study contributes to this literature analyzing the effects of peer evaluations. Because one of the two TV shows also reports the evaluations of an independent expert whose task is to report an objective review and who therefore does not compete in the tournament, we can also analyze how age diversity affects superior evaluations. To our knowledge, we are not aware of a study

[1] See the second section of this study for a more comprehensive review of the literature.

that investigates whether age diversity in groups affects the quality of the peer and the superior evaluations.

This study also contributes to the literature analyzing performance effects of age diversity in groups by using data from professional golf tournaments of the PGA Golf Tour. Even though the tournament organizers assign the players randomly to groups of three in which the golfers have to play together for the first two rounds, only individual performance determines the winner of the tournament and the second, third etc. best. This means that no group price is available. This setting differs from most of the previous literature as we do not analyze group performance, but individual performance in response to changing the diversity composition of the group. It has practical relevance for working situations where individuals perform individual tasks, but companies assign them to departments or divisions solely because of organizational reasons and not because working in groups is necessary to produce the output. Our analyses open the black box of whether and how age diversity of groups affects individual behavior in these settings.

The results indicate that individuals in age heterogenous groups award significantly fewer points to their peers. This is true regardless of using data from the cooking or the shopping show, meaning two settings that differ in the tasks that individuals have to perform and evaluate. This result spills over to the evaluation of an independent expert, suggesting that age diversity also affects the superior evaluations negatively. Further analyses reveal that the reason for this evaluation differential is likely due to individual performance that is found to be lower in age diverse groups as well. Importantly, these results apply to settings where individuals have recently gotten to know each other. Therefore, they probably represent initial effects because further suggestive evidence shows that once group members have gotten familiar with another the performance effect vanishes.

The paper is organized as follows. The next section summarizes the theories and reviews the previous empirical literature on age diversity to derive the hypotheses of the study. Section 3 presents the data, the empirical strategy and the results on the relationship between age diversity and performance evaluations. Section 4 answers the question on how age diversity affects individual performance. The last section concludes the study.

2. Theoretical background and literature review

The effect of age diversity on peer and superior evaluations

Individuals seek to identify themselves as a member of a group based on self-categorization to gain social identity (Tajfel and Turner 1986, Turner 1987). The similarity-attraction theory

states that groups have a stronger social cohesion if group member's attributes such as demographic characteristics like individuals' age are similar within a group (Berscheid and Walster 1969, Byrne 1971). Kelly and Presslee (2017) find that individuals in groups with stronger identification compete less and are, thus, less willing to win against another. Such behavior could be explained by the other-regarding preference theory that states that people have concerns for the well-being of others (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2002).[2] Chen and Li (2009) confirm empirically that individuals have higher concerns for the well-being of others in their own group, if group identity is higher.

Several empirical studies demonstrate that age diversity within the group deteriorates communication (Zenger and Lawrence 1989, Ellwart et al. 2014, De Meulenaere and Kunze 2020) and increases conflicts (Jehn et al. 1997, Knight et al. 1999, Pelled et al. 2001, Colquitt et al. 2002, Luksyte et al. in press). This could explain why another stream of the literature shows that age diverse groups also suffer from higher turnover of its members (Wagner et al. 1984, O'Reilly et al. 1989, Jackson et al. 1991, Tsui et al. 1992, Wiersema and Bird 1993, Kunze et al. 2021). For the reasons of lower other-regarding preferences, communication problems and higher conflict potential, we expect lower peer evaluations in age diverse groups within tournaments, which leads to the first hypothesis:

Hypothesis 1: *Age diversity leads to lower peer evaluations among the group members.*

Murphy (2008) and Scullen et al. (2000) suggests that peer evaluations can suffer from biases. Prendergast and Topel (1996) argue that even supervisors who evaluate the performance of their subordinates and do not compete with them in a tournament provide biased ratings, e. g. caused by social preferences. Bandiera et al. (2009) demonstrate that supervisors, who receive a fixed compensation that is independent of group performance, favor individuals who have e. g. similar demographic characteristics when choosing the members of their team. Thus, the second hypothesis is:

Hypothesis 2: *Age diversity among the group members leads to lower superior evaluations.*

The effect of age diversity on individual performance

When individuals within a diverse group suffer from a worse atmosphere, this could spill over to a negative performance effect. The empirical literature confirms that age diversity has indeed a negative impact on performance (Zajac et al. 1991, West et al. 1999, Timmerman 2000, Ely

[2] See Cooper and Kegal (2006) as well as Fehr and Schmidt (2006) for a review of this literature.

2004, Leonard et al. 2004, Kearney and Gebert 2009, Kunze et al. 2011, Hafsi and Turgut 2013, Ali et al. 2014, De Meulenaere et al. 2016, De Meulenaere and Kunze 2020, Kunze et al. 2021, Luksyte et al. in press). However, this literature is not conclusive because other studies find positive (Kilduff et al. 2000, Wegge et al. 2008, Kearney et al. 2009, Gong et al. 2021) or insignificant effects (Bantel and Jackson 1989, Wiersema and Bantel 1992, Simons et al. 1999, Bunderson and Sutcliffe 2002, van der Vegt and Bunderson 2005). Comprehensive reviews summarize that the effect sign depends e. g. on the analyzed performance task (Milliken and Martins 1996, Williams and O'Reilly 1998, Jackson et al. 2003, Joshi and Roh 2009, Wegge and Schmidt 2009, Bell et al. 2011, Boehm et al. 2011, van Dijk et al. 2012, Schneid et al. 2016).

