

Hett, Florian

Conference Paper

Dynamic Contests with Feedback and Heterogeneity: Empirical Evidence from a Gaming App

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik
- Session: Risk and Ambiguity I, No. D06-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Hett, Florian (2014) : Dynamic Contests with Feedback and Heterogeneity: Empirical Evidence from a Gaming App, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Risk and Ambiguity I, No. D06-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/100494>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Dynamic Contests with Feedback and Heterogeneity: Empirical Evidence from a Gaming App*

Florian Hett[†]

February 2014

Abstract

This paper empirically analyzes how performance feedback and information on heterogeneity affect behavior in dynamic contests, using data on two-player-contests from a smartphone/tablet application called “Wordblitz for Friends”. We find that players increase output as underdogs and decrease output as favorites, where being an underdog (favorite) refers to being matched against an opponent who was more (less) successful in previous contests. Performance feedback, i. e. information about past performance within one particular contest, induces players to reduce output when being ahead and increase output when being behind. The strength of these effects increases in output differences. An analysis of the interaction of information on heterogeneity and performance feedback reveals that the output reducing effect of being ahead is entirely driven by favorites, as players generally increase their output when being an underdog, independent of being ahead or behind.

Keywords: Dynamic Contests, heterogeneity

JEL Classification Codes:

*I thank Mark Bernard, Philipp Denter, Markus Kroell, Mario Mechtel, Felix Schmidt, Isabel Schnabel, Daniel Schunk and Jakob Schwab for helpful comments and discussion and Jens Abke and the team of Lotum for providing the data.

[†]GSEFM, University of Mainz and Goethe University Frankfurt, Germany, e-mail: research@flo-hett.de.

1 Introduction

Contests, i. e. situations where participants spend effort to compete for a given prize, are ubiquitous in today’s institutions. In the educational system, for example, the admission process to schools and universities as well as the common practice of “grading on a curve” can be understood as contests whose features impact learning outcomes of students. Hiring and promotion decisions in organizational settings as well as success in innovation and entrepreneurship are also often determined by contest structures. A thorough understanding of the implications of different contest designs is therefore crucial to assess the consequences of different institutional arrangements in many areas of public policy.

While there is already a well-developed economic literature on static contests¹, many real world settings are characterized by dynamic structures, i. e. scenarios where decisions are made repeatedly and may depend on previous actions and information. For example, promotion decisions in organizations typically depend on behavior over a longer period of time during which an employee receives feedback on his performance. However, taking these dynamic aspects into account induces a large host of potential theoretical mechanisms whose interactions undermine the predictive power of models in this class. This problem is even more severe once the heterogeneity among participants is taken into account. Given this lack of robust theoretical results, empirical evidence is required to improve our understanding of the consequences of contest schemes actually reflected in existing institutional settings.

This paper provides such evidence. We exploit a data set on two-player contests from a smartphone/tablet application called “Wordblitz for Friends” to empirically investigate the impact of information regarding the heterogeneity of participants (i. e. differences in perceived ability) and performance feedback on the behavior in contests. First, we study if and how behavior depends on information about the opponent’s success in previous contests (information on heterogeneity). Second, we study if and how behavior *within* the contest depends on the relative performance in previous rounds in the same contest (dynamic effects). Finally, we study the interaction of these two mechanisms, i. e. whether the reaction to being behind or ahead in a particular contest depends on whether one entered this contest as a favorite or an underdog.

Disentangling the different mechanisms that are potentially driving behavior in dynamic contests between heterogeneous participants is empirically challenging for the following reasons: First, it is likely that dynamic patterns in performance in repeated tasks exist even in the absence of contest structures. Such autocorrelation may then be hard to distinguish from the actual dynamics induced by the contest. Second, previous research has shown

¹See Tullock (1980); Lazear and Rosen (1981); Baye et al. (1996); Moldovanu and Sela (2001) for important contributions and Konrad (2009) for a recent textbook treatment.

that *individual* behavior in contests is extremely heterogeneous and that the sources of this heterogeneity are mostly unknown (Dechenaux et al., 2012). As a result, linking the heterogeneity in characteristics (like ability) to systematic heterogeneity in behavior is hindered by the huge variation in behavior due to other mechanisms and requires well-structured and/or large data sets. Third, identifying the effect of heterogeneity in previous success is typically hampered by the fact that having been more or less successful than an opponent in previous contests is likely to be related to ability and other individual-specific characteristics and thus creates endogeneity problems.

The data at hand is particularly suitable to address these issues due to the following specific features: (i) Players receive information on their own and their opponent’s previous performance prior to the game and (ii) each game lasts for three rounds including intermediate feedback on the performance in previous rounds. These two features introduce heterogeneity and dynamics in the setup and therefore allow to address their effects on behavior. (iii) The game takes place in a controlled environment - the application user surface. While it is therefore still a “natural setting” as compared to a laboratory, it maintains a high level of control. (iv) Players play the game repeatedly against different players, allowing to control for between-player differences in characteristics like ability by exploiting within-player variation. Finally, (v) the matching of opponents is exogenous conditional on a player’s level which allows for a clean identification of the effect of heterogeneity among opponents on behavior.

This chapter contributes to the literature in two ways. First, it allows to test some of the behavioral predictions that the still scarce theoretical literature on dynamic contests (Denter and Sisak, 2013; Ederer, 2010) provides, thereby potentially guiding the further theoretical development in this area. Second, it adds to the growing empirical literature on the effects of contest incentives. Feedback in dynamic contests (Gill and Prowse, 2012; Delfgaauw et al., 2012; Casas-Arce and Martinez-Jerez, 2009; Barankay, 2012a,b) as well as heterogeneity of players (Dechenaux et al., 2012) have already been extensively studied separately in the literature.² However, the empirical evidence on how these two effects interact is missing so far. In addition, while there are studies looking at the effect of feedback in field settings, the analysis of heterogeneity among players is so far typically restricted to studies in the lab.

Our results can be summarized as follows. First, looking at the entire contest (total output over three rounds) and the overall effect of information with respect to heterogeneity, we find that players have a higher output when entering the contest as an underdog compared to being a favorite. Furthermore, we document that this reallocation of output between games is harmful to winning-rates. Second, we investigate the effect of dynamics, i. e. the

²See chapter ?? for a more extensive review of the theoretical and empirical literature on contests.

reaction of a player to being behind or ahead in the second or third round. Here the results show that players adjust output in accordance to relative standing as they find more words when being behind while finding fewer words when being ahead. In addition, both of these effects become stronger the larger the respective lead or trail. Third, looking at the interaction of the two previous aspects, we find that the output-reducing effect of being ahead is only driven by favorites, as underdogs increase their output even in a leading position.

This chapter is structured as follows: First, section 2 describes the setting, including a description of the main variables. Section 3 then describes how this setting can be captured theoretically and describes mechanisms from the existing theoretical literature related to the setting at hand. Section 4 describes the empirical strategy and presents the results. Finally, section 5 discusses how the empirical results relate to existing theory and hints at potential avenues for future research.

2 Setup and Data

2.1 Environment: Wordblitz for Friends

The data used in this study stems from a game on cell phone or tablet applications and is called “Wordblitz for Friends”. Given an array of letters, the aim of the game is to find as many words as possible in a given time period (2 minutes), where the letters of each word have to be connected on the array (see figure 1 for an impression of the layout of the game). Depending on its length and the type of letter used, one gains points for each word. One match consists of three rounds including three different arrays of letters and two minutes time each.

Crucially for the analysis in this chapter, each match is played against an opponent who faces the same three arrays of letters and has the same task. After these three rounds, the player who accumulated more points over the entire match is declared to be the winner and gains a “star” while the other player loses a star. Every time a player has gained five stars his user level increases by one.

