

On the Descriptive Legitimacy of Status Differences

Janina Kraus and Stefan Traub

WISSENSCHAFTSRAUM VERHALTENSÖKONOMIK
UND GESELLSCHAFTLICHE TRANSFORMATION



Gefördert durch

zukunft.
niedersachsen

Ein Förderprogramm des Niedersächsischen
Ministeriums für Wissenschaft und Kultur
und der VolkswagenStiftung.

On the Descriptive Legitimacy of Status Differences

Janina Kraus*

Stefan Traub[†]

May 28, 2026

Abstract

Despite the existence of a variety of mechanisms for assigning roles and building groups that have been established in laboratory experiments, there is a dearth of knowledge regarding the extent to which these mechanisms are perceived as legitimate. This empirical study examines four different mechanisms for inducing status differences with regard to their descriptive legitimacy and finds that a random classification is evaluated as the most legitimate mechanism. Conversely, the university entrance grade is evaluated as the least legitimate mechanism. Women perceive greater differences in legitimacy between the mechanisms.

Keywords: Legitimacy, Status Differences, Group-building Mechanism

JEL classification: C91, C92, D90

*Corresponding author: Leuphana University Lüneburg, Institut of Economics, Universitaetsallee 1, 21335 Lüneburg, janina.kraus@leuphana.de, +49 4131 677 2938

[†]Helmut Schmidt University, Department of Economics, Holstenhofweg 85, 22043 Hamburg, traub@hsu-hh.de

1 Introduction

In laboratory experiments, a variety of different mechanisms are employed to divide participants into groups, thereby inducing status differences. As the acceptance of status differences and subsequent behavior may depend on their descriptive, i.e., perceived, legitimacy (among others, see e.g., Turner and Brown, 1978; Bettencourt et al., 2001; Olson and Hafer, 2001), it is fundamental to understand how the descriptive legitimacy of these mechanisms is assessed. For instance, if status differences are based on mechanisms that are not perceived as legitimate, this may have a substantial impact on the acceptance of decisions made by people with higher status and their outcomes.

Although there are numerous definitions of status, they share three common dimensions. First, status is the individual's ranking within an entity, which is based on different factors, such as skills, characteristics, wealth, or norms. Second, the perception and acceptance of status by the entity is necessary. Third, status is associated with access to or the attribution of resources and/or preferential treatment by others. Summarized, status is conceptualized as a relational construct, wherein individuals engage in comparisons with others within their respective entity (Ball et al., 2001; Henrich and Gil-White, 2001).

Closely linked to status is entitlement, which is conceptualized as "legally enforceable claims [...] to economic resources" (p. 260), reflecting subjective beliefs in one's right to earn and defend such claims (Hoffman and Spitzer, 1985). Rooted in Lerner's (1987) notion of a just world, entitlements justify outcomes as fair when based on effort or personal characteristics (Hülle et al., 2018). Consequently, individuals may perceive greater rewards or resource access as legitimate – a phenomenon known as the entitlement effect (Demiral and Mollerstrom, 2018).

Thus, entitlement signifies the right to determine the allocation of resources, whereas status denotes one's role or position within a societal structure. Given our focus on group differences and the extent to which these are perceived as legitimate, we refer to status differences.

The concept of legitimacy is linked to the expectation that a person or institution is legitimate, so their actions and decisions are accepted voluntarily. As such, legitimacy constitutes an important element of many social interactions and processes (Tyler, 2001; Zelditch, 2001). Typically, two perspectives of legitimacy are considered. Normative legitimacy encompasses objective criteria that delineate what is considered legitimate and the standards by which legitimacy is determined. In contrast, descriptive legitimacy refers to the perception of legitimacy among the general public (Hinsch, 2008). An entity may be perceived as normative legitimate, yet not be descriptively legitimate (Hough et al., 2013). A well-known example where the descriptive and normative standards of legitimacy are not met is the Nazi regime. Despite the regime's support among the German population (descriptive legitimacy), it failed to achieve a basic level of normative legitimacy (Hough

et al., 2013).

If status differences are regarded legitimate, they imply a legitimate decision-maker whose decisions are (voluntarily) accepted (French and Raven, 1959). Similarly, the acceptance of unpleasant outcomes and inequalities is more likely to occur, particularly by groups with lower status (see e.g., Olson and Hafer, 2001; Levin et al., 2002). If a high-status group decides to implement an outcome that is disadvantageous or less profitable for the low-status group, it is more likely to be accepted because the high-status group is entitled to a superior outcome (Turner and Brown, 1978). Furthermore, individuals in a low-status position are more likely to identify with their lower status if it has been implemented through a legitimate procedure, while non-legitimate procedures lead to more competitive behavior between status groups (Ellemers et al., 1993). Thus, the advantages of a higher status position are only realized when the underlying system is perceived as legitimate.

Consequently, in the context of laboratory experiments, where status differences are deliberately introduced and the subsequent decision-making behavior is analyzed, it is imperative to ascertain whether these status differences are perceived as legitimate and accepted by all parties involved. There are usually a variety of procedures and mechanisms used to induce status differences in laboratory experiments. In many cases status groups are randomly determined, i.e., by the computer (see e.g., Bhatia et al., 2024) or by a coin toss (see e.g., Hoffman and Spitzer, 1985). Another mechanism of inducing status differences is through the completion of “real-effort tasks,” i.e. the slider task (see e.g., Gill and Prowse, 2019). Furthermore, knowledge quizzes (see e.g., Fleiß, 2015) and cognitive ability tests (see e.g., Paetzel and Sausgruber, 2018) are also frequently used. The allocation of a status position is then based on the performance observed in the real-effort-task or quiz. Based on this, five categories of tasks are distinguishable: (1) random procedures, (2) real-effort tasks, (3) cognitive ability tasks, (4) quizzes, and (5) characteristics acquired outside the laboratory.¹ In our study, we use four different mechanism corresponding to four of the five categories, in terms of their legitimacy. We describe the decision situation in more detail in the next section.

A consensus among experimental researchers regarding the definition of a legitimate status-building mechanism remains elusive. Nevertheless, this is an important area of study as understanding the factors that influence the formation of groups is crucial to understanding the acceptance of status differences as described above. The present study aspires to contribute to this understanding by exploring which mechanism of group formation is perceived as more or less legitimate in creating status differences. This methodological contribution offers a novel perspective on the mechanisms through

¹Table 3 in the Appendix provides an overview of tasks used in economic laboratory experiments and names exemplary papers for each tasks. This overview focuses on experiments from economics; however, it should be noted that there are more procedures from other disciplines, such as psychology.

which legitimate status differences emerge in laboratory experiments.

2 Data and Method

We collected the data with an online survey in July 2020. The survey was programmed in oTree (Chen et al., 2016) and participants were recruited via hroot (Bock et al., 2014), to evaluate all four mechanisms in a within-subjects design. A total of 105 students from the university of Hamburg participated in the study, comprising approximately a third of whom were male (34.29%).

At the beginning of the study, the decision situation was presented in which six individuals were to be divided into three *decision-makers* and three *recipients*. The decision-makers were then permitted to determine the division of a sum of money between themselves and the recipients.² Here, participants are not explicitly declared as having a higher status. Rather, the instructions explain that decision-makers have the right to determine the final allocation for the whole group. Thus, status differences are induced indirectly as this measures the effect of status differences on the assigned responsibility and not identity implications.

Subsequently, the participants were presented with the following four different mechanisms in a randomized sequence, accompanied by an illustrative example: (i) a random assignment, the most common method to induce status differences, (ii) the real-effort slider task (Gill and Prowse, 2019), (iii) Raven’s matrices, sometimes referred to as an IQ test (see e.g., Putterman et al., 2011), and (iv) the university entrance grade. Participants had to evaluate the legitimacy of each mechanism for dividing the six individuals into two status groups on a seven-point Likert scale (1 = completely not legitimate and 7 = completely legitimate). The study concluded with a short questionnaire on demographic information.

In this study, we deliberately abstained from financially incentivizing the decision, as our objective was to avoid creating an incentive for participants to select the mechanism that best suited their preferences. The participants’ decision should be as unbiased as possible and should not reflect expectations of future profits.

3 Results

Figure 1 depicts the mean rating of the four mechanisms, with higher values indicating a higher legitimacy assessment.

²Our investigation of the descriptive legitimacy of status assignment mechanisms served as a preliminary study for another experiment that tested a modification of Bartling et al. (2015). The decision-making situation described in the instructions here was based on their experiment (see Appendix ?? for the exact wording).

