Research Policy 50 (2021) 104212

Contents lists available at ScienceDirect

Research Policy

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

Incentives, competition, and inequality in markets for creative production

Stefano Balietti [a,b], Christoph Riedl [c,d,e,f,*]

[a] Mannheim Center for European Social Science Research (MZES), A5, 6, Bauteil A, 68159 Mannheim, Germany
[b] Alfred-Weber Institute of Economics, Heidelberg University, Bergheimer Str. 58, 69115 Heidelberg, Germany
[c] Network Science Institute, Northeastern University, Boston, MA, United States
[d] D'Amore-McKim School of Business, Northeastern University, Boston, MA, United States
[e] Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States
[f] Institute for Quantitative Social Science, Harvard University, Cambridge, MA, United States

ARTICLE INFO

Keywords:
Reward stratification
Creativity
Innovation
Peer review
Market design
Competition
Inequality

ABSTRACT

Incentive structures and the intensity of competition play a key role in shaping the quality and direction of creative work. Organizing incentives as stratified rewards has emerged as a universal feature of modern society. However, this has implications for the producers and consumers of creative work that are not fully understood. We test the effects of reward stratification on producers, reviewers, and consumers of creative work by using data from a large online experiment of an artificial market for creative products. We find that competition induced by stratified rewards shapes the evolution of creative production. The quality of each tier in a stratified market is consistent with its position in the hierarchy. The top tier maintains high quality standards by attracting many submissions and then filtering its output, operating as an effective sorting device for budget-constrained consumers. However, reward stratification leads to higher levels of inequality and market exit among producers who fall behind in earnings, despite producing high-quality work. We discuss the broad implications of reward stratification across individual and market aggregate levels. This discussion contributes to the cumulative advantage debate in creative industries specifically, and to the creative aspects of scientific fields and industrial markets more broadly.

1. Introduction

Stratified reward structures have emerged as a recurring feature of many markets for creative products (Piketty & Saez, 2014; Rivera, 2011), resulting in an "age of hypercompetition" (D'Aveni, 1994). These structures are often designed—implicitly or explicitly—to keep elites in place and they contribute to growing inequality (Markovits, 2019). Reward stratification takes the form of market structures[1] with high and low payoff tiers, while a market that is not stratified does not distinguish between these tiers. Reward stratification has surged in creative markets such as art and music (Fraiberger, Sinatra, Resch, Riedl, & Barabási, 2018), innovation in business (Di Lorenzo & Almeida, 2017), and science (Azoulay, Graff Zivin, & Manso, 2011; Larson, Ghaffarzadegan, & Xue, 2014). Across those different domains, those who succeed are granted considerable rewards in terms of career advancement, reputation, and resources, in reinforcing cycles of cumulative advantage often termed a "Matthew effect" (Azoulay, Stuart, & Wang, 2014; Perc, 2014; Siler, Lee, & Bero, 2015).

However, we currently do not yet fully understand the complex and interrelated consequences of stratified reward structures on producers, reviewers, and consumers in creative production markets. Empirical evidence to better understand and learn about unintended consequences in markets with stratified reward structures is scant and mixed. In the context of science, for example, creativity seems to flourish when competitive pressure is removed (Azoulay, Graff Zivin and Manso, 2011), yet competition is also found to fuel productivity and innovation (Feller, 1996). Furthermore, there are concerns about the increasing inequality in recognition and access to resources between the haves and have nots (Markovits, 2019). There are signs that the intense competition for the few but highly valuable prizes in stratified markets can be damaging because they amplify self-interested behavior, lead to the loss of creativity, and increase conformism (Heckman & Moktan, 2020;

* Corresponding author.
E-mail addresses: stefano.balietti@mzes.uni-mannheim.de (S. Balietti), c.riedl@neu.edu (C. Riedl).

[1] By market structure, we mean the ensemble of institutions judging the novelty and appropriateness of creative products (Amabile, Hennessey, & Grossman, 1986) and either directly or indirectly distributing author credits in the form of reputational and/or monetary compensation.

https://doi.org/10.1016/j.respol.2021.104212
Received 3 February 2020; Received in revised form 15 January 2021; Accepted 19 January 2021
Available online 20 February 2021
0048-7333/

Seeber, Cattaneo, Meoli, & Malighetti, 2019).