The literature finding positive or insignificant effects investigates performance tasks that require an innovative solution (Bantel and Jackson 1989, Wiersema and Bantel 1992, Kearney et al. 2009, Luksyte et al. in press) or can be characterized as a complex problem which can be solved best by group members who have acquired knowledge in different areas of expertise (Simons et al. 1999, Kilduff et al. 2000, Bunderson and Sutcliffe 2002, van der Vegt and Bunderson 2005, Wegge et al. 2008, Gong et al. 2021). Because our study analyzes golf performance which does neither require an innovative solution nor knowledge in different areas of expertise, the third hypothesis is:

Hypothesis 3: *Members of age diverse groups exhibit a lower performance in our setting.*

Individuals divide themselves into social categories (Tajfel et al. 1971, Tajfel and Turner 1986, Turner 1987). These categories include amongst other things demographics (e.g., age, gender) as well as attitudes, beliefs and values (Mannix and Neale 2005). While group members can immediately recognize some visible categories (age, gender), differences in others like attitudes, beliefs and values only become apparent over time (Harrison et al. 1998, Harrison et al. 2002). Fiske and Neuberg's (1990) continuum model states that the first impression automatically or unconsciously leads to categorization based on demographics (age, gender) or other visible characteristics. After getting to know each other, other categories become more important. The same prediction arises from the dual process model stating that individuals always choose the simpler information processing (Brewer and Feinstein 1999). Allport's (1954) contact-hypothesis states that getting to know each other more closely can reduce categorical prejudices. Thus, the fourth hypothesis is:

Hypothesis 4: *Members of age diverse groups exhibit a lower performance in our setting mainly in the short run.*

3. The effect of age diversity on peer and superior evaluations

3.1 Data and empirical model

Data

This section uses data from two different sources. First, it exploits data from the German TV show "*Das perfekte Dinner*" based on the concept of the British TV series "Come Dine with Me". Each calendar week, five volunteer contestants from one city compete against each other by cooking a three-course dinner of their own choice for the others. Every contestant serves a dinner in his or her apartment at another day of the workweek. The competitors do not know each other upfront and meet for the first time on Monday which is the day when the first player prepares his or her dinner. All competitors aim to create the best dinner to win the cash prize of 3,000 euro. The contestants determine the winner by peer evaluations. After the respective dinner evening, each of the four competitors anonymously awards scores on a scale from 0 to 10 to the chef. Thus, every chef can reach a maximum score of 40 points. The announcement of the winner takes place at the award ceremony at the end of the week after the last chef has served his or her dinner. While each contestant has the information about his or her overall score at the ceremony, the individual ratings for each candidate remain anonymously and are announced not until the TV broadcasting. The data includes TV episodes covering the period from January 2007 through January 2021 and encompasses 529 weeks with 2,645 candidates as well as 10,580 peer evaluations.[3]

Second, this section uses data from the TV show "*Shopping Queen*" where five voluntary competitors compete in each calendar week by going shopping for at most four hours to find a stylish outfit that is in line with a prescribed theme. The limit of the budget for the outfit is at most 500 euro, including clothing, shoes, accessories, hairstyle and make-up. Each contestant must find, buy and present the outfit at another day of the workweek. The candidates receive the prescribed theme on the day where all players meet each other for the first time. The other four competitors evaluate the performance anonymously on a scale from 0 to 10 after the presentation of the outfit. In contrast to the cooking show, a superior rating also exists. At the end of the week, the German fashion expert Guido Maria Kretschmer gives a score for each of the five candidates using the same scale from 0 to 10. Thus, the external superior rating influences the outcome of the competition. The superior evaluation of the candidates is based on

[3] Some episodes are excluded, e. g., where only four contestants participated in the show (e.g. in calendar weeks with a holiday), where shows with other participants/ rules took place (e. g. with celebrities as contestants), where episodes or information like the candidates' age were unavailable.

the video recordings made, without including the awarded critique or scores of the other candidates. The superior gives his rating to each candidate at the award ceremony and announces the winner as well as the overall scores, while the individual ratings remain anonymously until the TV broadcasting. The maximum overall score is 50 points. Only the participant with the highest scores will win a cash prize of 1,000 euros.[4] The data covers the period from April 2013 through March 2021. In total, 322 weeks, 1,610 candidates and 6,440 peer evaluations as well as 1,610 superior evaluations are available.[5]

Table 1 contains the summary statistics for both data sources. The average of the peer evaluations given by the other players is 7.6 in the cooking and 7.8 in the shopping show. In the shopping show, the superior evaluation has a slightly lower mean of 7.5, but a similar standard deviation as the peer evaluations. While almost equal numbers of men and women are in the cooking show, the participants in the shopping show are almost exclusively women because of the 1,610 players there were only two men. The chefs are older than the shopping contestants and their standard deviation is slightly lower. Generating the standard deviation of the age of the contestants at the group level gives a more detailed view on the extent of age diversity in both data sources. Again, the age variation of the contestants is slightly lower in the cooking show. Nevertheless, both data sources show that age diversity varies greatly over the groups. While some groups are age homogenous (with a minimum value of 2.2 in the cooking show and with 3.9 in the shopping show), others are rather heterogenous (where the maximum is as high as 24.7 and 23.2, respectively). In particular, the group with the lowest overall variation consists of players having the following ages: 20, 23, 25, 25 and 25 years. In contrast, the most heterogenous group comprises players being 23, 32, 53, 63 and 85 years old.

[4] In rare cases, the winner receives only a non-cash prize (e. g., a bag worth 4,000 euros).
[5] Again, this study excludes shows with missing information and irregular shows like those with four competitors or special shows.

Table 1: Summary statistics of the cooking and the shopping show

Variables	Cooking show				Shopping show			
	Obs.	Mean	Min	Max	Obs.	Mean	Min	Max
Peer Evaluation	10,580	7.64 (1.31)	0	10	6,440	7.83 (1.25)	0	10
Superior Evaluation	Not available				1,610	7.46 (1.26)	1	10
Indicator for men (1-yes, 0- no)	2,645	0.47	0	1	1,610	0.001	0	1
Number of men per group	529	2.37	0	4	322	0.01	0	1
Age of the contestants in years	2,645	39.66 (11.68)	18	85	1,610	36.50 (11.85)	18	81
Standard deviation of the age in years per group	529	11.34	2.2	24.7	322	12.52	3.9	23.2

Notes: The standard deviations of the means are shown in parentheses.