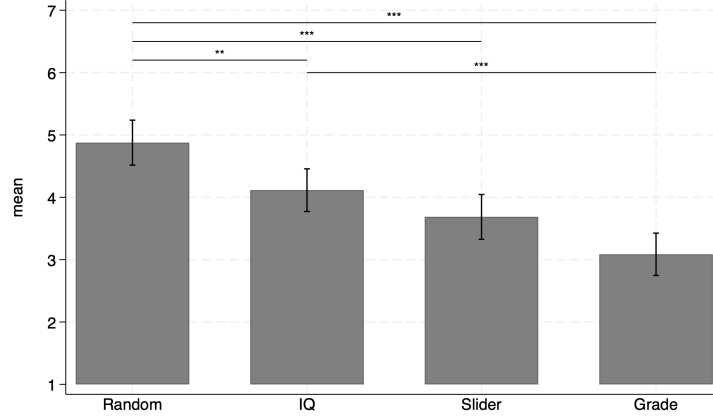


Figure 1: Rating of the Four Mechanisms

Notes: Rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). The p-values are from two-tailed t-tests with adjusted p-values (Bonferroni). Bars represent the 95% confidence interval. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

The most legitimate mechanism is a random procedure, which is rated on average with 4.88 out of 7. IQ and slider are the second and third most legitimate mechanisms, respectively, while the university entrance grade is the least legitimate mechanism (3.09). The difference between random and the other three mechanisms is highly significant, as is the difference between IQ and university entrance grade. Table 1 displays the mean values, standard deviations, medians and p-values of the significance tests, summarizing our descriptive results.

Table 1: Summary of the Four Mechanisms

Rating	N	Mean	Std. Dev.	Median	IQ	Slider	Grade
Random	105	4.88	1.86	5	$p = 0.016$	$p < 0.001$	$p < 0.001$
					$p = 0.015$	$p < 0.001$	$p < 0.001$
IQ	105	4.11	1.77	4	/	$p = 0.529$	$p < 0.001$
						$p = 0.585$	$p < 0.001$
Slider	105	3.69	1.86	4	/	/	$p = 0.102$
							$p = 0.086$
Grade	105	3.09	1.75	3	/	/	/

Table notes: Rating scale from 1 (completely not legitimate) to 7 (completely legitimate). The upper p-value results from two-tailed t-tests, assuming equal variances, while the lower p-value results from sign-rank tests. Both tests are controlled for multiple hypothesis testing (Bonferroni corrected p-values).

Figure 2 shows the cumulative distribution functions of the four mechanisms. As anticipated, the university entrance grade is deemed illegitimate by approximately half of the respondents (rating less than 3), while the random mechanism is regarded as

illegitimate by less than one-fifth. The remaining two mechanisms are positioned at the midpoint between these two extremes. The distribution function of the random mechanism exhibits a marked increase from a value of 3, which is indicative of a high legitimacy assessment. Conversely, the university entrance grade is deemed legitimate by only approximately 20% of respondents who allocated a rating of 5 or above.

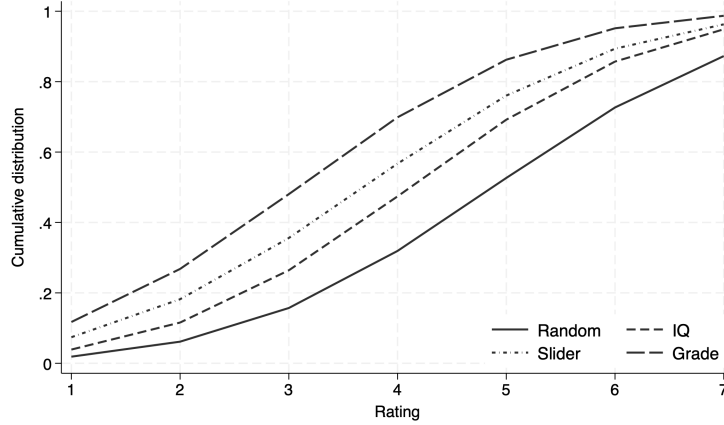


Figure 2: Cumulative Distribution of Rating Behavior

Notes: Rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). Smoothed gaussian probability curve.

Next, we econometrically analyze the rating behavior by ordered logit regression using the rating as the dependent variable (see Table 2).³ In Models (1) to (3), the entire sample is utilized; in Models (4) and (5), the sample is divided according to gender.

Model (1) confirms that IQ, slider, and university entrance grade receive significantly lower ratings than random. Furthermore, the mutual differences between the four mechanisms are significantly different at least at the 10%-level, except for IQ and Slider (see Table 4 in Appendix A). Controlling for gender, age, and risk attitude in Model (2) does not change the results. None of the covariates prove to be significant. In Model (3), we interact the gender dummy with the mechanism, thereby providing a more nuanced picture. Although women and men (as shown in Model (2)) find the mechanisms equally legitimate overall, women tend to find the random mechanism significantly more legitimate than men and the other three mechanisms significantly less legitimate. The university entrance grade stands out in particular with -1.276 points. Overall, women seem to react more strongly to the differences between the mechanisms in their perception of legitimacy.

³As a robustness check, we ran an additional tobit regression (see Table 5 in the Appendix), which yields similar results.

Table 2: Ordered Logit Regression

Rating Behavior	(1)	(2)	(3)	(4) only women	(5) only men
<i>Baseline: Random</i>					
IQ	-0.900*** (0.287)	-0.915*** (0.299)	-0.267 (0.445)	-1.244*** (0.403)	-0.092 (0.477)
Slider	-1.299*** (0.272)	-1.305*** (0.271)	-0.798 (0.546)	-1.603*** (0.316)	-0.755 (0.546)
Grade	-1.921*** (0.321)	-1.934*** (0.322)	-1.102** (0.514)	-2.397*** (0.422)	-1.030** (0.490)
Female		-0.222 (0.236)	0.554 (0.451)		
<i>Baseline: Random#Female</i>					
IQ#Female			-0.988* (0.572)		
Slider#Female			-0.788 (0.623)		
Grade#Female			-1.276** (0.642)		
Age		-0.025 (0.020)	-0.025 (0.020)	-0.039* (0.023)	0.034 (0.050)
<i>Baseline: Riskneutral</i>					
Riskaffine		0.436 (0.320)	0.455 (0.326)	0.231 (0.419)	0.757 (0.530)
Riskaverse		-0.084 (0.262)	-0.071 (0.267)	-0.315 (0.352)	0.227 (0.396)
Round dummies	No	Yes	Yes	Yes	Yes
Constant	0.399** (0.183)	0.361** (0.164)	0.387** (0.166)	0.330* (0.199)	0.519 (0.319)
N	420	420	420	276	144
<i>Table notes:</i> Ordered logit regression with panel estimator and robust standard errors. Rating from 1 (= completely not legitimate) to 7 (= completely legitimate) as dependent variable. Covariates: Female (0=no, 1=yes), Risk Attitude (Baseline=risk neutral). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.					

To further investigate gender differences, we split our dataset and estimate the regression separately for women (Model 4) and men (Model 5). The evaluation behavior of women is more dispersed than that of men. More specifically, the difference between the most legitimate mechanism random and the least legitimate mechanism university entrance grade is 2.40 (out of seven), which is highly statistically significant ($p < 0.01$). In addition, IQ and slider are also considered significantly less legitimate than random. Turning to Model (5), reveals that men do not rate IQ and slider as significantly more (less) legitimate than a random procedure. Only the university entrance grade is evaluated as less legitimate ($p < 0.05$).

Taken these results together, a random procedure is the most legitimate mechanism for inducing status differences, whereas the university entrance grade is regarded as the least legitimate mechanism. Furthermore, women exhibit a more dispersed rating behavior and perceive the university entrance grade as even less legitimate than men do.

4 Discussion and Conclusion

This paper analyzes the descriptive legitimacy of four different group-building mechanisms to induce status differences in laboratory experiments. A random assignment is considered the most legitimate mechanism for determining status, whereas the university entrance grade is the least legitimate. Thus, a normatively legitimate mechanism (the university entrance grade) is not inherently perceived or accepted as descriptively legitimate, while a non-normatively legitimate mechanism (random) is perceived as descriptively legitimate. Additionally, women perceive greater differences in legitimacy between the mechanisms. Although the university entrance grade is determined by pre-defined standards, it represents an objective criterion that fulfills the requirements of normative legitimacy. However, this study has demonstrated that it is not descriptively legitimate, as it lacks support from the individuals concerned. In contrast, a random mechanism lacks specific, objective criteria that must be fulfilled for a person to become a decision-maker. From a normative perspective, this mechanism is therefore not legitimate, while this study has demonstrated that it is descriptively legitimate.