During the course of one match, players receive several pieces of information. Prior to the first round, each player is informed about the username and userlevel of the opponent he has been matched against. In addition, after each round he receives a list of words found, the number of words and the amount of points gained in the previous round, as well as the points accumulated over all finished rounds separately for himself and his opponent. Finally, after the three rounds the outcome of the overall match is presented as well as the gain or loss of a star and potentially the userlevel, respectively.

For each match, there are two possibilities to be matched with an opponent. First, via a connection to Facebook, one can deliberately challenge a friend. This match is then carried out independently of the level difference between the two contestants. Second, one can rely on the random matching mechanism the application provides. In this case, one is randomly matched with another contestant whose level is as close as possible to the own level, but the difference must not be greater than 14 levels.³ Importantly, this implies that whether one has a higher or a lower level than the opponent is random and entirely exogenous conditional on the level of the user, given that the user chooses the second matching mechanism.

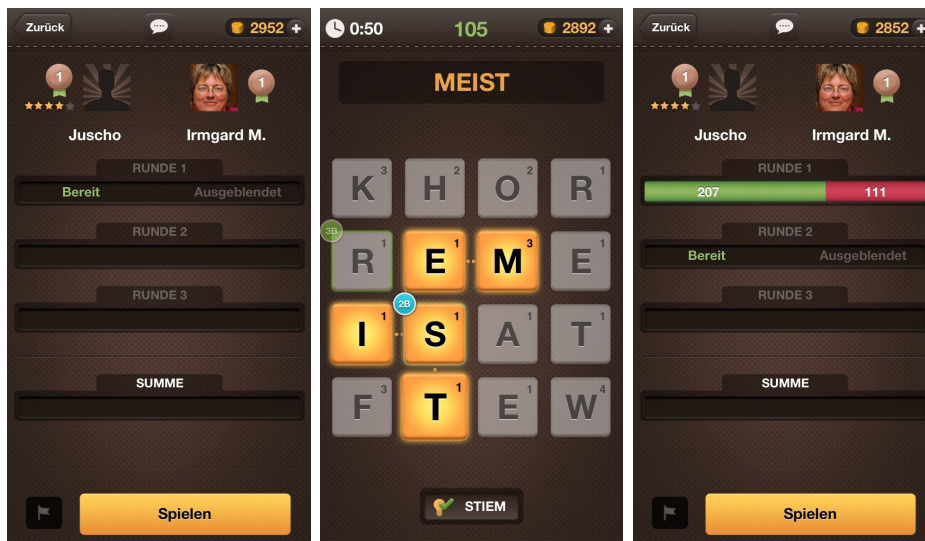


Figure 1: In-game screenshots

2.2 Data

As already noted in the introduction, the possibility of disentangling the effects of heterogeneity and dynamics requires specific data. In accordance with the description of the environment above, the data in this study fit this requirement as they have the following structure. Each observation can be classified with respect to three dimensions: The first player i in round r of game g , where rounds are nested within games and each game shows up for two players. Overall, there are $i = 1, \dots, 94637$ different players carrying out $g = 1, \dots, 200000$ games with $r = 1, 2, 3$ rounds each. Overall, this results in 200000 game observations, with 400000 player-game observations and 1200000 player-round observations. The average number of games per player is 4.22.

³Note that once a player is matched with another player he does not have the option to decline the contest, ruling out selection issues in this regard.

We restrict the data being used in the following way. First, we only consider observations of players for which we have information of at least 12 games in the sample. This reduces the sample size by 305,166 player-round observations. Second, we drop all games that have been arranged via Facebook. These games seem to be special in a number of dimensions, most importantly with respect to the interpretation of information about heterogeneity as provided by the difference in user levels. This eliminates another 29,064 observations from the sample. Finally, we keep only observations in which a player’s user level is between 5 and 30. This adjustment is done as being at the ends of the distribution might result in unbalanced matching simply due to the unavailability of players with respective levels. Additionally, the lower bound is motivated by the notion of unexperienced players still having to learn the game, which might introduce a lot of noise. Thereby, another 247,710 observations are dropped. In total, this leaves 618,600 player-round observations. Table 1 present descriptive statistics on the remaining sample.

$Words_{i,g,r}$ is the main variable of interest, used throughout the chapter in various forms and transformations. It measures the number of words a player has entered correctly. In the analysis, this variable is interpreted as productivity and is used to determine the relative output of the two contest participants and the winner. Note that words do not directly enter the contest success function, i. e. the relative number of words does not perfectly determine the winner of the game. Rather, each word provides a certain number of points, based on its length and letters used. In principle, this gives rise to an “intensive margin” of output by finding longer words in addition to the “extensive margin” of simply finding more words. However, due to data availability issues, in this study we only consider the extensive margin.⁴ Table 1 shows descriptive statistics on this variable. On average, a player finds about 36 words per round, with the maximum taking on the value of 194. The variation of output across observations is substantial as measured by the standard deviation of approximately 9 words. Also the within-player variation is considerable as indicated by the within-standard deviation of 4.95. This facilitates basing the analysis on within-player variation in order to control for unobserved heterogeneity.

$Userlevel_{i,g}$ is the second main variable. As previously described, the user level is determined by the number of stars gathered and therefore a function of past wins and losses. Hence, the user level is an informative signal about past performance. In the analysis, the variable will show up in two respects. First, it will serve as a control variable: Even in the case where only within-player variation is exploited, user level should be controlled for as it is probably an important determinant even of within-player output and thus valuable to

⁴Theoretically, in this setting shifts between the relative importance of the intensive and extensive margin could be related to risk aversion in an optimal stopping problem with respect to word search.

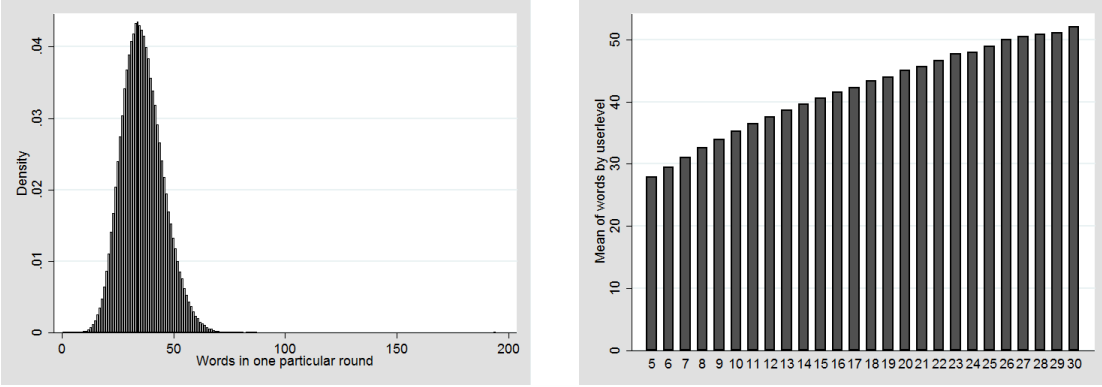


Figure 2: Distribution of $words_{i,g,r}$ overall and by $userlevel_{i,g}$,

Variable	Obs	Mean	Std. Dev.	Min	Max
$words_{i,g,r}$	618,600	36.179	9.326	1	194
$userlevel_{i,g}$	618,600	11.534	5.584	5	30
$leveldifference_{i,g}$	618,600	0.178	2.713	-10	8
$leveldifference_g$	618,600	2.236	1.545	0	10
$favorite_{i,g}$	618,600	0.476	0.499	0	1
$underdog_{i,g}$	618,600	0.395	0.489	0	1
$largefavorite_{i,g}$	618,600	0.105	0.306	0	1
$largeunderdog_{i,g}$	618,600	0.101	0.301	0	1
$words_demeaned_{i,g,r}$	618,600	0.027	6.088	-67	113
$position_{i,g,r}$	412,400	0.567	12.899	-84	89
$ahead_{i,g,r}$	618,600	0.668	0.471	0	1
$behind_{i,g,r}$	618,600	0.309	0.462	0	1

Table 1: Descriptive statistics: Selected sample

reduce the residual variance and increase estimation efficiency. Second, conditional on one player's user level, the *difference* in user levels of opponents is exogenous. This allows to treat the difference in user levels as an exogenous variation of the heterogeneity of players. Table 1 shows the descriptive statistics of user level and of the difference in user levels. The mean level is 11.5 with a maximum level of 30, while the average difference in user level is approximately 2.2 levels.