Empirical Model

The following OLS regression analysis estimates the effect of age diversity on the evaluation scores that each contestant *i* gets from each other player *j* in group (or equivalently week) *k*:

$$evaluation\ score_{i,j,k} = \alpha + \beta\ age\ diversity_k + X_i'\delta + X_j'\gamma + X_k'\eta + \varepsilon_{i,j,k} \quad (1)$$

where the dependent variable *evaluation score* refers to the peer evaluations. The independent variable of interest is *age diversity*. The regression controls for the characteristics of the player (X_i), the competitors (X_j) as well as group-specific covariates (X_k). The age of the contestant and the age of the other players account for potential correlations between age diversity and the players' ages. Weekday fixed effects that control for the day of the week where a contestant cooks or shops absorb day-related scoring effects.[6] In the cooking show data, the corresponding regression additionally incorporates gender covariates. The group characteristics include the average age and for the cooking show the share of men per group. ε is the idiosyncratic error term. Inference is based on robust standard errors.

The identification strategy exploits the exogenous variation in age diversity across groups to estimate the effect on the players' evaluation score. The variation derives from the setting of the TV show. The TV production team assigns the contestants randomly to their competitors, who could thereby not self-select themselves into another group. As already explained in the data section, this is because the locations of the series change from week to week to another

[6] Previous research shows that day-related effects can have a significant impact on the performance rating (Schüller et al. 2014, Blum and Wenskat 2020).

city where the contestants have to have an apartment. In addition, the competitors meet for the first time on the day when the recording of the first episode begins.

To find out whether age diversity only affects the peer evaluations or also spills over to the evaluation of an independent superior, the model estimates equation (1) again using the evaluation score given by the superior from the shopping show as the independent variable. Apart from that, this specification also changes the control variables. Because the superior is the same in all shows, controlling for the age of the evaluator becomes obsolete. As in the analysis of the shopping data before, the model cannot take gender into account because almost all players are women and the superior is always a man.

Because the previous literature analyzes alternative definitions of age diversity (Harrison and Klein 2007), we define *age diversity* in different ways to prove the robustness of our results. The most straightforward measure is to use the standard deviation of the contestants' age per group that varies tremendously across groups (see Table 1 for summary statistics). This measure considers absolute age differences. To analyze relative differences, we additionally define age diversity at the group level as the coefficient of variation which refers to the standard deviation divided by the average age (Timmerman 2000) as well as the standard deviation of the logarithm of participants' age (Leonard et al. 2004). These measures shed light on the question whether differences at older ages might be less pronounced than at younger ages. Furthermore, we allow for a more flexible functional form by estimating quartiles and tertiles of the group-specific standard deviation of age. This means estimating the standard deviation, sorting them in descending order and dividing them into four or three groups of equal size, respectively. Put differently, the lowest (highest) quartile contains the groups that belong to the bottom (top) 25% percent of all groups being characterized by the lowest (highest) age diversity. By doing so, this specification analyzes whether the effect is linear over age diversity or whether some parts of the distribution are affected more severely.

Further sensitivity analyses include the estimation of alternative standard errors. Because the players interact with each other at the group level, their errors might be serially correlated (Moulton 1990). Without accounting for the correlated errors, the models would underestimate standard errors and even non-significant effects could become statistically significant. Therefore, one robustness specification clusters the standard errors at the group level, i. e. using week-year clusters and bootstrapped standard errors. Re-estimating the main results by applying the ordered Logit and the ordered Probit model serves as a sensitivity check to find out whether the linear probability model is the adequate model specification. Furthermore, Schüller et al. (2014) demonstrate that the age difference between the player and the rater affects the

performance ratings which is why we introduce the age difference as an additional control variable.

3.2 Results

Table 2 shows that age diversity reduces the peer evaluation scores that a player gets. Column (1) presents the results from the cooking show, documenting that an increase in the standard deviation by one unit reduces the evaluation scores given by each peer by -0.021. Related to the fact that always four players evaluate the remaining contestant, the overall score is 0.084 points lower. As the variation between the most homogenous and the most heterogenous group is 22.5 ($= 24.7 - 2.2$) as can be seen from Table 1, this relates to an overall penalty of the most heterogenous group compared to the most homogeneous group of almost two points ($= 22.5 \times -0.084$). This is a non-negligible effect size. In the shopping show, the corresponding estimate of the age diversity is -0.026 points for every player (column (2)). The finding that age diversity decreases peer evaluations is robust to analyzing alternative measures of age diversity. Interpreting the results of the quartiles and the tertiles further reveals that the effect of age diversity is slightly more pronounced when comparing groups with the lowest compared to groups with the highest age diversity.

The third column reveals that age diversity even influences the evaluation scores given by an independent superior who does not benefit himself from up- or downgrading others. In every of the used definitions of age diversity, the superior gives on average a lower evaluation score to players from more age diverse groups. When investigating quartiles or tertiles, groups being on the top of the diversity distribution experience the largest deduction of the superior evaluation. Nevertheless, the estimates of the superior evaluation are smaller in magnitude compared to the peer evaluations and lack statistical precision.

Table 2: Results of age diversity on peer and superior evaluation

	Peer evaluation		Superior evaluation
	Cooking show	**Shopping show**	**Shopping show**
	(1)	**(2)**	**(3)**
Age diversity defined as…			
the standard deviation of the age by group	-0.021 ***	-0.026 ***	-0.024 *
	(0.004)	(0.006)	(0.012)
the coefficient of variation: standard deviation divided by average age by group	-0.843 ***	-1.010 ***	-0.975 **
	(0.151)	(0.236)	(0.468)
the standard deviation of the logarithm of the age by group	-0.953 ***	-1.611 ***	-1.087 **
	(0.163)	(0.275)	(0.535)
… quartiles of the standard deviation of the age by group			
First (lower) quartile		*Reference group*	
Second quartile	-0.021	-0.057	-0.009
	(0.036)	(0.045)	(0.087)
Third quartile	-0.168 ***	-0.055	-0.148
	(0.037)	(0.044)	(0.091)
Fourth (upper) quartile	-0.194 ***	-0.268 ***	-0.205 **
	(0.037)	(0.052)	(0.102)
… tertiles of the standard deviation of the age by group			
First (lower) tertile		*Reference group*	
Second tertile	-0.109 ***	-0.077 **	-0.002
	(0.031)	(0.038)	(0.075)
Third (upper) tertile	-0.182 ***	-0.238 ***	-0.201 **
	(0.032)	(0.042)	(0.090)
Age of peer awarding scores	Yes	Yes	No
Age of player being evaluated	Yes	Yes	Yes
Gender of peer awarding scores	Yes	No	No
Gender of player being evaluated	Yes	No	No
Weekday fixed effects	Yes	Yes	Yes
Average age per group	Yes	Yes	Yes
Share of men per group	Yes	No	No
Obervations	10,580	6,440	1,610

Notes: The table reports the OLS results from regressing peer evaluation (from the cooking and the shopping show in columns (1) and (2), respectively) and from regressing the superior evaluation in the shopping show (column (3)) on various measures of age diversity in addition to further covariates. See equation (1) for further information. Statistical significance: $p<0.1$ *, $p<0.05$ **, $p<0.01$ ***.