Our data does not permit an exhaustive response to the question of how these differences between normative and descriptive legitimacy of the tested mechanisms came about. However, there are already some studies that offer insights into the acceptance and assessment of both mechanisms.

The university entrance grade reflects a longer learning process that requires effort and persistence (Wentzel, 1991) and is a significant factor in achieving high social status (Weiss and Fershtman, 1998). Additionally, it is a significant factor in university and job applications (Fleiß, 2015; Burnham, 2018), and serves as the most reliable indicator of academic success (Brandstätter and Farthofer, 2002; Gold and Souvignier, 2005). In Germany, the criteria for university student selection are defined by the *Hochschulrahmengesetz* (Framework Act for Higher Education). According to this regulation, the final school-leaving grade is a mandatory component of the admission process, whereas additional factors—such as weighted individual grades, entrance examinations, interviews, or prior professional experience (e.g., vocational training)—are optional. Within this framework, the final grade is considered the most reliable predictor of academic success (see, among others, Brandstätter and Farthofer, 2002; Gold and Souvignier, 2005).

However, it is also the subject of considerable criticism, as it is contingent upon a relatively brief period (two years in Germany) encompassing numerous examinations

across diverse courses (Camara, 2005). Furthermore, subjective and non-academic factors frequently exert a significant influence, such as the subjective assessments of teachers (Süß, 2001; Ingenkamp, 1997), which could provide an explanation for the non-legitimate assessment.

The high descriptive legitimacy of randomization is consistent with findings from experimental studies indicating that individuals tend to prefer a random procedure that determines outcomes, particularly when another person is affected by the outcome (see e.g., Sandroni et al., 2013; Cettolin and Riedl, 2019). In contrast, the university entrance grade is not a direct reflection of the task itself, which may contribute to the perception that it is descriptively unrelated.

Our second finding concerns gender differences. Relating to this, previous research has demonstrated that men tend to exhibit greater competitiveness and a heightened sensitivity to status differences relative to women (see e.g., Campbell, 2002; Burnham, 2018). Consequently, it is not unexpected that women evaluated randomness as more legitimate, as this mechanism is the least competitive. In a cross-national meta-analysis, Voyer and Voyer (2014) find that women outperform men in school grades across all educational levels, despite men achieving higher scores on standardized tests, particularly in math and science. Although women seem to perform better at school and therefore would benefit from the university entrance grade as inducing status difference, they perceive it as less legitimate than all other tested mechanisms.

A notable limitation of our study is the exclusive use of a student sample, which may limit the generalizability of the findings to a broader population. Additionally, the university entrance grades currently serve as a more pertinent metric in this context, as they are directly linked to the academic performance and success of students in higher education institutions. It would be worthwhile to examine how the assessment of legitimacy would perform in a representative sample, particularly given that grades should be less pertinent in this context than for students.

Summarized, this study makes a methodological contribution to the acceptance of status differences in laboratory experiments. Thus, it contributes to a better understanding how (legitimate) status differences can be implemented, which is fundamental to understand, as the acceptance of decisions depends on their legitimacy perceptions. It is to be expected that, e.g., distribution decisions made by individuals with different statuses will be influenced not only by self-interest and social preferences, but also by the perceived legitimacy of the status difference. Therefore, this study provides a foundation for future research, particularly in the realm of normative and descriptively legitimate mechanisms that can be employed to induce status differences.

CrediT authorship contribution statement

Janina Kraus: Conceptualization, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration

Stefan Traub: Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Supervision, Visualization

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Studies in humans

All procedures were performed in compliance with relevant laws and institutional guidelines. The privacy rights of human subjects have been observed and informed consent was obtained for experimentation with human subjects.

Data availability

The Stata do-file and data set to reproduce the results are available at <https://osf.io/untjf/>.

Acknowledgements

We thank Fabian Paetzel, as well as the participants of the economics workshop at the Helmut Schmidt University, TU Clausthal, the GfeW Conference 2021 in Magdeburg, and the 2022 ESA World Meeting in Boston.

References

- Ball, S., Eckel, C.C., 1998. The economic value of status. *Journal of Socio-Economics* 27, 495–514.
- Ball, S., Eckel, C.C., Grossman, P.J., Zame, W., 2001. Status in markets. *The Quarterly Journal of Economics* 116, 161–188. doi:10.1162/003355301556374.
- Bartling, B., Fischbacher, U., Schudy, S., 2015. Pivotality and responsibility attribution in sequential voting. *Journal of Public Economics* 128, 133–139. doi:10.1016/j.jpubeco.2015.03.010.
- Bettencourt, B.A., Charlton, K., Dorr, N., Hume, D.L., 2001. Status differences and in-group bias: A meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. *Psychological Bulletin* 127, 520–542. doi:10.1037/0033-2909.127.4.520.
- Bhatia, D., Fischbacher, U., Hausfeld, J., Stumpf, R., 2024. Blame and praise: responsibility attribution patterns in decision chains. *Experimental Economics* 27, 637–663.
- Blue, P.R., Hu, J., Wang, X., Dijk, E.v., Zhou, X., 2016. When do low status individuals accept less? the interaction between self- and other-status during resource distribution. *Frontiers in Psychology* 7. doi:10.3389/fpsyg.2016.01667.
- Bock, O., Baetge, I., Nicklisch, A., 2014. hroot: Hamburg registration and organization online tool. *European Economic Review* 71, 117–120. doi:10.1016/j.euroecorev.2014.07.003.
- Brandstätter, H., Farthofer, A., 2002. Studienerfolgsprognose - konfigurativ oder linear additiv? *Zeitschrift für Differentielle und Diagnostische Psychologie* 23(4), 381–391.
- Burnham, T.C., 2018. Gender, punishment, and cooperation: Men hurt others to advance their interests. *Socius: Sociological Research for a Dynamic World* 4, 237802311774224. doi:10.1177/2378023117742245.
- Camara, W.J., 2005. Broadening predictors of college success and the impact of cognitive predictors, in: Wayne J. Camara, Ernest W. Kimmel (Eds.), *Choosing students: Higher Education Admissions Tools for the 21st Century*. Routledge, pp. 53–79.
- Campbell, A., 2002. *A Mind Of Her Own*. Oxford University Press. doi:10.1093/acprof:oso/9780198504986.001.0001.
- Carmit Segal, 2012. Working when no one is watching: Motivation, test scores, and economic success. *Management Science* 58, 1438–1457. doi:10.1287/mnsc.1110.1509.
- Catherine C. Eckel, Philip J. Grossman, 2001. Chivalry and solidarity in ultimatum games. *Economic Inquiry* 39, 171–188. doi:10.1111/j.1465-7295.2001.tb00059.x.
- Cettolin, E., Riedl, A., 2019. Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization. *Journal of Economic Theory* 181, 547–585. doi:10.1016/j.jet.2019.03.002.

- Chen, D.L., Schonger, M., Wickens, C., 2016. otree\textmdashan open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9, 88–97. doi:10.1016/j.jbef.2015.12.001.
- Daniel Voyer, Susan D. Voyer, 2014. Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin* 140, 1174–1204. doi:10.1037/a0036620.
- Demiral, E.E., Mollerstrom, J., 2018. The entitlement effect in the ultimatum game \textmdash does it even exist? *Journal of Economic Behavior & Organization* 175, 341–352. doi:10.1016/j.jebo.2018.08.022.
- Elizabeth Hoffman, Kevin McCabe, Keith Shachat, Vernon Smith, 1994. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* 7, 346–380.
- Ellemers, N., Wilke, H., Knippenberg, A.v., 1993. Effects of the legitimacy of low group or individual status on individual and collective status-enhancement strategies. *Journal of Personality and Social Psychology* 64, 766–778. doi:10.1037/0022-3514.64.5.766.
- Essen, E.v., Ranehill, E., 2011. Dominance and submission: Social status biases economic sanctions. *SSE/EFI Working Paper Series in Economics and Finance* .
- Fleiß, J., 2015. Merit norms in the ultimatum game: An experimental study of the effect of merit on individual behavior and aggregate outcomes. *Central European Journal of Operations Research* 23, 389–406. doi:10.1007/s10100-015-0385-8.
- French, J.R.P., Raven, B., 1959. The bases of social power, in: Dorwin Cartwright (Ed.), *Studies in Social Power*. The University of Michigan, Ann Arbor, pp. 150–167.
- Gill, D., Prowse, V., 2019. Measuring costly effort using the slider task. *Journal of Behavioral and Experimental Finance* 21, 1–9. doi:10.1016/j.jbef.2018.11.003.
- Gold, A., Souvignier, E., 2005. Prognose der studierfähigkeit. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 37, 214–222. doi:10.1026/0049-8637.37.4.214.
- Henrich, J., Gil-White, F.J., 2001. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior* 22, 165–196. doi:10.1016/s1090-5138(00)00071-4.
- Hett, F., Mechtel, M., Kröll, M., 2020. The structure and behavioral effects of revealed social identity preferences. *The Economic Journal* 130, 2569–2595. doi:10.1093/ej/ueaa055.
- Hinsch, W., 2008. Legitimacy and justice, in: *Political Legitimation without Morality?*. Springer Netherlands, pp. 39–52. doi:10.1007/978-1-4020-8576-5\textunderscore4.
- Hoffman, E., Spitzer, M.L., 1985. Entitlements, rights, and fairness: An experimental examination of subjects\textquotesingle concepts of distributive justice. *The Journal of Legal Studies* 14, 259–297. doi:10.1086/467773.