3 Mechanisms and Theoretical Predictions

In order to apply existing theoretical contributions to the setting studied in this chapter, we first classify the environment at hand in a more abstract and theoretically tractable way. The contest as described above is dynamic in the sense of consisting of repeated

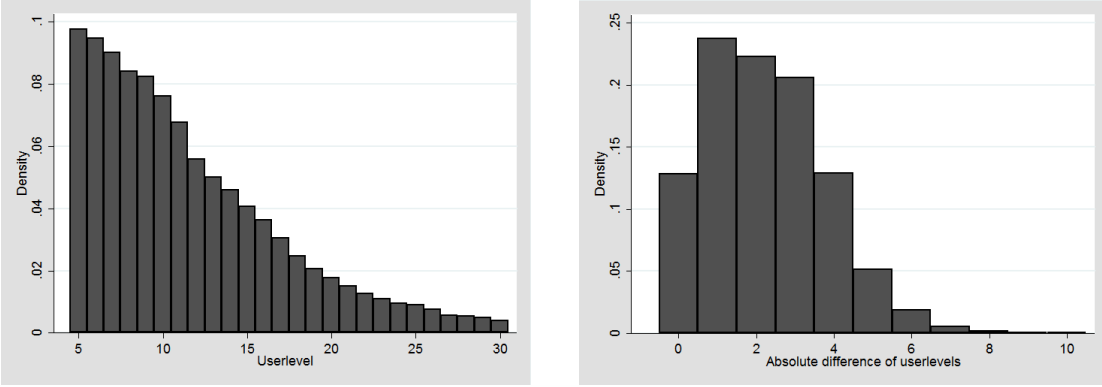


Figure 3: Distribution of $userlevel_{i,g}$, and $leveldifference_{i,g}$,

simultaneous actions (choice of effort), where after each action participants receive feedback on the consequences of previous actions of both players in form of output levels. This feedback then introduces the possibility of conditioning behavior on information that depends on previous actions, making the problem dynamic. Second, the evaluation of the contest only takes place after the final round and depends on the sum of individual output aggregated over all three rounds. This is different from another class of dynamic contests, in which the overall winner of the contest depends on the number of intermediate static contests a participant wins.⁵ This distinction is important because it implies that dynamic incentives arise not due to the nature of the contest success function but rather due to the intermediate revelation of information and its potential strategic implications. To see this, note that in the setting here output in all rounds are perfect substitutes with respect to the contest success function, which is not necessarily the case in other types of dynamic contests. Finally, the fact that information on the opponents’ user levels is given prior to the contest can theoretically be interpreted in at least two ways. First, if one allows for asymmetry with respect to ability or productivity of participants, information about success in previous contests represents an informative, albeit noisy signal about its respective level. This allows a participant to condition his initial behavior in the first round as well as his conditional behavior in later rounds on his belief about relative abilities based on the received signal(s). Second, even in the absence of asymmetries in ability, information on past performance might induce a reference point for the winning probability which in turn might lead to different behavior within the contest.

Given this abstract description of the setting, we now relate this to existing theoretical

⁵Obvious examples for the latter structure are playoffs in US sports, where teams move on another round once they won a particular number of games (“Best-Of-Seven” for instance). In contrast, the structure analyzed here rather illustrates the dynamic contest nature within a game, where each move represents a new round.

studies on dynamic contests to derive the potential mechanisms that matter in the empirical analysis.

Denter and Sisak (2013) discuss dynamic contests in the context of political campaign spending as a means to win elections and explicitly analyze the strategic effects of being ahead or behind. Their setting is similar to the one described above in so far as output in all rounds are perfect substitutes with respect to the contest success function. A key difference is that players are otherwise symmetric, i. e. no differences in productivity or effort costs exist. One key result is the existence of “momentum”, which refers to the positive effect of being ahead on subsequent output in an intermediate round. The intuition for this result follows a strategic reasoning: One basic result in the theory of static contests is that equilibrium effort levels are higher the more equal the contest is. Given that the prize of the contest is fixed this implies that, *ceteris paribus*, players prefer unequal contests to equal ones as this implies lower effort costs in equilibrium. Note that the final round of a dynamic contest can always be interpreted as a static contest whose asymmetry is determined by previous behavior. Hence, players should consider how their actions in previous rounds affect the characteristics of the “static” last round contest and try to make it more unequal in expected terms in order to reduce expected effort costs. This logic then implies an explanation for “momentum”: If a player is ahead, the asymmetry of the contest in the last round increases in individual effort, while it decreases in individual effort if the player is behind. Momentum would therefore predict a positive effect of the intermediate position in the contest on output. However, given that this reasoning only applies to intermediate rounds, one should not find this effect in the last round of the contest.

Similarly, Ederer (2010) also analyzes a setup where output levels in both rounds are perfect substitutes and looks at the effect of feedback after the first round. The so-called “evaluation effect” of feedback resembles the already described result from static settings that less balanced standings lead to lower effort levels. However, the model also allows for asymmetric abilities that are unobservable to contest participants and assumes complementarity of effort and ability.⁶ This induces additional effects on optimal effort levels. First, the “motivation effect” refers to the informational content of feedback with respect to the individual’s as well as the opponent’s unobservable ability. Revealed differences in output represent a noisy signal of relative ability. In turn, this affects optimal subsequent effort choices due to the complementarity of effort and ability. As a result, higher relative output increases subsequent effort as it indicates a higher relative productivity of effort and vice versa. Second, the “implicit incentives” effect describes the strategic incentives the motivation effect implies. Given that a higher perceived ability difference leads an opponent to

⁶So ability can be interpreted as productivity.

reduced subsequent effort, each player wants to appear as able as possible. This induces a “signal jamming” effect, i. e. an increase in effort in early rounds. Taken together, the two effects imply conflicting predictions regarding the setup analyzed here: First, effort should increase in output differences as well as in the difference of user levels (motivation effect). Second, effort should decrease in the absolute value of differences of user levels and output (evaluation effect).

Gill and Prowse (2012) look at the effect of disappointment aversion in sequential contests in a theoretical and experimental setup. While the previous two papers highlight potential strategic incentives, i. e. they explicitly consider the effect of behavior on the incentives of the opponent, Gill and Prowse, in contrast, rule out these effects by experimental design and thereby focus on the individual decision perspective. They do so in order to identify whether behavior in dynamic contests is affected by “disappointment aversion”, which results from expectation-based reference-dependent preferences with loss aversion. The intuition behind this is that participants evaluate outcomes relative to the outcome they expected (expectation-based reference dependence) and that positive deviations from this expectation produce a smaller utility increase than the utility lost due an equally sized negative deviation from this expectation (loss aversion). Therefore, individuals again want to avoid balanced contests as they imply larger uncertainty and thus larger expected utility losses due to the loss aversion effect. As a result, players increase effort when in front and decrease effort when being behind.