Table A1 contains the results of the sensitivity analyses. The sensitivity checks for the peer evaluation demonstrate that the results are robust to clustering and bootstrapping the standard

errors, different estimation models (ordered Logit, ordered Probit) and using the age difference as further control variable. The same conclusion applies to analyzing superior evaluations, except for results using clustered and bootstrapped standard errors. In conclusion, peers understate their contestants' performance in age diverse groups in high-stakes tournaments. This reinforces our first hypothesis. There is also evidence for the second hypothesis, nevertheless, two sensitivity checks reveal that the estimates are less precise compared to the estimates of the peer evaluations.

4. The effect of age diversity on individual performance

4.1 Data and empirical model

Data

This section uses PGA Golf Tour 2002 data from Guryan et al. (2009). The PGA Golf Tour organizes various tournaments with professional golfers mainly in the USA. Each tournament has four rounds. The organizer assigns the golfers randomly into groups of 3 people which remain unchanged until the end of the second round. This is why this study analyzes the first and second round only. When the organizers assign the golfers to the groups, they consider the players' performance by creating groups of golfers who belong to the same performance category. Category 1 refers to the best and category 3 to the worst players. This categorization remains almost unchanged during the season. To test our hypotheses three and four, we expect that members of groups with category 3 players do not previously know each other. This is because these groups consist of participants such as local qualifiers who have participated few times (five years) on the PGA Golf Tour and played few tournaments (3.9) within the PGA Golf Tour. Furthermore, category 3 includes approximately 30% more players compared to category 1. For this reason, we refer to category 3 players as "previously unknown group members". Category 1 players include tournament winners and the top 25 money earners from the previous year, as well as PGA Golf Tour life members like Tiger Woods who have achieved outstanding career achievements. We expect that category 1 golfers more often previously know each other. On average, a category 1 golfer has experienced twelve years on the PGA Golf Tour and participated in 12.6 tournaments. We refer to category 1 players as "previously known group members" in the remaining study.

All players compete against each other, including the players within each group, to win the high-stakes price.[7] The golfer who attains the lowest cumulative number of strokes wins. Thus,

[7] On average, a tournament allocates 3.7 million dollars in prize money of which the winner gets about 18% and the top 10 approximately 60% (ESPN 2021).

player A with 70 strokes performs better than golfer B with 72 strokes. For ease of interpretation, we define a new variable, henceforth the performance score, which multiplies the number of strokes with -1. The higher the performance of a player is (because of having fewer strokes), the higher is the corresponding performance score. We merge data of the age of the golfers to the PGA Golf Tour 2002.[8] Using data on the first and second round, leaves us with 193 players from 992 groups and 2,976 performance scores.

Table 3 presents the summary statistics for the PGA Golf Tour 2002 data separately for the previously unknown and the previously known group members. The previously unknown group members have the lowest and the previously known group members have the highest performance score because they require the most (72.2) and the fewest (70.5) strokes on average, respectively. This also corresponds to the handicap (ability score), which is better for known group members. Female golfers did not participate in the tournaments studied. The group of unknown members is younger than the group of known members. Generating the standard deviation of the age of the player at the group level provides a more detailed view on the extent of age diversity. Both groups show a wide range of age diversity. While some groups are age homogenous (with a minimum value of 0.6), others are more heterogenous (where the maximum is as high as 17.8 and 15.3, respectively). The group with the lowest variation consists of players having for example the following ages: 23, 24 and 24 or 39, 40 and 40 years. In contrast, the most heterogenous group of previously unknown members comprises players being 24, 47 and 59 years old.

Table 3: Summary statistics of the PGA Golf Tour 2002 data

Variables	**Group members who are previously …**							
	unknwon				**known**			
	Obs.	Mean	Min	Max	Obs.	Mean	Min	Max
Performance Score (negative number of strokes)	462	-72.24 (3.51)	-84	-63	2,514	-70.48 (3.08)	-84	-61
Ability (handicap)	462	1.2 (1.5)	-1.5	8.1	2,514	-0.3 (0.7)	-2.5	4.1
Age of the players in years	462	31.94 (7.44)	17	59	2,514	36.47 (5.88)	22	59
Standard deviation of the age in years per group	154	6.37	0.6	17.8	838	5.31	0.6	15.3

Notes: We defined the performance score as the number of strokes multiplied by -1. This guarantees that a better golfer with fewer strokes has a higher value on the performance score. The standard deviations of the means are shown in parentheses.

[8] The age data are mainly from the following websites: Wikipedia, PGA, BlueGolf, ESPN, Yahoo!Sports.

Empirical Model

The following OLS regression analysis estimates the effect of age diversity on the performance scores that each player *i* achieves in group *k*:

$$performance\ score_{i,k} = \alpha + \beta\ age\ diversity_k + X_i'\delta + \eta X_k + \varepsilon_{i,k} \quad (2)$$

The dependent variable *performance score* refers to the players' golf performance which is better the higher the score is. Following our analyses of peer evaluations in section 3.1, the model uses various measures of *age diversity* as independent variable where the main specification uses the standard deviation of the players' age per group. The regression controls for the age of the player (X_i) to account for potential correlations between the age diversity and the players' ages. X_k comprises round fixed effects that absorb round-related scoring effects including weather conditions. X_k also includes the average age of the group. ε is the idiosyncratic error term. Inference is based on robust standard errors. Exploiting the fact that golfers are randomly assigned to groups, the identification strategy uses the exogenous group-specific variation in age diversity in the groups to estimate the effect on players' performance scores.