- Hough, M., Jackson, J., Bradford, B., 2013. Legitimacy, trust and compliance: An empirical test of procedural justice theory using the European Social Survey, in: Justice Tankebe (Ed.), *Legitimacy and Criminal Justice: An International Exploration*. Oxford University Press, pp. 326–352.
- Hülle, S., Liebig, S., May, M.J., 2018. Measuring attitudes toward distributive justice: The basic social justice orientations scale. *Social Indicators Research* 136, 663–692. doi:10.1007/s11205-017-1580-x.
- Ingenkamp, K.H., 1997. *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Jeffrey V. Butler, 2014. Trust, truth, status and identity: An experimental inquiry. *The B.E. Journal of Theoretical Economics* 14, 293–338. doi:10.1515/bejte-2013-0026.
- Levin, S., Federico, C.M., Sidanius, J., Rabinowitz, J.L., 2002. Social dominance orientation and intergroup bias: The legitimization of favoritism for high-status groups. *Personality and Social Psychology Bulletin* 28, 144–157. doi:10.1177/0146167202282002.
- Melvin J. Lerner, 1987. Integrating societal and psychological rules of entitlement: The basic task of each social actor and fundamental problem for the social sciences. *Social Justice Research* 1, 107–125. doi:10.1007/bf01049386.
- Nick Feltovich, 2019. Is earned bargaining power more fully exploited? *Journal of Economic Behavior & Organization* 167, 152–180. doi:10.1016/j.jebo.2019.09.021.
- Olson, J.M., Hafer, C.L., 2001. Tolerance of personal deprivation, in: John T. Jost, Brenda Major (Eds.), *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Cambridge University Press, pp. 157–175.
- Paetzel, F., Sausgruber, R., 2018. Cognitive ability and in-group bias: An experimental study. *Journal of Public Economics* 167, 280–292. doi:10.1016/j.jpubeco.2018.04.006.
- Putterman, L., Tyran, J.R., Kamei, K., 2011. Public goods and voting on formal sanction schemes. *Journal of Public Economics* 95, 1213–1222. doi:10.1016/j.jpubeco.2011.05.001.
- Sandroni, A., Ludwig, S., Kircher, P., 2013. On the difference between social and private goods. *The B.E. Journal of Theoretical Economics* 13, 151–177. doi:10.1515/bejte-2012-0008.
- Süß, H.M., 2001. Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich, in: E. Stern (Ed.), *Perspektiven der Intelligenzforschung*. Lengerich: Pabst, pp. 109–136.
- Turner, J.C., Brown, R.J., 1978. Status, position, and legitimacy in intergroup behaviour: The effects of different status relationships on in-group bias and group creativity, in: *Differentiation between social groups: Studies in the social psychology of intergroup relations*. London: Academic Press Inc, pp. 201–234.

- Tyler, T.R., 2001. A psychological perspective on the legitimacy of institutions and authorities, in: John T. Jost, Brenda Major (Eds.), *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Cambridge University Press, pp. 416–436.
- Vanessa Mertins, Henrik Egbert, Tanja Könen, 2013. The effects of individual judgments about selection procedures: Results from a power-to-resist game. *The Journal of Socio-Economics* 42, 112–120. doi:10.1016/j.soc.2012.11.010.
- Weiss, Y., Fershtman, C., 1998. Social status and economic performance. *European Economic Review* 42, 801–820. doi:10.1016/s0014-2921(97)00137-2.
- Wentzel, K.R., 1991. Relations between social competence and academic achievement in early adolescence. *Child Development* 62, 1066. doi:10.2307/1131152.
- Zelditch, M., 2001. Theories of legitimacy, in: John T. Jost, Brenda Major (Eds.), *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Cambridge University Press, pp. 33–53.

Table 3: Overview of Tasks used in Laboratory Experiments

Category	Task	Exemplary paper
Random	Computer / Experimenter	Bartling et al. (2015), Bhatia et al. (2024)
	Coin flip	Hoffman and Spitzer (1985)
	Color chips	Butler (2014)
	Klee-Kandinsky	Paetzel and Sausgruber (2018)
Effort	Slider	Gill and Prowse (2019)
	Hash-mark game	Hoffman and Spitzer (1985)
	Encoding words / coding speed test	Feltovich (2019), Segal (2012)
Cognitive ability	Raven’s progressive matrices	Paetzel and Sausgruber (2018)
	Math/arithmetic task	Blue (2016), Demiral and Mollerstrom (2018)
	Summation task	Hett et al. (2020)
Quiz	Current event quiz	Hoffman et al. (1994)
	General knowledge quiz	Fleiß (2015)
	Economic trivia quiz	Ball et al. (1998; 2001)
Characteristics	Gender	Eckel and Grossman (2001), Campbell (2002)
	Name	Essen and Ranehill (2011)
	Age	Mertins et al. (2013)

A Further Results

Table 4: Wald-F Tests

Mechanism	Random		IQ		Slider	
	χ^2	$p(\chi^2)$	χ^2	$p(\chi^2)$	χ^2	$p(\chi^2)$
IQ	9.83	$p = 0.010$	-		-	
Slider	22.72	$p < 0.001$	3.99	$p = 0.274$	-	
Grade	35.91	$p < 0.001$	23.18	$p < 0.001$	6.15	$p = 0.079$

Table notes: The table shows the χ^2 statistics and significance levels p of a Wald-F test for equality of coefficients from Model 1 of the regression shown in Table 2 in the respective cells. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Tobit Regression: Rating Behavior

Rating Behavior	(1)	(2)	(3)	(4) only women	(5) only men
<i>Baseline: Random</i>					
IQ	-0.762*** (0.235)	-0.780*** (0.236)	-0.253 (0.403)	-1.019*** (0.282)	-0.125 (0.422)
Slider	-1.190*** (0.235)	-1.194*** (0.234)	-0.747* (0.398)	-1.422*** (0.282)	-0.756* (0.408)
Grade	-1.790*** (0.235)	-1.793*** (0.235)	-0.977** (0.401)	-2.211*** (0.281)	-0.960** (0.416)
Female		-0.208 (0.211)	0.468 (0.369)		
<i>Baseline: Random#Female</i>					
IQ#Female			-0.793 (0.496)		
Slider#Female			-0.677 (0.492)		
Grade#Female			-1.233** (0.494)		
Age		-0.015 (0.020)	-0.015 (0.020)	-0.027 (0.021)	0.034 (0.048)
<i>Baseline: Riskneutral</i>					
Riskaffine		0.368 (0.282)	0.368 (0.282)	0.150 (0.359)	0.657 (0.442)
Riskaverse		-0.090 (0.244)	-0.090 (0.244)	-0.322 (0.298)	0.228 (0.418)
Round dummies	No	Yes	Yes	Yes	Yes
Constant	4.876*** (0.176)	5.443*** (0.590)	5.034*** (0.617)	5.792*** (0.646)	3.872*** (1.300)
N	420	420	420	276	144

Table notes: Tobit regression with panel estimator. Rating from 1 (= completely not legitimate) to 7 (= completely legitimate) as dependent variable. Covariates: Female (0=no, 1=yes), Risk Attitude (Baseline=riskneutral). Baseline: random mechanism. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

On the Descriptive Legitimacy of Status Differences

Janina Kraus*

Stefan Traub[†]

May 28, 2026

Abstract

Despite the existence of a variety of mechanisms for assigning roles and building groups that have been established in laboratory experiments, there is a dearth of knowledge regarding the extent to which these mechanisms are perceived as legitimate. This empirical study examines four different mechanisms for inducing status differences with regard to their descriptive legitimacy and finds that a random classification is evaluated as the most legitimate mechanism. Conversely, the university entrance grade is evaluated as the least legitimate mechanism. Women perceive greater differences in legitimacy between the mechanisms.