4 Empirical analysis

In a first step, we analyze the effect of ex-ante heterogeneity on final outcomes of individual contests after all three rounds. We test whether and how individuals differ in behavior when they enter a contest as a favorite or an underdog compared to being matched with a player of an equal level. At this point, we ignore the actual mechanisms through which these differences materialize. We will also look at whether these adjustments of behavior to information on heterogeneity increase the probability of winning the contest, i. e. whether players react “optimally” to the information provided. In a second step, we then disentangle the overall effect and isolate the effects of heterogeneity and dynamics as well as their interaction separately.

4.1 The Overall Effect of Heterogeneity

In a typical data set, contest participants who are informed to be the favorite in the upcoming contest most likely differ systematically from those who are informed to be the underdog with

respect to important characteristics, most obviously their ability. Therefore, identifying the behavioral effect of information about the heterogeneity of previous success (which might be interpreted as ability) typically suffers from the problem of disentangling the general heterogeneity of behavior from the effect of information about heterogeneity. However, the data at hand are particularly suited to solve this problem due to the combination of its panel structure and the matching procedure: The fact that there are multiple games for many participants in the data allows to use within- instead of between-subject variation in order to estimate the effect, thereby controlling for all time-invariant individual-specific characteristics like ability. However, this might still lead to a biased estimate as players are more likely to enter a contest as a favorite the better they become, such that even the within-player variation in ability becomes correlated with the probability of being a favorite or an underdog. In this setting, the nature of the matching procedure overcomes this concern, as it randomly assigns opponents within a symmetric range around the own user level. Nevertheless, there might still be spurious effects due to asymmetries in the distribution of user levels. To address this issue, all estimations will also include the user level of a player as an explanatory variable, such that the identifying assumption is only that, conditional on the user level, being a favorite or an underdog is randomly assigned, which is a technical fact of the application itself.

The overall effect of heterogeneity is estimated using the following baseline equation:

$$totalwords_{i,g} = \alpha_i + \beta_1 userlevel_{i,g} + \beta_2 favorite_{i,g} + \beta_3 underdog_{i,g} + \epsilon_{i,g}. \quad (1)$$

$totalwords_{i,g}$ is the overall number of correct words player i achieved in game g over all three rounds. α_i is a player fixed effect, absorbing the average number of correct words by player i over all his games, thereby reducing the considered variation to be only within-player. $userlevel_{i,g}$ is the level user i has in the current game g , and $favorite_{i,g}$ ($underdog_{i,g}$) is a dummy variable, denoted 1 if the user level of player i is higher (lower) than the user level of his opponent and zero otherwise. Equation 1 is estimated using a fixed effects OLS regression, clustering standard errors on the player-level.

We subsequently adjust the equation in the following ways: First, we will add two additional dummy variables $largefavorite_{i,g}$ and $largeunderdog_{i,g}$, which take on the value of 1 if the differences in user levels is larger than 3, or smaller than -3, respectively, which corresponds to one standard deviation of the user level difference. Second, instead of looking at dummy variables to capture variation in differences in user levels, we use the variable $leveldifference_{i,g}$, which is the difference in user levels between player i and his opponent.

Table 2 shows the respective results. In column (2) we see a statistically highly significant

	(1)	(2)	(3)	(4)	(5)	(6)
	$totalwords_{i,g}$	$totalwords_{i,g}$	$totalwords_{i,g}$	$totalwords_{i,g}$	$totalwords_{i,g}$	$totalwords_{i,g}$
					if $underdog_{i,g} = 1$	if $favorite_{i,g} = 1$
$userlevel_{i,g}$	0.035 (0.040)	0.033 (0.040)	0.034 (0.040)	0.037 (0.040)	0.070 (0.053)	-0.048 (0.060)
$favorite_{i,g}$	-0.349*** (0.082)		-0.309*** (0.084)			
$largefavorite_{i,g}$			-0.178* (0.100)	-0.384*** (0.095)		
$underdog_{i,g}$	0.223*** (0.084)		0.145* (0.086)			
$largeunderdog_{i,g}$			0.304*** (0.098)	0.547*** (0.090)		
$leveldifference_{i,g}$		-0.114*** (0.010)			-0.124*** (0.035)	-0.154*** (0.034)
Constant	108.211*** (0.463)	108.176*** (0.459)	108.226*** (0.463)	108.091*** (0.458)	109.981*** (0.649)	107.719*** (0.668)
Observations	206,200	206,200	206,200	206,200	98,237	81,500
R^2	0.001	0.001	0.001	0.000	0.000	0.000
Number of players	23,632	23,632	23,632	23,632	21,631	21,806

Notes: This table shows fixed effects regressions of overall output ($totalwords_{i,g}$) on variables reflecting information regarding heterogeneity. Variable definitions are given in the text. Standard errors clustered at the player level in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 2: The overall effect of heterogeneity

decrease of output in the difference of user levels, so that players perform *ceteris paribus* the better the worse their relative level. Column (1) allows to break up this effect to check whether it is driven only by being a favorite or an underdog. It becomes apparent that there is a significant effect in both domains: Players reduce their output when being a favorite by approximately 0.35 words and increase their output when being an underdog by approximately 0.2 words. Column (3) and (4) now additionally test whether these effects are stronger when the gap in user levels is particularly large by considering the dummy variables $largefavorite_{i,g}$ and $largeunderdog_{i,g}$. While column (3) shows that the difference in the effect is weakly statistically significant, column (4) confirms that the overall effect for large favorites and underdogs remains highly statistically significant. Column (5) and (6) show how the effect of being a favorite or an underdog changes if the difference in user levels becomes larger. However, instead of including additional dummy variables, it makes use of the continuous variable $leveldifference_{i,g}$. In column (5) the sample is restricted to games in which player i is the favorite. Thus, the significant negative coefficient of $leveldifference_{i,g}$ indicates that the reduction of output due to being a favorite becomes stronger the wider the level gap between the participants is. Analogously, column (6) restricts the sample to games in which player i is an underdog. Here the negative coefficient indicates that the increase in output is the larger the larger the level difference between the players is.

To summarize, there is a clear and systematic relationship between information on heterogeneity of players and productivity: Players increase (decrease) their output when they

play against opponents with a higher (lower) user level and these effects become stronger the larger the difference in levels is. This result raises the natural question what the consequences of this behavior are with respect to the outcomes of the contest. On the one hand, given that favorites generally have higher baseline levels of output, this *intra-player reallocation of output between games* might resemble an optimal response in order to achieve a higher number of contests won overall. On the other hand, this readjustment could lead to lower winning rates, if the increase of the number of games lost as a favorite are not outweighed by the additional number of games won as an underdog. Note that these effects do not necessarily cancel out, as the point difference by which one wins a contest does not matter. Thus, if for example as a large underdog the basic gap in average performance is very large, then even a substantial increase in productivity might simply be lost, as it is not sufficient to overcome the gap. Equivalently, if a player reduces output as a favorite although the productivity gap is already rather small, then this small reduction of output might be enough to lose the contest. As a result, it is not clear whether the documented output adjustments of players conditional on relative user levels change their winning probabilities and if so to the better or the worse.

The setup at hand includes no cases in which information on heterogeneity is absent. Thus, a straightforward test of the benefits of output adjustment conditional on relative user levels is not possible. To nevertheless address this issue we proceed as follows: The systematic deviation of output depending on being a favorite or an underdog represents a specific form of *conditioning* output on specific characteristics of the game. Thus, a natural way to create a counterfactual is to look at output following the *unconditional* output choice, i. e. similar output for all games, no matter what relative user levels actually are. Whether output reallocation across games benefits winning rates can then be tested in the following way: Matching the hypothetical output levels that would follow from unconditional output with the actual output of the opponent allows to determine a counterfactual outcome of the game. Doing this for all games then permits to compare the number of contests won due to the actual output levels, following the conditional output choices, to the counterfactual number of contests won that would have followed from unconditional output choices. Beneficial output adjustments then correspond to a higher number of contests won in the second case and vice versa.