Further sensitivity analyses reveal whether age diversity affects the objective performance scores. Clustering the standard errors at the group level which refers to group-round clusters and bootstrapping the standard errors examines whether alternative standard errors produce robust results. An ordered Logit and an ordered Probit model re-test whether a non-linear model can confirm the OLS results. In contrast to the dinner and the shopping show, a further sensitivity analyses adds control variables of players' ability (handicap) and experience (years on tour). To support the fourth hypothesis, another model specification examines first round effects only, because players have just recently gotten to know each other in the first round.

4.2 Results

Table 4 summarizes the effects of age diversity on the golf performance score. Column (1) shows that age diversity has significant negative effects on the performance for previously unknown group members. An increase in the standard deviation by one unit decreases the performance for each group member who is previously unknown by 0.17 additional strokes per round. Because every golfer is playing two rounds in the same group, they perform 0.34 strokes worse. The deviation between the most homogeneous and heterogeneous group is 17.2 (= 17.8 – 0.6). Thus, the most age diverse group is at a disadvantage of about 5.87 strokes. If one takes the respective tournament winner as a reference, a golfer will earn about $544,000 less in prize money. This conclusion is robust to analyzing alternative measures of age diversity. In contrast,

the results of the previously known members in column (2) are statistically insignificant regardless of using alternative measures of age diversity. Importantly, the magnitude of the coefficients is much smaller compared to the results of the unknown members. This indicates that the difference in the results from column 1 and 2 is not due to a lack of efficiency, but shows that no performance effect for players with previously known members exists. The sensitivity analysis shown in Table A2 reinforces this conclusion in all specifications. In conclusion, the results confirm our third and fourth hypotheses.

Table 4: Results of age diversity on the performance score

	Performance Score of group members who are previously …	
	unknown	**known**
	(1)	**(2)**
Age diversity defined as…		
… the standard deviation of the age by group	-0.171 ***	0.011
	(0.050)	(0.027)
… the standard deviation divided by average age by group	-5.677 ***	0.598
	(1.649)	(0.962)
… the standard deviation of the logarithm of the age by group	-5.374 ***	0.551
	(1.622)	(0.904)
… quartiles of the standard deviation of the age by group		
First (lower) quartile		
Second quartile	-0.154	0.271
	(0.417)	(0.176)
Third quartile	-0.480	-0.073
	(0.471)	(0.173)
Fourth (upper) quartile	-1.178 **	0.143
	(0.485)	(0.181)
… tertiles of the standard deviation of the age by group		
First (lower) tertile		
Second tertile	-0.064	0.199
	(0.383)	(0.148)
Third (upper) tertile	-1.384 ***	-0.018
	(0.425)	(0.157)
Age of player	Yes	Yes
Average age by group	Yes	Yes
Round fixed effects	Yes	Yes
Obervations	462	2,514

Notes: The table shows the effects of the OLS regression of the performance score which is the negative number of a players' strokes for the PGA Golf Tour 2002 data. The independent variable of interest is based on different definitions of age diversity. Besides the model includes further fixed effects. Column (1) refers to previously unknown group members while column (2) focuses on previously known group members. Statistical significance: $p<0.1$ *, $p<0.05$ **, $p<0.01$ ***.

5. Conclusion

This study analyzes the effects of age diversity on peer and superior evaluations as well as on individual performance in settings where the performance does not require teamwork or diverse expert knowledge. It is comparable to settings where the organizational structure divides individuals' group-wise, e. g. into sections or departments, but every single worker individually performs his/her own tasks. The results show that age diverse groups award significantly fewer points in peer evaluations compared to homogeneous groups. The same applies to the superior evaluations of independent experts who do not participate in the tournament themselves, albeit the estimates are less precise. The reason for this might stem from subjective biases in the evaluation due to a worse group atmosphere. Alternatively, the group atmosphere could affect the work effort which in turn could spill over to the individual performance. Further results indicate that age diversity in tournaments indeed spills over to individuals' performance, when group members do not previously know each other. Once the group members have gotten to know each other, this performance effect renders statistically insignificant. This evidence is in favor of theories predicting that self-categorization at first build upon visible factors like demographics.

From a management perspective, it is relevant to know that superior evaluations are biased when members from age diverse groups compete e. g. for a promotion with members from homogenous groups. The same applies to peer evaluations when using the 360° Feedback to determine the bonus or a performance-related pay for groups. Showing that age diversity affects individual performance is also of great relevance to companies seeking to improve productivity. Even though this disadvantage of age diverse groups mirrors performance differences which could explain pay differentials, it is exogenous from an individual point of view and only represents the "luck" of being assigned to an age homogeneous group. This could have implications for individuals' motivation. Last, further results suggest that these effects are more pronounced in settings where group members have just recently gotten to know each other, meaning during the time when managers or supervisors built new teams. It remains a topic for future research to provide more direct evidence on when and how group diversity affects individuals' behavior in the longer run.

Bibliography

Ali, Muhammad, Yin Lu Ng and Carol T. Kulik (2014), "Board Age and Gender Diversity: A Test of Competing Linear and Curvilinear Predictions", *Journal of Business Ethics*, 125 (3), pp. 497-512.

Allport, G. W. (1954): *The Nature of Prejudice*, Addison-Wesley: Boston.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2009), "Social connections and incentives in the workplace: evidence from personnel data", *Econometrica*, 77 (4), pp. 1047-1094.

Bantel, Karen A. and Susan E. Jackson (1989), "Top Management and Innovations in Banking: Does the Composition of the Top Team Make a Difference?", *Strategic Management Journal*, 10 (S1), pp. 107-124.

Baroda, Satyawan, Chhavi Sharma and Jyoti Kandpal Bhatt (2012), "360 Degree Feedback Appraisals- An Innovative Approach of Performance Management System", *International Journal of Management & Information Technology*, 1 (2), pp. 53-66.