Keywords: Legitimacy, Status Differences, Group-building Mechanism

JEL classification: C91, C92, D90

*Corresponding author: Leuphana University Lüneburg, Institut of Economics, Universitaetsallee 1, 21335 Lüneburg, janina.kraus@leuphana.de, +49 4131 677 2938

[†]Helmut Schmidt University, Department of Economics, Holstenhofweg 85, 22043 Hamburg, traub@hsu-hh.de

1 Introduction

In laboratory experiments, a variety of different mechanisms are employed to divide participants into groups, thereby inducing status differences. As the acceptance of status differences and subsequent behavior may depend on their descriptive, i.e., perceived, legitimacy (among others, see e.g., Turner and Brown, 1978; Bettencourt et al., 2001; Olson and Hafer, 2001), it is fundamental to understand how the descriptive legitimacy of these mechanisms is assessed. For instance, if status differences are based on mechanisms that are not perceived as legitimate, this may have a substantial impact on the acceptance of decisions made by people with higher status and their outcomes.

Although there are numerous definitions of status, they share three common dimensions. First, status is the individual's ranking within an entity, which is based on different factors, such as skills, characteristics, wealth, or norms. Second, the perception and acceptance of status by the entity is necessary. Third, status is associated with access to or the attribution of resources and/or preferential treatment by others. Summarized, status is conceptualized as a relational construct, wherein individuals engage in comparisons with others within their respective entity (Ball et al., 2001; Henrich and Gil-White, 2001).

Closely linked to status is entitlement, which is conceptualized as "legally enforceable claims [...] to economic resources" (p. 260), reflecting subjective beliefs in one's right to earn and defend such claims (Hoffman and Spitzer, 1985). Rooted in Lerner's (1987) notion of a just world, entitlements justify outcomes as fair when based on effort or personal characteristics (Hülle et al., 2018). Consequently, individuals may perceive greater rewards or resource access as legitimate – a phenomenon known as the entitlement effect (Demiral and Mollerstrom, 2018).

Thus, entitlement signifies the right to determine the allocation of resources, whereas status denotes one's role or position within a societal structure. Given our focus on group differences and the extent to which these are perceived as legitimate, we refer to status differences.

The concept of legitimacy is linked to the expectation that a person or institution is legitimate, so their actions and decisions are accepted voluntarily. As such, legitimacy constitutes an important element of many social interactions and processes (Tyler, 2001; Zelditch, 2001). Typically, two perspectives of legitimacy are considered. Normative legitimacy encompasses objective criteria that delineate what is considered legitimate and the standards by which legitimacy is determined. In contrast, descriptive legitimacy refers to the perception of legitimacy among the general public (Hinsch, 2008). An entity may be perceived as normative legitimate, yet not be descriptively legitimate (Hough et al., 2013). A well-known example where the descriptive and normative standards of legitimacy are not met is the Nazi regime. Despite the regime's support among the German population (descriptive legitimacy), it failed to achieve a basic level of normative legitimacy (Hough

et al., 2013).

If status differences are regarded legitimate, they imply a legitimate decision-maker whose decisions are (voluntarily) accepted (French and Raven, 1959). Similarly, the acceptance of unpleasant outcomes and inequalities is more likely to occur, particularly by groups with lower status (see e.g., Olson and Hafer, 2001; Levin et al., 2002). If a high-status group decides to implement an outcome that is disadvantageous or less profitable for the low-status group, it is more likely to be accepted because the high-status group is entitled to a superior outcome (Turner and Brown, 1978). Furthermore, individuals in a low-status position are more likely to identify with their lower status if it has been implemented through a legitimate procedure, while non-legitimate procedures lead to more competitive behavior between status groups (Ellemers et al., 1993). Thus, the advantages of a higher status position are only realized when the underlying system is perceived as legitimate.

Consequently, in the context of laboratory experiments, where status differences are deliberately introduced and the subsequent decision-making behavior is analyzed, it is imperative to ascertain whether these status differences are perceived as legitimate and accepted by all parties involved. There are usually a variety of procedures and mechanisms used to induce status differences in laboratory experiments. In many cases status groups are randomly determined, i.e., by the computer (see e.g., Bhatia et al., 2024) or by a coin toss (see e.g., Hoffman and Spitzer, 1985). Another mechanism of inducing status differences is through the completion of “real-effort tasks,” i.e. the slider task (see e.g., Gill and Prowse, 2019). Furthermore, knowledge quizzes (see e.g., Fleiß, 2015) and cognitive ability tests (see e.g., Paetzel and Sausgruber, 2018) are also frequently used. The allocation of a status position is then based on the performance observed in the real-effort-task or quiz. Based on this, five categories of tasks are distinguishable: (1) random procedures, (2) real-effort tasks, (3) cognitive ability tasks, (4) quizzes, and (5) characteristics acquired outside the laboratory.¹ In our study, we use four different mechanism corresponding to four of the five categories, in terms of their legitimacy. We describe the decision situation in more detail in the next section.

A consensus among experimental researchers regarding the definition of a legitimate status-building mechanism remains elusive. Nevertheless, this is an important area of study as understanding the factors that influence the formation of groups is crucial to understanding the acceptance of status differences as described above. The present study aspires to contribute to this understanding by exploring which mechanism of group formation is perceived as more or less legitimate in creating status differences. This methodological contribution offers a novel perspective on the mechanisms through

¹Table 3 in the Appendix provides an overview of tasks used in economic laboratory experiments and names exemplary papers for each tasks. This overview focuses on experiments from economics; however, it should be noted that there are more procedures from other disciplines, such as psychology.

which legitimate status differences emerge in laboratory experiments.

2 Data and Method

We collected the data with an online survey in July 2020. The survey was programmed in oTree (Chen et al., 2016) and participants were recruited via hroot (Bock et al., 2014), to evaluate all four mechanisms in a within-subjects design. A total of 105 students from the university of Hamburg participated in the study, comprising approximately a third of whom were male (34.29%).

At the beginning of the study, the decision situation was presented in which six individuals were to be divided into three *decision-makers* and three *recipients*. The decision-makers were then permitted to determine the division of a sum of money between themselves and the recipients.² Here, participants are not explicitly declared as having a higher status. Rather, the instructions explain that decision-makers have the right to determine the final allocation for the whole group. Thus, status differences are induced indirectly as this measures the effect of status differences on the assigned responsibility and not identity implications.

Subsequently, the participants were presented with the following four different mechanisms in a randomized sequence, accompanied by an illustrative example: (i) a random assignment, the most common method to induce status differences, (ii) the real-effort slider task (Gill and Prowse, 2019), (iii) Raven’s matrices, sometimes referred to as an IQ test (see e.g., Putterman et al., 2011), and (iv) the university entrance grade. Participants had to evaluate the legitimacy of each mechanism for dividing the six individuals into two status groups on a seven-point Likert scale (1 = completely not legitimate and 7 = completely legitimate). The study concluded with a short questionnaire on demographic information.

In this study, we deliberately abstained from financially incentivizing the decision, as our objective was to avoid creating an incentive for participants to select the mechanism that best suited their preferences. The participants’ decision should be as unbiased as possible and should not reflect expectations of future profits.

3 Results

Figure 1 depicts the mean rating of the four mechanisms, with higher values indicating a higher legitimacy assessment.

²Our investigation of the descriptive legitimacy of status assignment mechanisms served as a preliminary study for another experiment that tested a modification of Bartling et al. (2015). The decision-making situation described in the instructions here was based on their experiment (see Appendix ?? for the exact wording).

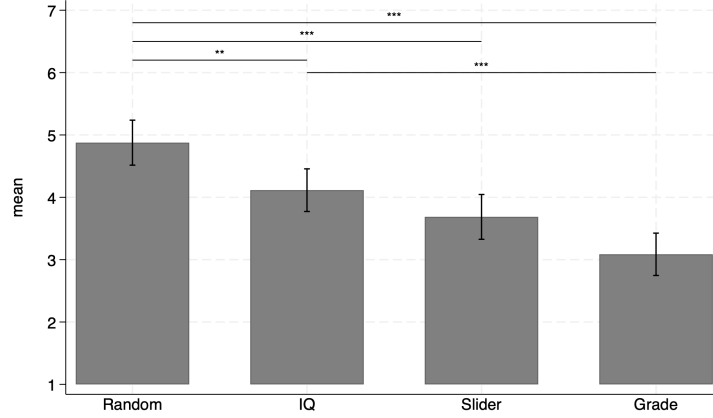


Figure 1: Rating of the Four Mechanisms

Notes: Rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). The p-values are from two-tailed t-tests with adjusted p-values (Bonferroni). Bars represent the 95% confidence interval. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

The most legitimate mechanism is a random procedure, which is rated on average with 4.88 out of 7. IQ and slider are the second and third most legitimate mechanisms, respectively, while the university entrance grade is the least legitimate mechanism (3.09). The difference between random and the other three mechanisms is highly significant, as is the difference between IQ and university entrance grade. Table 1 displays the mean values, standard deviations, medians and p-values of the significance tests, summarizing our descriptive results.