To implement this approach, we first define a game as won by player i if the total amount of words by player i exceeds those by his opponent.⁷ Second, we then determine the hypothetical outcome of the game by matching the average output of player i with the

⁷Note again that number of words only serve as a proxy for actual points achieved, as the latter differ with respect to the characteristics of words.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\Delta Wins$	$\Delta Wins$	$\Delta Wins$	$\Delta Wins$	$\Delta Wins$	$\Delta Wins$
<i>favorite_{i,g}</i>		-0.010*** (0.003)	-0.022*** (0.001)		-0.008*** (0.003)	-0.020*** (0.001)
<i>underdog_{i,g}</i>		0.013*** (0.003)	0.001 (0.001)		0.010*** (0.003)	-0.002 (0.002)
<i>largefavorite_{i,g}</i>				-0.028*** (0.003)	-0.008*** (0.003)	-0.008*** (0.003)
<i>largeunderdog_{i,g}</i>				0.012*** (0.003)	0.014*** (0.003)	0.014*** (0.003)
Constant	-0.011*** (0.001)	-0.012*** (0.003)			-0.012*** (0.003)	
Observations	206,200	206,200	206,200	206,200	206,200	206,200
\bar{R}^2	0.000	0.001	0.001	0.001	0.001	0.002

Notes: This table shows regressions of overall output ($\Delta Wins$) on variables reflecting information regarding heterogeneity. Variable definitions are given in the text. Standard errors clustered at the player level in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 3: Counterfactual analysis of winning rates

actual output of his opponent and define game g to be “counterfactually won” by player i if his average output exceeds the actual output of his opponent. Finally, we then generate the variable $\Delta Wins$, which is one if the game was actually won but not counterfactually won, 0 if the outcome in both cases is equal, and -1 if the game was actually lost, but counterfactually won.

Table 3 now presents a series of regressions using $\Delta Wins$ as a dependent variable. Column (1) simply shows that the average of $\Delta Wins$ is significantly different from zero and negative, meaning that players would increase the number of games won by not adjusting output at all to characteristics of their opponent. The point estimate of the coefficient implies that the probability of winning a game is 1.1 percentage points higher under the counterfactual scenario. Beyond this baseline effect, the question remains to what degree players might still benefit from these readjustments by at least increasing the probability of winning as an underdog. To look at this, column (2) looks at $\Delta Wins$ for games as underdog or favorite separately. It reveals that the increase of output as an underdog does not lead to a significant increase of the winning probability as compared to the counterfactual, while the probability of winning as a favorite reduces by 2.2 percentage points due to the conditionality of productivity. Finally, column (4) looks at the difference in the effect for games with large differences in user levels. These regressions reveal that the stronger adjustment of output in games with strong level differences also translate to stronger changes in winning probabilities. As a result, large favorites lose significantly more games due to their reduction of output, while for large underdogs the strong adjustment of output now even leads to a statistically significantly positive effect: The probability of winning reduces for

	(1)	(2)	(3)	(4)	(5)	(6)
	$words_{i,g,1}$	$words_{i,g,1}$	$words_{i,g,1}$	$words_{i,g,1}$	$words_{i,g,1}$	$words_{i,g,1}$
$userlevel_{i,g}$	0.039** (0.020)	0.039** (0.020)	0.037* (0.020)	0.038** (0.020)	0.044 (0.027)	0.024 (0.030)
$favorite_{i,g}$		-0.087* (0.046)	-0.080* (0.046)			
$largefavorite_{i,g}$			-0.017 (0.056)	-0.074 (0.054)		
$underdog_{i,g}$		0.096** (0.046)	0.050 (0.048)			
$largeunderdog_{i,g}$			0.184*** (0.054)	0.255*** (0.050)		
$leveldifference_{i,g}$	-0.037*** (0.006)				-0.020 (0.020)	-0.061*** (0.019)
Constant	36.004*** (0.226)	35.995*** (0.229)	36.014*** (0.229)	35.982*** (0.225)	36.642*** (0.337)	35.696*** (0.339)
Observations	206,200	206,200	206,200	206,200	98,237	81,500
R^2	0.000	0.000	0.000	0.000	0.000	0.000
Number of players	23,632	23,632	23,632	23,632	21,631	21,806

Notes: This table shows regressions of output in the first round ($words_{i,g,1}$) on variables reflecting information regarding heterogeneity. Variable definitions are given in the text. Standard errors clustered at the player level in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 4: The effect of heterogeneity in the first round

large favorites by 2.8 percentage points, while it increases by 1.2 percentage points for large underdogs.

Overall, the results on the total effect of information on heterogeneity are as follows: First, players increase their productivity as underdogs and decrease productivity as favorites. Second, this effect becomes stronger the larger the differences in user levels between participants are. Third, compared to the constructed counterfactual, players suffer from these adjustments by significantly decreasing the probability of winning. Fourth, the only positive effect of these adjustments materializes for large underdogs whose probability of winning significantly increases as compared to the counterfactual.

4.2 The effect of information on heterogeneity

The previous section only looked at the overall effects of heterogeneity on productivity, thereby ignoring the potential channels through which they emerge. As indicated in Section 3, there are several mechanisms that potentially explain different sorts of behavior in the respective setting. In the following we therefore try to disentangle the effect of information on heterogeneity from the effect of dynamics to finally also consider the interaction of the two effects.

Analyzing the effect of information on heterogeneity requires to look at situations in which effects due to feedback on past performance can be ruled out. Output in the first round of the game qualifies in this regard: Prior to round 1, a player only receives information about his and the opponent's user level. Given that no previous round has been played at this

point, there is no past performance on which feedback could be given and behavior could be conditioned on. However, as user levels are already revealed, systematic differences in output levels in the first round can unambiguously be attributed to the effect of playing as a favorite or an underdog. To allow for comparability with respect to the analysis of the overall effect of heterogeneity, the analysis in this section considers the same methodology as in Table 2, the only difference being that instead of using total words we here use the number of words in the first round as a dependent variable.

Table 4 shows the results. One can see that the effects in general go into the same direction as reported in Table 2: As favorites, players have a lower output, while as underdogs their output is higher. The strength of these effects again increases in the difference of user levels. However, in contrast to the previous results, not all of these effects are statistically significant. In line with this, one can also see that the size of these effects is considerably smaller. For example, the effect of being a favorite leads on average to a reduction of words in the first round of just 0.09 words, while overall, favorites' productivity reduces by 0.35 words.

As a result, one can conclude that the overall effects of information on heterogeneity already seem to be present in the first round in the absence of performance feedback. However, their size is too small to plausibly explain the entire effect, implying that also in later rounds and in the presence of performance feedback, information on heterogeneity matters. This raises the question to what extent information on heterogeneity interacts with the additional information on relative performance in later rounds.

4.3 The Effect of Performance Feedback

We now analyze the general effect of performance feedback. Doing so imposes several empirical challenges. To identify the effect of performance feedback on subsequent productivity requires to analyze the correlation between the current standing within the contest and the subsequently following output. However, it seems very likely that the variation of both these variables is to a large extent determined by the unobservable ability of the player. Thus, observing, for example, a positive correlation of performance feedback with subsequent productivity can simply be the result of the player being in general better in the game, i. e. having a higher ability or baseline productivity. A second concern regards the distribution of the idiosyncratic errors. Given that output levels in each round of one particular game typically resemble a repeated action, auto-correlation might occur, for example because of mean reversion. Given the dynamic nature of the analysis, this might induce a bias in the estimate. Another source of endogeneity is apparent if one considers output in the third round: Performance feedback is a function of output of both players in previous rounds.