Becker, Florian (2016): *Teamarbeit, Teampsychologie, Teamentwicklung. So führen Sie Teams!*, Springer-Verlag: Berlin, Heidelberg.

Bell, Suzanne T., Anton J. Villado, Marc A. Lukasik, Larisa Belau and Andrea L. Briggs (2011), "Getting Specific about Demographic Diversity Variable and Team Performance Relationships: A Meta-Analysis", *Journal of Management*, 37 (3), pp. 709-743.

Berscheid, Ellen and Elaine Hatfield Walster (1969): *Interpersonal Attraction*, Addison-Wesley: Boston.

Blum, Peter and Marc Wenskat (2020), *Why Monday Never Wins. An Example of the Secretary Problem*, June.

Boehm, Stephan A., Miriam K. Baumgaertner, David J. G. Dwertmann and Florian Kunze (2011): Age diversity and its performance implications - Analyzing a major future workforce trend, in: S. Kunisch, S. Boehm and M. Boppel (ed.), *From Grey to Silver,* Springer-Verlag: Berlin, Heidelberg, pp. 123-144.

Bolton, Gary E. and Axel Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity, and Competition", *The American Economic Review*, 90 (1), pp. 166-193.

Brewer, Marilynn B. and Amy S. Harasty Feinstein (1999): Dual Processes in the Cognitive Representation of Person and Social Categories, in: S. Chaiken and Y. Trope (ed.), *Dual-process theories in social psychology,* The Guilford Press: New York and London, pp. 255-270.

Bunderson, J. Stuart and Kathleen M. Sutcliffe (2002), "Comparing Alternative Conceptualizations of Functional Diversity in Management Teams: Process and Performance Effects", *Academy of Management Journal*, 45 (5), pp. 875-893.

Byrne, Donn (1971): *The attraction paradigm*, Academic Press: New York and London.

Charness, Gary and Matthew Rabin (2002), "Understanding Social Preferences with Simple Tests", *The Quarterly Journal of Economics*, 117 (3), pp. 817-869.

Chen, Yan and Sherry Xin Li (2009), "Group Identity and Social Preferences", *The American Economic Review*, 99 (1), pp. 431-457.

Colquitt, Jason A., Raymond A. Noe and Christine L. Jackson (2002), "Justice in Teams: Antecedents and Consequences of Procedural Justice Climate", *Personnel Psychology*, 55 (1), pp. 83-109.

Cooper, David J. and John H. Kagel (2006): Other-Regarding Preferences, in: J. H. Kagel and A. E. Roth (ed.), *The Handbook of Experimental Economics, Volume 2,* Princeton University Press: Princeton and Oxford, 1st Edition, pp. 217-289.

De Meulenaere, Kim, Christophe Boone and Tine Buyl (2016), "Unraveling the impact of workforce age diversity on labor productivity: The moderating role of firm size and job security", *Journal of Organizational Behavior*, 37 (2), pp. 193-212.

De Meulenaere, Kim and Florian Kunze (2020), "Distance matters! The role of employees' age distance on the effects of workforce age heterogeneity on firm performance", *Human Resource Management*, 60 (4), pp. 499-516.

Edwards, Mark R. and Ann J. Ewen (1996), "How to Manage Performance and Pay With 360-Degree Feedback", *Compensation & Benefits Review*, 28 (3), pp. 41-46.

Ellwart, Thomas, Silke Bündgens and Oliver Rack (2014), "Managing knowledge exchange and identification in age diverse teams", *Journal of Managerial Psychology*, 28 (7/8), pp. 950-972.

Ely, Robin J. (2004), "A field study of group diversity, participation in diversity education programs, and performance", *Journal of Organizational Behavior*, 25 (6), pp. 755-780.

Fehr, Ernst and Klaus M. Schmidt (1999), "A Theory of Fairness, Competition, and Cooperation", *The Quarterly Journal of Economics*, 114 (3), pp. 817-868.

Fehr, Ernst and Klaus M. Schmidt (2006): The economics of fairness, reciprocity and altruism - experimental evidence and new theories, in: S. -C. Kolm and J. M. Ythier (ed.), *Handbook of the Economics of Giving, Altruism and Reciprocity, Volume 1,* North Holland: w. p., 1st Edition, pp. 615-691.

Fiske, Susan T. and Steven L. Neuberg (1990): A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation, in: M. P. Zanna (ed.), *Advances in Experimental Social Psychology, Volume 23,* Academy Press: San Diego, New York, Berkeley, Boston, London, Sydney, Tokyo and Toronto, pp. 1-74.

Li, Yixuan, Yaping Gong, Anne Burmeister, Mo Wang, Valeria Alterman, Alexander Alonso and Samuel Robinson (2021), "Leveraging Age Diversity for Organizational Performance: An Intellectual Capital Perspective", *Journal of Applied Psychology*, 106 (1), pp. 71-91.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo (2009), "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments", *American Economic Journal: Applied Economics*, 1 (4), pp. 34-68.

Hafsi, Taïeb and Gokhan Turgut (2013), "Boardroom Diversity and its Effect on Social Performance: Conceptualization and Empirical Evidence", *Journal of Business Ethics*, 112 (3), pp. 463-479.

Harrison, David A., and Katherine J. Klein (2007), "What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations", *Academy of Management Review*, 32 (4), pp. 1199-1228.

Harrison, David A., Kenneth H. Price and Myrtle P. Bell (1998), "Beyond Relational Demography: Time and the Effects of Surface- and Deep-Level Diversity on Work Group Cohesion", *Academy of Management Journal*, 41 (1), pp. 96-107.

Harrison, David A., Kenneth H. Price, Joanne H. Gavin and Anna T. Florey (2002), "Time, Teams, and Task Performance: Changing Effects of Surface- and Deep-Level Diversity on Group Functioning", *Academy of Management Journal*, 45 (5), pp. 1029-1045.

Jackson, Susan E., Joan F. Brett, Valerie I. Sessa, Dawn M. Cooper, Johan A. Julin and Karl Peyronnin (1991), "Some Differences Make a Difference: Individual Dissimilarity and Group Heterogeneity as Correlates of Recruitment, Promotions, and Turnover", *Journal of Applied Psychology*, 76 (5), pp. 675-689.