Table 1: Summary of the Four Mechanisms

Rating	N	Mean	Std. Dev.	Median	IQ	Slider	Grade
Random	105	4.88	1.86	5	$p = 0.016$	$p < 0.001$	$p < 0.001$
					$p = 0.015$	$p < 0.001$	$p < 0.001$
IQ	105	4.11	1.77	4	/	$p = 0.529$	$p < 0.001$
						$p = 0.585$	$p < 0.001$
Slider	105	3.69	1.86	4	/	/	$p = 0.102$
							$p = 0.086$
Grade	105	3.09	1.75	3	/	/	/

Table notes: Rating scale from 1 (completely not legitimate) to 7 (completely legitimate). The upper p-value results from two-tailed t-tests, assuming equal variances, while the lower p-value results from sign-rank tests. Both tests are controlled for multiple hypothesis testing (Bonferroni corrected p-values).

Figure 2 shows the cumulative distribution functions of the four mechanisms. As anticipated, the university entrance grade is deemed illegitimate by approximately half of the respondents (rating less than 3), while the random mechanism is regarded as

illegitimate by less than one-fifth. The remaining two mechanisms are positioned at the midpoint between these two extremes. The distribution function of the random mechanism exhibits a marked increase from a value of 3, which is indicative of a high legitimacy assessment. Conversely, the university entrance grade is deemed legitimate by only approximately 20% of respondents who allocated a rating of 5 or above.

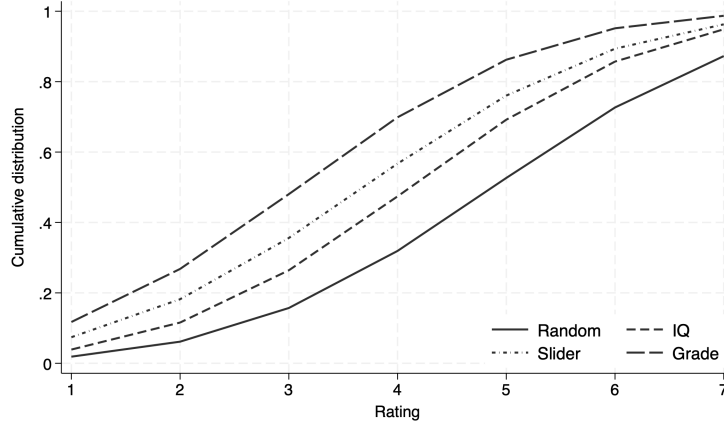


Figure 2: Cumulative Distribution of Rating Behavior

Notes: Rating scale range from 1 (completely not legitimate) to 7 (completely legitimate). Smoothed gaussian probability curve.

Next, we econometrically analyze the rating behavior by ordered logit regression using the rating as the dependent variable (see Table 2).³ In Models (1) to (3), the entire sample is utilized; in Models (4) and (5), the sample is divided according to gender.

Model (1) confirms that IQ, slider, and university entrance grade receive significantly lower ratings than random. Furthermore, the mutual differences between the four mechanisms are significantly different at least at the 10%-level, except for IQ and Slider (see Table 4 in Appendix A). Controlling for gender, age, and risk attitude in Model (2) does not change the results. None of the covariates prove to be significant. In Model (3), we interact the gender dummy with the mechanism, thereby providing a more nuanced picture. Although women and men (as shown in Model (2)) find the mechanisms equally legitimate overall, women tend to find the random mechanism significantly more legitimate than men and the other three mechanisms significantly less legitimate. The university entrance grade stands out in particular with -1.276 points. Overall, women seem to react more strongly to the differences between the mechanisms in their perception of legitimacy.

³As a robustness check, we ran an additional tobit regression (see Table 5 in the Appendix), which yields similar results.

Table 2: Ordered Logit Regression

Rating Behavior	(1)	(2)	(3)	(4) only women	(5) only men
<i>Baseline: Random</i>					
IQ	-0.900*** (0.287)	-0.915*** (0.299)	-0.267 (0.445)	-1.244*** (0.403)	-0.092 (0.477)
Slider	-1.299*** (0.272)	-1.305*** (0.271)	-0.798 (0.546)	-1.603*** (0.316)	-0.755 (0.546)
Grade	-1.921*** (0.321)	-1.934*** (0.322)	-1.102** (0.514)	-2.397*** (0.422)	-1.030** (0.490)
Female		-0.222 (0.236)	0.554 (0.451)		
<i>Baseline: Random#Female</i>					
IQ#Female			-0.988* (0.572)		
Slider#Female			-0.788 (0.623)		
Grade#Female			-1.276** (0.642)		
Age		-0.025 (0.020)	-0.025 (0.020)	-0.039* (0.023)	0.034 (0.050)
<i>Baseline: Riskneutral</i>					
Riskaffine		0.436 (0.320)	0.455 (0.326)	0.231 (0.419)	0.757 (0.530)
Riskaverse		-0.084 (0.262)	-0.071 (0.267)	-0.315 (0.352)	0.227 (0.396)
Round dummies	No	Yes	Yes	Yes	Yes
Constant	0.399** (0.183)	0.361** (0.164)	0.387** (0.166)	0.330* (0.199)	0.519 (0.319)
N	420	420	420	276	144
<i>Table notes:</i> Ordered logit regression with panel estimator and robust standard errors. Rating from 1 (= completely not legitimate) to 7 (= completely legitimate) as dependent variable. Covariates: Female (0=no, 1=yes), Risk Attitude (Baseline=risk neutral). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.					

To further investigate gender differences, we split our dataset and estimate the regression separately for women (Model 4) and men (Model 5). The evaluation behavior of women is more dispersed than that of men. More specifically, the difference between the most legitimate mechanism random and the least legitimate mechanism university entrance grade is 2.40 (out of seven), which is highly statistically significant ($p < 0.01$). In addition, IQ and slider are also considered significantly less legitimate than random. Turning to Model (5), reveals that men do not rate IQ and slider as significantly more (less) legitimate than a random procedure. Only the university entrance grade is evaluated as less legitimate ($p < 0.05$).

Taken these results together, a random procedure is the most legitimate mechanism for inducing status differences, whereas the university entrance grade is regarded as the least legitimate mechanism. Furthermore, women exhibit a more dispersed rating behavior and perceive the university entrance grade as even less legitimate than men do.

4 Discussion and Conclusion

This paper analyzes the descriptive legitimacy of four different group-building mechanisms to induce status differences in laboratory experiments. A random assignment is considered the most legitimate mechanism for determining status, whereas the university entrance grade is the least legitimate. Thus, a normatively legitimate mechanism (the university entrance grade) is not inherently perceived or accepted as descriptively legitimate, while a non-normatively legitimate mechanism (random) is perceived as descriptively legitimate. Additionally, women perceive greater differences in legitimacy between the mechanisms. Although the university entrance grade is determined by pre-defined standards, it represents an objective criterion that fulfills the requirements of normative legitimacy. However, this study has demonstrated that it is not descriptively legitimate, as it lacks support from the individuals concerned. In contrast, a random mechanism lacks specific, objective criteria that must be fulfilled for a person to become a decision-maker. From a normative perspective, this mechanism is therefore not legitimate, while this study has demonstrated that it is descriptively legitimate.

Our data does not permit an exhaustive response to the question of how these differences between normative and descriptive legitimacy of the tested mechanisms came about. However, there are already some studies that offer insights into the acceptance and assessment of both mechanisms.

The university entrance grade reflects a longer learning process that requires effort and persistence (Wentzel, 1991) and is a significant factor in achieving high social status (Weiss and Fershtman, 1998). Additionally, it is a significant factor in university and job applications (Fleiß, 2015; Burnham, 2018), and serves as the most reliable indicator of academic success (Brandstätter and Farthofer, 2002; Gold and Souvignier, 2005). In Germany, the criteria for university student selection are defined by the *Hochschulrahmengesetz* (Framework Act for Higher Education). According to this regulation, the final school-leaving grade is a mandatory component of the admission process, whereas additional factors—such as weighted individual grades, entrance examinations, interviews, or prior professional experience (e.g., vocational training)—are optional. Within this framework, the final grade is considered the most reliable predictor of academic success (see, among others, Brandstätter and Farthofer, 2002; Gold and Souvignier, 2005).