However, if performance feedback actually influences output, then output of player j in round 2 depends on output of player i in round 1. As a result, the idiosyncratic error of output of player i in round 1 affects output of player j in round 2 through performance feedback. Output of player j in round 2 in turn again affects output of player i in round 3 through performance feedback, thus eventually inducing a correlation between the idiosyncratic errors in round 1 and 3. As a result, it seems unfeasible to simply assume errors to be i.i.d. within games. Put differently, to identify the effect of performance feedback requires to control for the general dynamics of repeated activities, which might lead to systematic output patterns even in the absence of performance feedback, as well as for endogeneity due to strategic interdependencies of output of both players.

We deal with these problems in the following way: First, to measure productivity in a particular round we use the demeaned number of words ($words_demeaned_{i,g,r}$) as dependent variable. This variable is constructed by subtracting the average number of words of player i over all rounds in all games from the number of words achieved in the round considered. Thereby, the dependent variable reflects the deviation of output from the individual-specific mean output. Second, we choose an instrumental variable approach in order to isolate “clean” variation of performance feedback. As an instrument for performance feedback in rounds 2 and 3, we use the output of the opponent in the first round. This variable satisfies both criteria required to hold in order to serve as an instrument: First, output of the opponent in a previous round obviously has an substantial impact on performance feedback a player receives in subsequent rounds. Second, the exclusion restriction holds as the only mechanism through which output of player j in round 1 affects output of player i in subsequent rounds is through its effect on performance feedback. The following equations formalize this intuition for round 2:

$$\begin{aligned}
words_demeaned_{i,g,2} &= f(score_{i,g,2}) + \epsilon_{i,g,2} \\
score_{i,g,2} &= g(words_{i,g,1}; words_{j,g,1}) \\
words_{i,g,1} &= f(score_{i,g,1}; ability_i) + \epsilon_{i,g,1} \\
words_{j,g,1} &= f(score_{j,g,1}; ability_j) + \epsilon_{j,g,1} \\
score_{i,g,1} &= score_{j,g,1} = 0 \\
words_demeaned_{i,g,2} &= f(g(f(score_{i,g,1}; ability_i) + \epsilon_{i,g,1}; words_{j,g,1})) + \epsilon_{i,g,2}
\end{aligned}$$

One can see that by using the number of words achieved by the opponent in the first round ($words_{j,g,1}$) as an instrument for the relative performance before the start of the second round ($score_{i,g,2}$) the effect of performance feedback on subsequent output is identified

	(1)	(2)	(3)	(4)	(5)
	<i>words_dem.i,g,r</i>	<i>words_dem.i,g,r</i>	<i>words_dem.i,g,r</i>	<i>words_dem.i,g,r</i>	<i>words_dem.i,g,r</i>
				if <i>ahead</i> _{<i>i,g,r</i>} = 1	if <i>behind</i> _{<i>i,g,r</i>} = 1
<i>position</i> _{<i>i,g,r</i>}	-0.009*** (0.001)			-0.028*** (0.005)	-0.020*** (0.004)
<i>userlevel</i> _{<i>i,g</i>}	-0.018*** (0.001)	-0.018*** (0.001)	-0.018*** (0.001)	-0.013*** (0.002)	-0.022*** (0.002)
<i>ahead</i> _{<i>i,g,r</i>}		-0.286*** (0.040)			
<i>behind</i> _{<i>i,g,r</i>}			0.285*** (0.040)		
Constant	0.104*** (0.016)	0.242*** (0.025)	-0.033 (0.025)	0.315*** (0.057)	-0.040 (0.050)
Observations	412,400	412,400	412,400	206,766	191,241
R ²	0.000	0.000	0.000	0.000	0.000

Notes: This table shows IV 2SLS regressions of demeaned output (*words_demeaned*_{*i,g,r*}) on variables reflecting performance feedback. *position*_{*i,g,r*}, *ahead*_{*i,g,r*} and *behind*_{*i,g,r*} are instrumented by *words*_{*j,g,1*}. Variable definitions are given in the text. Standard errors clustered at the player level in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 5: The effect of performance feedback: full sample

as long as $\epsilon_{i,g,2}$ has no effect on $\epsilon_{j,g,1}$.

The analysis proceeds as follows: The baseline specification is

$$words_demeaned_{i,g,r} = \alpha + \beta_1 position_{i,g,r} + \beta_2 + userlevel_{i,g} \epsilon_{i,g,r}, \quad (2)$$

where *position*_{*i,g,r*} is the difference in words player *i* and player *j* achieved in rounds prior to round *r* and thus a measure of performance feedback. The regression equation is estimated using 2SLS with *words*_{*j,g,1*} as an instrument for *position*_{*i,g,r*}. This specification allows to test for the average effect of performance feedback on subsequent performance. However, it implicitly assumes a monotonous relationship between these two variables and thus does not allow the effect of an increase in relative performance to be different for being ahead or behind. To test whether this is the case, as a second and third specification we use the variable *ahead*_{*i,g,r*} and *behind*_{*i,g,r*}, where either is again instrumented by *words*_{*j,g,1*} respectively. The former takes on the value of 1 if position is positive, i. e. player *i* is ahead in round *r* of game *g*, while the latter takes on the value of 1 if position is negative, i. e. player *i* is behind in round *r* of game *g*. Finally, to test whether the effect of being ahead or behind varies in the size of the difference in words after previous rounds, we run separate regressions for both cases, using the variable *position*_{*i,g,r*}, which is instrumented by *words*_{*j,g,1*}.

Table 5 presents the results. In column (1) one can see that the average effect of performance feedback on subsequent productivity is significantly negative, implying a stronger reduction of productivity the better relative performance turns out. Column (2) and (3) now disentangle the effect and test for the average effect of being ahead or behind, respectively.

	(1)	(2)	(3)	(4)	(5)
	$words_dem_{i,g,r}$	$words_dem_{i,g,r}$	$words_dem_{i,g,r}$	$words_dem_{i,g,r}$	$words_dem_{i,g,r}$
				if $ahead_{i,g,r} = 1$	if $behind_{i,g,r} = 1$
position	-0.016*** (0.002)			-0.038*** (0.008)	-0.037*** (0.007)
$userlevel_{i,g}$	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.002 (0.003)	-0.011*** (0.004)
ahead		-0.377*** (0.052)			
behind			0.376*** (0.051)		
Constant	0.083*** (0.026)	0.263*** (0.036)	-0.096*** (0.036)	0.286*** (0.073)	-0.120* (0.065)
Observations	206,200	206,200	206,200	102,106	94,946
R^2	0.000	0.000	0.000	0.000	0.000

Notes: This table shows IV 2SLS regressions of demeaned output ($words_demeaned_{i,g,r}$) on variables reflecting performance feedback. $position_{i,g,r}$, $ahead_{i,g,r}$ and $behind_{i,g,r}$ are instrumented by $words_{j,g,1}$. Variable definitions are given in the text. Standard errors clustered at the player level in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 6: The effect of performance feedback: round 2

It becomes apparent that when being ahead, players reduce their productivity on average by 0.29 words per round, while they increase their productivity by the same amount when being behind. The fact that the two effects are similar in absolute value is a consequence of there being almost no rounds in which previous performance for both players was exactly similar. Column (4) and (5) now look at how effect sizes change when the difference in previous performance increases. Column (4) thereby reveals that the reduction in output when being ahead becomes the stronger the larger the lead is, i. e. the higher the value of $position_{i,g,r}$. Similarly, column (5) shows the increase in output due to being behind becoming the stronger the larger the gap in previous performance. Tables 6 and 7 break down the results by rounds, with Table 6 referring only to observations from round 2, while Table 7 depicts the results for round 3. Although the size of the effects differs significantly, the overall magnitude and direction of the coefficients largely coincide. Regarding the relevance of the instrument, first stage results of 2SLS estimations show a substantial explanatory power of $words_{j,g,1}$ for $position_{i,g,r}$, as well as $ahead_{i,g,r}$ and $behind_{i,g,r}$.