Jackson, Susan E., Aparna Joshi and Niclas L. Erhardt (2003), "Recent Research on Team and Organizational Diversity: SWOT Analysis and Implications", *Journal of Management*, 29 (6), pp. 801-830.

Jehn, Karen A., Clint Chadwick, and Sherry M. B. Thatcher (1997), "To agree or not to agree: The effects of value congruence, individual demographic dissimilarity, and conflict on workgroup outcomes", *The International Journal of Conflict Management*, 8 (4), pp. 287-305.

Joshi, Aparna and Hyuntak Roh (2009), "The Role of Context in Work Team Diversity Research: A Meta-Analytic Review", *Academy of Management Journal*, 52 (3), pp. 599-627.

Kearney, Eric and Diether Gebert (2009), "Managing Diversity and Enhancing Team Outcomes: The Promise of Transformational Leadership", *Journal of Applied Psychology*, 94 (1), pp. 77-89.

Kearney, Eric, Diether Gebert and Sven C. Voelpel (2009), "When and How Diversity Benefits Teams: The Importance of Team Members' Need for Cognition", *Academy of Management Journal*, 52 (3), pp. 581-598.

Kelly, Khim, and Adam Presslee (2017), "Tournament group identity and performance: the moderating effect of winner proportion", *Accounting, Organizations and Society*, 56, pp. 21-34.

Kilduff, Martin, Reinhard Angelmar, Ajay Mehra (2000), "Top Management-Team Diversity and Firm Performance: Examining the Role of Cognitions", *Organization Science*, 11 (1), pp. 21-34.

Knight, Don, Craig L. Pearce, Ken G. Smith, Judy D. Olian, Henry P. Sims, Ken A. Smith and Patrick Flood (1999), "Top Management Team Diversity, Group Process, and Strategic Consensus", *Strategic Management Journal*, 20 (5), pp. 445-465.

Kunze, Florian, Stephan A. Boehm and Heike Bruch (2011), "Age diversity, age discrimination climate and performance consequences—a cross organizational study", *Journal of Organizational Behavior*, 32 (2), pp. 264-290.

Kunze, Florian, Stephan A. Boehm and Heike Bruch (2021), "It matters how old we feel in organizations: Testing a multilevel model of organizational subjective-age diversity on employee outcomes", *Journal of Organizational Behavior*, 42 (4), pp. 448-463.

Leonard, Jonathan S., David I. Levine, and Aparna Joshi (2004), "Do birds of a feather store together? The effects on performance of employees' similarity with one another and with customers", *Journal of Organizational Behavior*, 25, pp. 731-754.

Luksyte, Aleksandra, Derek R. Avery, Sharon K. Parker, Ying Wang, Lars U. Johnson and Loring Crepeau (in press), "Age diversity in teams: Examining the impact of the least agreeable member", *Journal of Organizational Behavior*.

Mannix, Elizabeth and Margaret A. Neale (2005), "What Differences Make a Difference? The Promise and Reality of Diverse Teams in Organizations", *Psychological Science in the Public Interest*, 6 (2), pp. 31-55.

Milliken, Frances J. and Luis L. Martins (1996), "Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups", *Academy of Management Review*, 21 (2), pp. 402-433.

Mone, Edward M. and Manuel London (2018): *Employee Engagement Through Effective Performance Management. A Practical Guide for Managers*, Routledge Taylor & Francis Group: New York and London, Second Edition.

Moulton, Brent R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units", *Review of Economics and Statistics*,72, pp. 334-338.

Murphy, Kevin R. (2008), "Explaining the Weak Relationship Between Job Performance and Ratings of Job Performance", *Industrial and Organizational Psychology*, 1, pp. 148-160.

O'Reilly III, Charles A. David F. Caldwell and William P. Barnett (1989), "Work Group Demography, Social Integration, and Turnover", *Administrative Science Quarterly*, 34 (1), pp. 21-37.

Pelled, Lisa Hope, Katherine R. Xin and Allen M. Weiss (2001), "No es como mí: Relational demography and conflict in a Mexican production facility", *Journal of Occupational and Organizational Psychology*, 74 (1), pp. 63-84.

Prendergast, Canice and Robert H. Topel (1996), "Favoritism in Organizations", *Journal of Political Economy*, 104 (5), pp. 958-978.

Rose, Dale S. and Jesse C. Biringer (2020): *Current Practices in 360 Feedback*, 3D Group: Emeryville, 6th Edition.

Schneid, Matthias, Rodrigo Isidor, Holger Steinmetz and Rüdiger Kabst (2016), "Age diversity and team outcomes: a quantitative review", *Journal of Managerial Psychology*, 31 (1), pp. 2-17.

Schüller, David, Harald Tauchmann, Thorsten Upmann, and Daniel Weimar (2014), "Pro-social behavior in the TV show "Come Dine With Me": An empirical investigation", *Journal of Economic Psychology*, 45, pp. 44-55.

Scullen, Steven E., Michael K. Mount and Maynard Goff (2000), "Understanding the Latent Structure of Job Performance Ratings", *Journal of Applied Psychology*, 85 (6), pp. 956-970.

Simons, Tony, Lisa Hope Pelled and Ken A. Smith (1999), "Making Use of Difference: Diversity, Debate, and Decision Comprehensiveness in Top Management Teams", *Academy of Management Journal*, 42 (6), pp. 662-673.

Tajfel, Henri, M. G. Billig and R. P. Bundy and Claude Flament (1971), "Social categorization and intergroup behaviour", *European Journal of Social Psychology*, 1 (2), pp. 149-178.

Tajfel, Henri and John C. Turner (1986): The Social Identity Theory of Intergroup Behavior, in S. Wochel and W.G. Austin (ed.), *Psychology of intergroup relations*, Nelson-Hall: Chicago, 2nd Edition, pp. 7-24.

Timmerman, Thomas A. (2000), "Racial Diversity, Age Diversity, Interdependence, and Team Performance", *Small Group Research*, 31 (5), pp. 592-606.