While past work on the "Matthew effect" has often focused on how past success increases access to resources, less is known about how subjective evaluations of creative production may be affected by reward structures (Fortunato et al., 2018). How do stratified rewards affect the quality and trajectory of creative work and do they automatically lead to inequality? Are consumers of creative work better off if producers compete for stratified rewards? In this study we take a multi-level perspective to investigate how reward stratification affects producers, reviewers, and consumers in markets for creative products. We investigate how individual-level behavior affects feedback loops that lead to the emergence of different quality standards and creative production trajectories.

We conducted large online, group-behavior experiment contrasting two stylized market structures: flat versus stratified. Both experimental market structures are competitive and comprise three institutions awarding prizes via peer review evaluation of creative products. In the flat market, all institutions are identical in terms of the number and size of awards; in the stratified market, the institutions are tiered: the top-tier institution bestows only one large prize, the median institution two medium-sized prizes, and the bottom-tier institution four small prizes.

We recruited participants from Amazon Mechanical Turk and engaged them in an interactive, multi-round game that allows for the emergence of style norms through feedback loops between producers of creative work and peer review evaluation. Our experimental design has three advantages over both theoretical models and observational studies. (i) It offers direct causal evidence for the effect of reward stratification as the experimentally manipulated variable without other confounding variables. (ii) The setting allows us to precisely measure innovation and to investigate both market-level outcomes and individual-level behavioral mechanisms behind them. (iii) We create multiple, parallel markets of each reward structure, with each evolving completely independently from identical starting positions. By studying a range of possible outcomes across multiple market instances rather than just a single market, we provide reliable evidence for reward stratification as the underlying causal mechanism. In particular, we can measure inequality: the extent to which multiple markets with identical initial conditions and indistinguishable populations generate different outcomes due to reward stratification.

We find that competition induced by stratified rewards shapes the evolution of creative production. Ultimately the quality of each tier in a stratified market is consistent with its position in the reward hierarchy. We show that the top-tier institution in the stratified market maintains its high quality standards by attracting many submissions and filtering its output. This filtering is an effective and beneficial sorting device for budget-constrained consumers. However, and quite worryingly, stratification also generates higher levels of inequality among producers. It leads to higher market exit rates of producers who earn low payoffs but who are not lower skilled. The main insights are that some individual-level differences and the response to accumulated slack resources are amplified by the hypercompetition in the stratified market.

Given the feedback loops in creative production markets, policy interventions run the risk of generating unintended consequences. We embrace the perspective of a policymaker dealing with practical questions of how markets should be organized: we consider the economic value of competition alongside its potential social costs. This holistic perspective is important since there may be unintended consequences that only become apparent when looking at different stakeholders or at a different level of analysis.

A key contribution of our work is the investigation of mechanisms that underlie the market-level outcomes. Specifically, we investigate how feedback loops driven by sorting and the emergence of style norms can lead to increased market exit and inequality. We contribute to the ongoing conversation about the role of cumulative advantage (Azoulay et al., 2014; Jones, Wuchty, & Uzzi, 2008; Merton, 1968) and its effect on creative production trajectories (Bourke & Butler, 1999; Bozeman & Youtie, 2017; Braun, 1998; Choi, Lee, & Sohn, 2009; Wang, Lee, & Walsh, 2018). We show that rapid cultural adaptation can lead to inequality between social groups. We develop a new framework that links the emergence of distinct innovation trajectories, sorting, and market failures to inequality. This framework complements innovation policy literature (e.g., Heinze et al., 2009). It expands our understanding of the complex relationship between reward structures, creative output, and cumulative advantage in competitive institutional environments.