To summarize, one can see that performance feedback leads players to adjust their behavior in the following way: When being ahead, players reduce productivity, while they increase productivity when being behind and both these effects become the stronger the larger the gap in previous performance is.

	(1)	(2)	(3)	(4)	(5)
	<i>words_dem.i.g,r</i>	<i>words_dem.i.g,r</i>	<i>words_dem.i.g,r</i>	<i>words_dem.i.g,r</i>	<i>words_dem.i.g,r</i>
				if <i>ahead</i> _{<i>i,g,r</i>} = 1	if <i>behind</i> _{<i>i,g,r</i>} = 1
position	-0.005*** (0.002)			-0.022*** (0.006)	-0.004 (0.005)
<i>userlevel</i> _{<i>i,g</i>}	-0.029*** (0.003)	-0.029*** (0.003)	-0.029*** (0.003)	-0.024*** (0.004)	-0.032*** (0.004)
ahead		-0.182*** (0.060)			
behind			0.182*** (0.059)		
Constant	0.125*** (0.029)	0.214*** (0.042)	0.037 (0.040)	0.345*** (0.084)	0.108 (0.074)
Observations	206,200	206,200	206,200	104,660	96,295
<i>R</i> ²	0.001	0.001	0.001	0.000	0.001

Notes: This table shows IV 2SLS regressions of demeaned output (*words_demeaned*_{*i,g,r*}) on variables reflecting performance feedback. *position*_{*i,g,r*}, *ahead*_{*i,g,r*} and *behind*_{*i,g,r*} are instrumented by *words*_{*j,g,1*}. Variable definitions are given in the text. Standard errors clustered at the player level in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 7: The effect of performance feedback: round 3

4.4 The Interaction of Information on Heterogeneity and Performance Feedback

After having discussed the effects of information about heterogeneity and performance feedback in isolation, we finally analyze the interaction of the two, i. e. how the strength of the effect of performance feedback differs with respect to differences in user levels. A straightforward way to look at this is by incorporating interaction terms of *ahead*_{*i,g,r*} with *favorite*_{*i,g*} and *underdog*_{*i,g*}.⁸ Doing so obviously raises similar concerns with respect to identification as in the case for the analysis of performance feedback: Given that *position*_{*i,g,r*} (and thus also *ahead*_{*i,g,r*} and *behind*_{*i,g,r*}) itself was prone to endogeneity and thus had to be instrumented, this issue also applies for its interactions with other variables. This therefore calls for additional instruments. Thus, we will add interactions of *favorite*_{*i,g*} and *underdog*_{*i,g*} with *words*_{*j,g,1*} to the first-stage regressions. This yields the following regression model:

$$\begin{aligned}
\text{words_demeaned}_{i,g,r} = & \alpha + \beta_1 \text{userlevel}_{i,g} \\
& + \beta_2 \text{favorite}_{i,g} + \beta_3 \text{underdog}_{i,g} + \beta_4 \text{ahead}_{i,g,r} \\
& + \beta_5 \text{ahead} * \text{underdog}_{i,g,r} + \beta_6 \text{ahead} * \text{favorite}_{i,g,r} \\
& + \epsilon_{i,g,r},
\end{aligned} \tag{3}$$

⁸Note that, as already mentioned previously, there are almost no rounds in which both players have the exact similar number of words and thus nobody is ahead or behind. Thus, being ahead or behind resemble inverse events and it is sufficient to include only one of the two respective variables, as one of them taking on the value of one almost always implies the other variable being zero and vice versa.

	(1)	(2)	(3)
	<i>words_demean</i> _{<i>i,g,r</i>}	<i>words_demean</i> _{<i>i,g,r</i>}	<i>words_demean</i> _{<i>i,g,r</i>}
<i>ahead</i> _{<i>i,g,r</i>}	-0.348** (0.153)	-0.162 (0.199)	-0.572** (0.235)
<i>ahead * favorite</i> _{<i>i,g,r</i>}	0.061 (0.176)	-0.254 (0.229)	0.438 (0.272)
<i>ahead * underdog</i> _{<i>i,g,r</i>}	0.286 (0.186)	-0.097 (0.236)	0.753*** (0.286)
<i>userlevel</i> _{<i>i,g</i>}	-0.017*** (0.001)	-0.006*** (0.002)	-0.028*** (0.003)
<i>favorite</i> _{<i>i,g</i>}	-0.118 (0.100)	0.099 (0.129)	-0.376** (0.154)
<i>underdog</i> _{<i>i,g</i>}	-0.097 (0.092)	0.111 (0.119)	-0.339** (0.141)
Constant	0.301*** (0.082)	0.149 (0.108)	0.473*** (0.128)
<i>Constant</i> (Player of same level being behind)	0.301*** (0.082)	0.149 (0.108)	0.473*** (0.128)
<i>Constant + ahead</i> _{<i>i,g,r</i>} (Player of same level being ahead)	-0.047 (0.081)	-0.013 (0.107)	-0.098 (0.122)
<i>Constant + underdog</i> _{<i>i,g</i>} (Underdog being behind)	0.204*** (0.037)	0.26*** (0.049)	0.135** (0.059)
<i>Constant + underdog</i> _{<i>i,g</i>} + <i>ahead</i> _{<i>i,g,r</i>} + <i>ahead * und</i> _{<i>i,g,r</i>} (Underdog being ahead)	0.142*** (0.061)	0.001 (0.078)	0.316*** (0.098)
<i>Constant + favorite</i> _{<i>i,g</i>} (Favorite being behind)	0.183*** (0.054)	0.248*** (0.07)	0.098 (0.087)
<i>Constant + favorite</i> _{<i>i,g</i>} + <i>ahead</i> _{<i>i,g,r</i>} + <i>ahead * fav</i> _{<i>i,g,r</i>} (Favorite being ahead)	-0.104*** (0.032)	-0.168*** (0.046)	-0.035 (0.05)
Observations	412,400	206,200	206,200
<i>R</i> ²	0.000	0.000	0.001

Notes: This table shows IV 2SLS regressions of demeaned output (*words_demeaned*_{*i,g,r*}) on variables reflecting performance feedback, information regarding heterogeneity and their interaction. *ahead*_{*i,g,r*}, *ahead * favorite*_{*i,g,r*} and *ahead * underdog*_{*i,g,r*} are instrumented by *favorite*_{*i,g*} * *words*_{*j,g,1*}, *underdog*_{*i,g*} * *words*_{*j,g,1*} and *words*_{*j,g,1*}. Variable definitions are given in the text. Standard errors clustered at the player level in parentheses. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 8: The interaction of information on heterogeneity and performance feedback

The model will be estimated by 2SLS with *ahead*_{*i,g,r*}, *ahead * underdog*_{*i,g,r*} and *ahead * favorite*_{*i,g,r*} being instrumented by the interactions of *favorite*_{*i,g*} and *underdog*_{*i,g*} with *words*_{*j,g,1*}.