Tsui, Anne S. Terri D. Egan and Charles A. O'Reilly III (1992), "Being Different: Relational Demography and Organizational Attachment", *Administrative Science Quarterly*, 37 (4), pp. 549-579.

Turner, John C. (1987): *Rediscovering the* social *group: A social categorization theory*, Blackwell: Oxford.

Van Der Vegt, Gerben S. and J. Stuart Bunderson (2005), "Learning and Performance in Multidisciplinary Teams: The Importance of Collective Team Identification", *Academy of Management Journal*, 48 (3), pp. 532-547.

van Dijk, Hans, Marloes L. van Engen and Daan van Knippenberg (2012), "Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance", *Organizational Behavior and Human Decision Processes*, 119 (1), pp. 38-53.

Wagner, W. Gary, Jeffrey Pfeffer and Charles A. O'Reilly III (1984), "Organizational Demography and Turnover in Top-Management Groups", *Administrative Science Quarterly*, 29 (1), pp. 74-92.

Wegge, Jürgen, Carla Roth, Barbara Neubach, Klaus-Helmut Schmidt and Ruth Kanfer (2008), "Age and Gender Diversity as Determinants of Performance and Health in a Public Organization: The Role of Task Complexity and Group Size", *Journal of Applied Psychology*, 93 (6), pp. 1301-1313.

Wegge, Jürgen and Klaus-Helmut Schmidt (2009): The impact of age diversity in teams on group performance, innovation and health, in: A.-S. G. Antoniou, C. L. Cooper, G. P. Chrousos, C. D. Spielberger and M. W. Eysenck (ed.), *New horizons in management. Handbook of managerial behavior and occupational health*, Edward Elgar Publishing: Cheltenham, pp. 79-94.

West, Michael, Malcom Patterson, Jeremy Dawson and Steve Nickell (1999), *The Effectiveness of Top Management Groups in Manufacturing Organisations*, Centre for Economic Performance, November.

Wiersema, Margarethe F. and Karen A. Bantel (1992), "Top Management Team Demography and Corporate Strategic Change", *Academy of Management Journal*, 35 (1), pp. 91-121.

Wiersema, Margarethe F. and Allan Bird (1993), "Organizational Demography in Japanese Firms: Group Heterogeneity, Individual Dissimilarity, and Top Management Team Turnover", *Academy of Management Journal*, 36 (5), pp. 996-1025.

Williams, Katherine Y. and Charles A. O'Reilly III (1998), "Demography and Diversity in Organizations: A Review of 40 Years of Research", *Research in Organizational Behavior*, 20, pp. 77-140.

Zajac, Edward J., Brian R. Golden and Stephen M. Shortell (1991), "New Organizational Forms for Enhancing Innovation: The Case of Internal Corporate Joint Ventures", *Management Science*, 37 (2), pp. 170-184.

Zenger, Todd R. and Barbara S. Lawrence (1989), "Organizational Demography: The Differential Effects of Age and Tenure Distributions on Technical Communication", *Academy of Management Journal*, 32 (2), pp. 353-376.

Appendix

Table A1: Sensitivity checks of age diversity on peer and superior evaluation

	Peer evaluation		Superior evaluation
	Cooking show	Shopping show	Shopping show
	(1)	(2)	(3)
Age diversity defined as the standard deviation of the age by group			
Model specification …			
Main results	-0.021 ***	-0.026 ***	-0.024 *
(taken from Table 2)	(0.004)	(0.006)	(0.012)
Ordinary Least Squares	-0.021 ***	-0.026 *	-0.024
(clustered standard errors)	(0.006)	(0.014)	(0.016)
Ordinary Least Squares	-0.021 ***	-0.026 *	-0.024
(bootstrapped standard errors)	(0.006)	(0.014)	(0.016)
Ordered Logit	-0.028 ***	-0.039 ***	-0.022
(robust standard errors)	(0.005)	(0.009)	(0.018)
Ordered Probit	-0.016 ***	-0.022 ***	-0.016
(robust standard errors)	(0.003)	(0.005)	(0.010)
Controling for the	-0.013 ***	-0.023 ***	-0.043 ***
age difference	(0.004)	(0.007)	(0.013)
Obervations	10,580	6,440	1,610

Notes: The table shows the results of the alternative OLS regressions which measure again the peer evaluation (from the cooking and the shopping data in columns (1) and (2), respectively) and the superior evaluation from the shopping show (third column) by different standard errors as well as a Logit and a Probit model. The last row adds the absolute age difference as a further control variable. Statistical significance: $p<0.1$ *, $p<0.05$ **, $p<0.01$ ***.

Table A2: Sensitivity checks of age diversity on the performance score

	Performance Score of group members who are previously …	
	unknown	**known**
	(1)	**(2)**
Age diversity defined as the standard deviation of the age by group		
Model specification …		
Main results(taken from Table 4)	-0.171 ***	0.011
	(0.050)	(0.027)
Ordinary Least Squares(clustered standard errors)	-0.171 ***	0.011
	(0.057)	(0.030)
Ordinary Least Squares(bootstrapped standard errors)	-0.171 ***	0.011
	(0.059)	(0.031)
Ordered Logit(robust standard errors)	-0.096 ***	0.001
	(0.027)	(0.015)
Ordered Probit(robust standard errors)	-0.048 ***	0.003
	(0.015)	(0.009)
Controling for ability (handicap) and experience (years on tour)	-0.181 ***	0.031
	(0.048)	(0.027)
First Round Effects	-0.231 ***	0.015
	(0.070)	(0.034)

Notes: The table presents the effect of age diversity on the PGA Golf Tour 2002 performance score for previously unknown and known group members, respectively. The performance score is the negative number of strokes per round. Row 1 for column (1) and (2) shows the baseline specifications as specified in equation (2). The other rows modify the model through alternative standard errors, ordered Logit and Probit regressions or expand the baseline regression by ability and experience fixed effects. The last row shows short-term effects by considering only the first round. The observations are always 462 (2,514) for unknown (known) group members except for the effects of the first round (240 and 1,299 observations, respectively). Statistical significance: $p<0.1$ *, $p<0.05$ **, $p<0.01$ ***.