However, it is also the subject of considerable criticism, as it is contingent upon a relatively brief period (two years in Germany) encompassing numerous examinations

across diverse courses (Camara, 2005). Furthermore, subjective and non-academic factors frequently exert a significant influence, such as the subjective assessments of teachers (Süß, 2001; Ingenkamp, 1997), which could provide an explanation for the non-legitimate assessment.

The high descriptive legitimacy of randomization is consistent with findings from experimental studies indicating that individuals tend to prefer a random procedure that determines outcomes, particularly when another person is affected by the outcome (see e.g., Sandroni et al., 2013; Cettolin and Riedl, 2019). In contrast, the university entrance grade is not a direct reflection of the task itself, which may contribute to the perception that it is descriptively unrelated.

Our second finding concerns gender differences. Relating to this, previous research has demonstrated that men tend to exhibit greater competitiveness and a heightened sensitivity to status differences relative to women (see e.g., Campbell, 2002; Burnham, 2018). Consequently, it is not unexpected that women evaluated randomness as more legitimate, as this mechanism is the least competitive. In a cross-national meta-analysis, Voyer and Voyer (2014) find that women outperform men in school grades across all educational levels, despite men achieving higher scores on standardized tests, particularly in math and science. Although women seem to perform better at school and therefore would benefit from the university entrance grade as inducing status difference, they perceive it as less legitimate than all other tested mechanisms.

A notable limitation of our study is the exclusive use of a student sample, which may limit the generalizability of the findings to a broader population. Additionally, the university entrance grades currently serve as a more pertinent metric in this context, as they are directly linked to the academic performance and success of students in higher education institutions. It would be worthwhile to examine how the assessment of legitimacy would perform in a representative sample, particularly given that grades should be less pertinent in this context than for students.

Summarized, this study makes a methodological contribution to the acceptance of status differences in laboratory experiments. Thus, it contributes to a better understanding how (legitimate) status differences can be implemented, which is fundamental to understand, as the acceptance of decisions depends on their legitimacy perceptions. It is to be expected that, e.g., distribution decisions made by individuals with different statuses will be influenced not only by self-interest and social preferences, but also by the perceived legitimacy of the status difference. Therefore, this study provides a foundation for future research, particularly in the realm of normative and descriptively legitimate mechanisms that can be employed to induce status differences.

CrediT authorship contribution statement

Janina Kraus: Conceptualization, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration

Stefan Traub: Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Supervision, Visualization

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Studies in humans

All procedures were performed in compliance with relevant laws and institutional guidelines. The privacy rights of human subjects have been observed and informed consent was obtained for experimentation with human subjects.

Data availability

The Stata do-file and data set to reproduce the results are available at <https://osf.io/untjf/>.

Acknowledgements

We thank Fabian Paetzel, as well as the participants of the economics workshop at the Helmut Schmidt University, TU Clausthal, the GfeW Conference 2021 in Magdeburg, and the 2022 ESA World Meeting in Boston.

References

- Ball, S., Eckel, C.C., 1998. The economic value of status. *Journal of Socio-Economics* 27, 495–514.
- Ball, S., Eckel, C.C., Grossman, P.J., Zame, W., 2001. Status in markets. *The Quarterly Journal of Economics* 116, 161–188. doi:10.1162/003355301556374.
- Bartling, B., Fischbacher, U., Schudy, S., 2015. Pivotality and responsibility attribution in sequential voting. *Journal of Public Economics* 128, 133–139. doi:10.1016/j.jpubeco.2015.03.010.
- Bettencourt, B.A., Charlton, K., Dorr, N., Hume, D.L., 2001. Status differences and in-group bias: A meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. *Psychological Bulletin* 127, 520–542. doi:10.1037/0033-2909.127.4.520.
- Bhatia, D., Fischbacher, U., Hausfeld, J., Stumpf, R., 2024. Blame and praise: responsibility attribution patterns in decision chains. *Experimental Economics* 27, 637–663.
- Blue, P.R., Hu, J., Wang, X., Dijk, E.v., Zhou, X., 2016. When do low status individuals accept less? the interaction between self- and other-status during resource distribution. *Frontiers in Psychology* 7. doi:10.3389/fpsyg.2016.01667.
- Bock, O., Baetge, I., Nicklisch, A., 2014. hroot: Hamburg registration and organization online tool. *European Economic Review* 71, 117–120. doi:10.1016/j.euroecorev.2014.07.003.
- Brandstätter, H., Farthofer, A., 2002. Studienerfolgsprognose - konfigurativ oder linear additiv? *Zeitschrift für Differentielle und Diagnostische Psychologie* 23(4), 381–391.
- Burnham, T.C., 2018. Gender, punishment, and cooperation: Men hurt others to advance their interests. *Socius: Sociological Research for a Dynamic World* 4, 237802311774224. doi:10.1177/2378023117742245.
- Camara, W.J., 2005. Broadening predictors of college success and the impact of cognitive predictors, in: Wayne J. Camara, Ernest W. Kimmel (Eds.), *Choosing students: Higher Education Admissions Tools for the 21st Century*. Routledge, pp. 53–79.
- Campbell, A., 2002. *A Mind Of Her Own*. Oxford University Press. doi:10.1093/acprof:oso/9780198504986.001.0001.
- Carmit Segal, 2012. Working when no one is watching: Motivation, test scores, and economic success. *Management Science* 58, 1438–1457. doi:10.1287/mnsc.1110.1509.
- Catherine C. Eckel, Philip J. Grossman, 2001. Chivalry and solidarity in ultimatum games. *Economic Inquiry* 39, 171–188. doi:10.1111/j.1465-7295.2001.tb00059.x.
- Cettolin, E., Riedl, A., 2019. Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization. *Journal of Economic Theory* 181, 547–585. doi:10.1016/j.jet.2019.03.002.

- Chen, D.L., Schonger, M., Wickens, C., 2016. otree\textmdashan open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9, 88–97. doi:10.1016/j.jbef.2015.12.001.
- Daniel Voyer, Susan D. Voyer, 2014. Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin* 140, 1174–1204. doi:10.1037/a0036620.
- Demiral, E.E., Mollerstrom, J., 2018. The entitlement effect in the ultimatum game \textmdash does it even exist? *Journal of Economic Behavior & Organization* 175, 341–352. doi:10.1016/j.jebo.2018.08.022.
- Elizabeth Hoffman, Kevin McCabe, Keith Shachat, Vernon Smith, 1994. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* 7, 346–380.
- Ellemers, N., Wilke, H., Knippenberg, A.v., 1993. Effects of the legitimacy of low group or individual status on individual and collective status-enhancement strategies. *Journal of Personality and Social Psychology* 64, 766–778. doi:10.1037/0022-3514.64.5.766.
- Essen, E.v., Ranehill, E., 2011. Dominance and submission: Social status biases economic sanctions. *SSE/EFI Working Paper Series in Economics and Finance* .
- Fleiß, J., 2015. Merit norms in the ultimatum game: An experimental study of the effect of merit on individual behavior and aggregate outcomes. *Central European Journal of Operations Research* 23, 389–406. doi:10.1007/s10100-015-0385-8.
- French, J.R.P., Raven, B., 1959. The bases of social power, in: Dorwin Cartwright (Ed.), *Studies in Social Power*. The University of Michigan, Ann Arbor, pp. 150–167.
- Gill, D., Prowse, V., 2019. Measuring costly effort using the slider task. *Journal of Behavioral and Experimental Finance* 21, 1–9. doi:10.1016/j.jbef.2018.11.003.
- Gold, A., Souvignier, E., 2005. Prognose der studierfähigkeit. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 37, 214–222. doi:10.1026/0049-8637.37.4.214.
- Henrich, J., Gil-White, F.J., 2001. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior* 22, 165–196. doi:10.1016/s1090-5138(00)00071-4.
- Hett, F., Mechtel, M., Kröll, M., 2020. The structure and behavioral effects of revealed social identity preferences. *The Economic Journal* 130, 2569–2595. doi:10.1093/ej/ueaa055.
- Hinsch, W., 2008. Legitimacy and justice, in: *Political Legitimation without Morality?*. Springer Netherlands, pp. 39–52. doi:10.1007/978-1-4020-8576-5\textunderscore4.
- Hoffman, E., Spitzer, M.L., 1985. Entitlements, rights, and fairness: An experimental examination of subjects\textquotesingle concepts of distributive justice. *The Journal of Legal Studies* 14, 259–297. doi:10.1086/467773.