Table 8 presents the results. The reference category to which the interpretation of the coefficient of the constant applies is a player whose opponent has the same user level and who is currently behind in the contest. In the bottom of Table 8 effect sizes and standard errors for the different dummy variable combinations are depicted. Column (1) refers to results based on the entire sample, i. e. both rounds 2 and 3, while column (2) shows results based only on observations in round 2 and column (3) in round 3. The dependent variable is again *words_demeaned*_{*i,g,r*} and can still be interpreted as the deviation of output from the individual-specific mean. The results show that when being behind players significantly increase their output, no matter whether they are playing against an opponent with a similar user level or they play as a favorite or an underdog. The insignificant coefficients of *favorite*_{*i,g*} and *underdog*_{*i,g*} as well as the absence of a statistically significant difference in their coefficients (p-value: 0.76) further imply that the effect of being behind also quantitatively does not depend on the player's relative user level. This result does not apply to cases

when a player is ahead in the contest. Here, favorites behave fundamentally different compared to underdogs, as they reduce output when being behind while underdogs in contrast increase output. This also implies that the overall reduction of output when being ahead documented in the previous section is entirely driven by favorites. The breakdown of the analysis distinguishing round 2 and round 3 behavior is in line with the aggregated results, as all signs of the respective effects point into the same direction, although not all of them are statistically significant.

5 Conclusion

In this chapter, we analyzed how performance feedback and information on heterogeneity of participants affect behavior in dynamic contests. A summary of the results reads as follows: First, over the entire course of a game, players increase output as underdogs and decrease output as favorites, which negatively affects winning rates as compared to the counterfactual of not conditioning output on opponent characteristics. The effect increases in the difference of opponents' user levels. Second, this effect is already present in the absence of performance feedback, i. e. in the first round in which the contest is still equal. However, first round differences are smaller than overall differences in output, so that they do not make up for the entire output gap between games as a favorite and an underdog. Third, performance feedback leads players to reduce output when being ahead and increase output when being behind. The strength of these effects increases in output differences. Finally, an analysis of the interaction of information on heterogeneity and performance feedback reveals that the output-reducing effect of being ahead is entirely driven by favorites, as players generally increase their output when being an underdog, independent of being ahead or behind.

Neither of the mechanisms described in Section 3 can fully rationalize these systematic patterns in the data. First, the intuition behind the momentum effect, i. e. the notion that players want to make contests as unequal as possible in order to commit to lower effort levels in later stages, would predict the opposite effect of performance feedback: Players being ahead should increase output, while those behind should reduce their output. Similarly, although this is not explicitly modelled, one would expect a similar effect with respect to information on ex ante heterogeneity, i. e. that players increase output as favorites and reduce output as underdogs. Second, the “evaluation effect” as described by Ederer (2010), i. e. that output should decrease in the absolute difference of relative past performance, is not entirely confirmed in the data: While the fact that favorites reduce output the stronger the larger they lead is consistent with this mechanism, players being behind increasing their output the larger the gap becomes is in stark contrast to what this theory predicts. Third, the

“motivation effect”, i.e. the idea that positive performance feedback resembles a positive signal about relative ability and thus induces an increase of output, cannot be confirmed.⁹ Finally, the implications of disappointment aversion as modelled by Gill and Prowse (2012) are rejected by the data as well.

This inconsistency of the empirical evidence with existing theoretical predictions from the literature has several potential explanations with different implications. First, one might wonder about the representability of the setting at hand and thus to what extent data on behavior in gaming applications on cell phones and tablets is actually likely to carry over to other settings in which contest structures occur and might be of higher importance from a policy perspective. Investigating this argument further would require to look at data from other settings and check whether the respective results accord more closely to existing theory or rather confirm the patterns found here. Such a result would raise the issue of the generality of behavioral patterns in dynamic contests: Comparing the setting at hand to, for example, political campaigning as studied by Denter and Sisak (2013), it becomes apparent that these settings probably differ strongly with respect to the underlying production technologies and also the specificities of the respective game forms, although appearing rather similar at first. For instance, opportunity costs of effort are likely to differ substantially. The fact that these differences in the characteristics of the contest coincide with differential behavioral patterns indicates that they should be investigated further, as they potentially have an substantial effect on outcomes.

Second, the existing theoretical literature might not yet have been able to capture the main determinants and mechanisms actually underlying behavior in competitive situations. In this regard, especially the systematic incorporation of alternative decision models and preference structures from the behavioral economics literature might prove valuable in explaining behavior in settings like the one studied here. A systematic analysis of the consequences of reference dependence in dynamic contests thereby seems particularly promising, given the multitude of potential reference points due to the different kinds of information typically revealed during dynamic contests.

Finally, another potential explanation for the incoherence of theory and empirical evidence is the likely heterogeneity of behavioral types among participants. As the study at hand only documents average patterns in the data, these can be interpreted as behavior of the “average individual” (in the sample). However, if individuals are actually sufficiently heterogeneous with respect to their behavioral patterns, averaging their behavior is unlikely

⁹However, assuming substitutability between effort and ability might generate predictions matching the patterns documented here.

to yield reasonable conclusions. Given the variety of plausible theoretical mechanisms in dynamic contests, type heterogeneity seems to be particularly likely. The question would then be to what extent at least behavior by the same individual adheres to a consistent behavioral type, or put differently, whether within-individual behavior can be rationalized by specific theoretical mechanisms. In principle, the data at hand is well suited for this endeavor, given the substantial number of individuals on which a reasonable amount of data exists. An endogenous classification of individuals to different behavioral types could therefore be reasonably tested. The relative frequency of different behavioral types could then in turn also inform theory about the relative relevancy of different theoretical mechanisms and thus provide guidance to future theoretical research.

References

- BARANKAY, I. (2012a): “Rank Incentives Evidence from a Randomized Workplace Experiment,” mimeo.
- (2012b): “Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment,” mimeo.
- BAYE, M., D. KOVENOCK, AND C. DE VRIES (1996): “The All-Pay Auction with Complete Information,” *Economic Theory*, 8, 291–305.
- CASAS-ARCE, P. AND A. MARTINEZ-JEREZ (2009): “Relative Performance Compensation, Contests, and Dynamic Incentives,” *Management Science*, 55, 1306–1320.
- DECHENAUX, E., D. KOVENOCK, AND R. SHEREMETA (2012): “A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments,” WZB Discussion Paper SP II 2012109.
- DELFGAAUW, J., R. DUR, A. NON, AND W. VERBEKE (2012): “The Effects of Prize Spread and Noise in Elimination Tournaments: A Natural Field Experiment,” IZA Discussion Paper 6480.
- DENTER, P. AND D. SISAK (2013): “Do Polls Create Momentum in Political Competition?” Tinbergen Institute Discussion Paper 13-169/VII.
- EDERER, F. (2010): “Feedback and Motivation in Dynamic Tournaments,” *Journal of Economics and Management Strategy*, 19, 733–769.
- GILL, D. AND V. PROWSE (2012): “A Structural Analysis of Disappointment Aversion in a Real Effort Competition,” *American Economic Review*, 102, 469–503.
- KONRAD, K. (2009): *Strategy and Dynamics in Contests*, Oxford University Press, USA.
- LAZEAR, E. AND S. ROSEN (1981): “Rank-Order Tournaments as Optimum Labor Contracts,” *Journal of Political Economy*, 89, 841–64.
- MOLDOVANU, B. AND A. SELA (2001): “The Optimal Allocation of Prizes in Contests,” *American Economic Review*, 91, 542–558.
- TULLOCK, G. (1980): “Efficient Rent-Seeking,” in *Toward a Theory of the Rent-Seeking Society*, ed. by J. M. Buchanan, R. D. Tollison, and G. Tullock, Texas A&M University Press, 97–112.