- Hough, M., Jackson, J., Bradford, B., 2013. Legitimacy, trust and compliance: An empirical test of procedural justice theory using the European social survey, in: Justice Tankebe (Ed.), *Legitimacy and Criminal Justice: An International Exploration*. Oxford University Press, pp. 326–352.
- Hülle, S., Liebig, S., May, M.J., 2018. Measuring attitudes toward distributive justice: The basic social justice orientations scale. *Social Indicators Research* 136, 663–692. doi:10.1007/s11205-017-1580-x.
- Ingenkamp, K.H., 1997. *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Jeffrey V. Butler, 2014. Trust, truth, status and identity: An experimental inquiry. *The B.E. Journal of Theoretical Economics* 14, 293–338. doi:10.1515/bejte-2013-0026.
- Levin, S., Federico, C.M., Sidanius, J., Rabinowitz, J.L., 2002. Social dominance orientation and intergroup bias: The legitimization of favoritism for high-status groups. *Personality and Social Psychology Bulletin* 28, 144–157. doi:10.1177/0146167202282002.
- Melvin J. Lerner, 1987. Integrating societal and psychological rules of entitlement: The basic task of each social actor and fundamental problem for the social sciences. *Social Justice Research* 1, 107–125. doi:10.1007/bf01049386.
- Nick Feltovich, 2019. Is earned bargaining power more fully exploited? *Journal of Economic Behavior & Organization* 167, 152–180. doi:10.1016/j.jebo.2019.09.021.
- Olson, J.M., Hafer, C.L., 2001. Tolerance of personal deprivation, in: John T. Jost, Brenda Major (Eds.), *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Cambridge University Press, pp. 157–175.
- Paetzel, F., Sausgruber, R., 2018. Cognitive ability and in-group bias: An experimental study. *Journal of Public Economics* 167, 280–292. doi:10.1016/j.jpubeco.2018.04.006.
- Putterman, L., Tyran, J.R., Kamei, K., 2011. Public goods and voting on formal sanction schemes. *Journal of Public Economics* 95, 1213–1222. doi:10.1016/j.jpubeco.2011.05.001.
- Sandroni, A., Ludwig, S., Kircher, P., 2013. On the difference between social and private goods. *The B.E. Journal of Theoretical Economics* 13, 151–177. doi:10.1515/bejte-2012-0008.
- Süß, H.M., 2001. Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich, in: E. Stern (Ed.), *Perspektiven der Intelligenzforschung*. Lengerich: Pabst, pp. 109–136.
- Turner, J.C., Brown, R.J., 1978. Status, position, and legitimacy in intergroup behaviour: The effects of different status relationships on in-group bias and group creativity, in: *Differentiation between social groups: Studies in the social psychology of intergroup relations*. London: Academic Press Inc, pp. 201–234.

- Tyler, T.R., 2001. A psychological perspective on the legitimacy of institutions and authorities, in: John T. Jost, Brenda Major (Eds.), *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Cambridge University Press, pp. 416–436.
- Vanessa Mertins, Henrik Egbert, Tanja Könen, 2013. The effects of individual judgments about selection procedures: Results from a power-to-resist game. *The Journal of Socio-Economics* 42, 112–120. doi:10.1016/j.soc.2012.11.010.
- Weiss, Y., Fershtman, C., 1998. Social status and economic performance. *European Economic Review* 42, 801–820. doi:10.1016/S0014-2921(97)00137-2.
- Wentzel, K.R., 1991. Relations between social competence and academic achievement in early adolescence. *Child Development* 62, 1066. doi:10.2307/1131152.
- Zelditch, M., 2001. Theories of legitimacy, in: John T. Jost, Brenda Major (Eds.), *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations*. Cambridge University Press, pp. 33–53.

Table 3: Overview of Tasks used in Laboratory Experiments

Category	Task	Exemplary paper
Random	Computer / Experimenter	Bartling et al. (2015), Bhatia et al. (2024)
	Coin flip	Hoffman and Spitzer (1985)
	Color chips	Butler (2014)
	Klee-Kandinsky	Paetzel and Sausgruber (2018)
Effort	Slider	Gill and Prowse (2019)
	Hash-mark game	Hoffman and Spitzer (1985)
	Encoding words / coding speed test	Feltovich (2019), Segal (2012)
Cognitive ability	Raven’s progressive matrices	Paetzel and Sausgruber (2018)
	Math/arithmetic task	Blue (2016), Demiral and Mollerstrom (2018)
	Summation task	Hett et al. (2020)
Quiz	Current event quiz	Hoffman et al. (1994)
	General knowledge quiz	Fleiß (2015)
	Economic trivia quiz	Ball et al. (1998; 2001)
Characteristics	Gender	Eckel and Grossman (2001), Campbell (2002)
	Name	Essen and Ranehill (2011)
	Age	Mertins et al. (2013)

A Further Results

Table 4: Wald-F Tests

Mechanism	Random		IQ		Slider	
	χ^2	$p(\chi^2)$	χ^2	$p(\chi^2)$	χ^2	$p(\chi^2)$
IQ	9.83	$p = 0.010$	-		-	
Slider	22.72	$p < 0.001$	3.99	$p = 0.274$	-	
Grade	35.91	$p < 0.001$	23.18	$p < 0.001$	6.15	$p = 0.079$

Table notes: The table shows the χ^2 statistics and significance levels p of a Wald-F test for equality of coefficients from Model 1 of the regression shown in Table 2 in the respective cells. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Tobit Regression: Rating Behavior

Rating Behavior	(1)	(2)	(3)	(4) only women	(5) only men
<i>Baseline: Random</i>					
IQ	-0.762*** (0.235)	-0.780*** (0.236)	-0.253 (0.403)	-1.019*** (0.282)	-0.125 (0.422)
Slider	-1.190*** (0.235)	-1.194*** (0.234)	-0.747* (0.398)	-1.422*** (0.282)	-0.756* (0.408)
Grade	-1.790*** (0.235)	-1.793*** (0.235)	-0.977** (0.401)	-2.211*** (0.281)	-0.960** (0.416)
Female		-0.208 (0.211)	0.468 (0.369)		
<i>Baseline: Random#Female</i>					
IQ#Female			-0.793 (0.496)		
Slider#Female			-0.677 (0.492)		
Grade#Female			-1.233** (0.494)		
Age		-0.015 (0.020)	-0.015 (0.020)	-0.027 (0.021)	0.034 (0.048)
<i>Baseline: Riskneutral</i>					
Riskaffine		0.368 (0.282)	0.368 (0.282)	0.150 (0.359)	0.657 (0.442)
Riskaverse		-0.090 (0.244)	-0.090 (0.244)	-0.322 (0.298)	0.228 (0.418)
Round dummies	No	Yes	Yes	Yes	Yes
Constant	4.876*** (0.176)	5.443*** (0.590)	5.034*** (0.617)	5.792*** (0.646)	3.872*** (1.300)
N	420	420	420	276	144

Table notes: Tobit regression with panel estimator. Rating from 1 (= completely not legitimate) to 7 (= completely legitimate) as dependent variable. Covariates: Female (0=no, 1=yes), Risk Attitude (Baseline=riskneutral). Baseline: random mechanism. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

STUDY: On the Descriptive Legitimacy of Status Differences

DOI 10.48548/pubdata-3703

Research Cluster Working Paper Series “Behavioral Economics and Societal Transformation”
ISSN 3055-1897

Version from: 12.06.2026

This working paper was produced within the context of the [Research Cluster “Behavioral Economics and Social Transformation”](#), a collaboration between seven universities in Lower Saxony (Clausthal University of Technology, Georg-August University of Göttingen, Leibniz University Hannover, Leuphana University Lüneburg, Carl von Ossietzky University of Oldenburg, University of Osnabrück, University of Vechta) and the Fraunhofer Institute for Applied Information Technology. As part of the “Wissenschaftsräume” funding line, the project is supported by the Volkswagen Foundation and the Lower Saxony Ministry of Science and Culture through the [zukunft.niedersachsen](#) program. Further information on the project and additional working papers can be found at [www.wissenschaftsraum-verhalten.de](#).

Contact:
Research Cluster “Behavioral Economics and Social Transformation”
Fon +49 4134 667 2938
Mail wissenschaftsraum-verhalten@leuphana.de

Janina Kraus and Stefan Traub

© 2026 This working paper is licensed under the Creative Commons CC BY 4.0 licence. For the full text of the licence terms, please visit the URL <https://creativecommons.org/licenses/by/4.0/>

Institute of Economics (IVWL)
Leuphana University Lüneburg
Universitätsallee 1
21335 Lüneburg