2. Theory and hypotheses

Reward and recognition systems have received much research attention, especially in sociology (Cole & Cole, 1974). The link between incentives and creativity has also been a research theme in different domains, including business and science. This research has often focused on comparing competitive and uncompetitive models of reward allocation and recognition (e.g., Balietti et al., 2016; Wang et al., 2018). However, we know little about how different forms of competition affect creativity (Azoulay et al., 2011). As a result, insights into the unintended consequences in markets with stratified reward structures are sometimes at odds with each other (Azoulay et al., 2011; Feller, 1996; Wang et al., 2018). It is unclear how different competitive systems affect the level and direction of creative work. A better understanding of the differences could be important to design policy for markets for creative cultural product, innovation in business, and science.

Research has also focused on different levels of analysis (e.g., Heinze et al., 2009). However, it is still unclear how the different levels interact. This is critical as there is an intricate interplay between producers, reviewers, and consumers. For example, in the context of science, while individual producers may care about receiving adequate compensation and recognition for their creative output, society as the consumer may care more about the quality and novelty of output produced collectively. Research has shown that the decisions made by individual scientists may not maximize collective discovery (Rzhetsky, Foster, Foster, & Evans, 2015). This points to the risk of unintended consequences. A policy that is beneficial for one set of stakeholders could have negative effects on another. Only research that investigates interrelationships will help us gain a better understanding of the advantages and disadvantages of different policies.

Here, we examine the effect of reward stratification on outcomes for producers and consumers alike. We focus specifically on the interrelationship between key market-level outcomes and the underlying individual-level behavioral mechanisms. We thus offer a more holistic analysis and focus specifically on the dynamic aspects in which these outcomes emerge and how they are interconnected (Fig. 1). In the next sections we review prior literature, broadly drawing on creative production fields including science, business, and cultural products such as art. We formulate hypotheses regarding (a) emergent market-level outcomes, (b) the behavior of producers, and (c) the behavior of reviewers. For the interested reader, the Appendix provides additional background on aspects of cultural evolution and our research design that cuts across levels of observation.

2.1. Market perspectives on the generation and selection of creative work

2.1.1. Cultural evolution

Research has shown that competitive rewards encouraging scientists to take more risks affect the direction of innovation (Azoulay et al., 2011; Jaffe, Trajtenberg, & Henderson, 1993). That is, the competitive nature of a market shapes the cumulative and cultural evolution of creative production (Boyd & Richerson, 1988). This suggests that creative work may take different trajectories in markets that differ in their level of competitiveness. If producers under a given reward structure were to make systematically similar decisions—such as highly risky leaps that are similar to the risky leaps by others in some way—even parallel instances of those markets could reliably evolve into similar

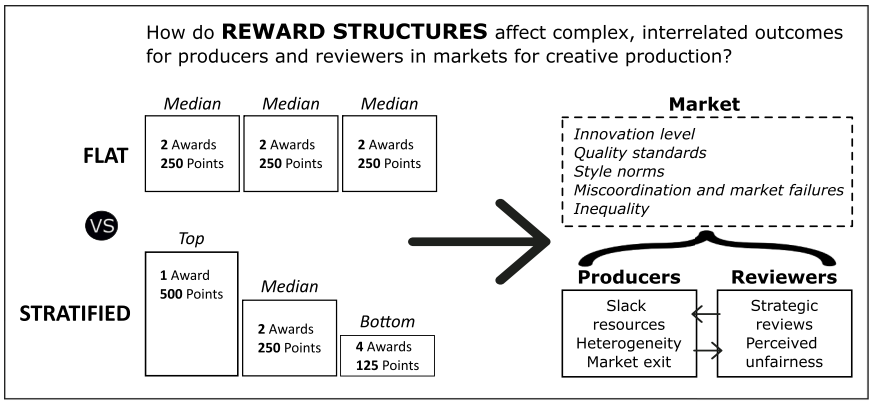

Fig. 1. Illustration of treatment conditions and research question. We perform a "parallel-worlds" experiment in which we vary the market reward structures: flat vs. stratified. Both worlds comprise three institutions, but the number and the size of the awards bestowed by the institutions differ across conditions. How do producers and reviewers of creative content adapt their actions in each world? We measure individual outcomes (solid-line boxes) and macro-market outcomes (dashed-line box).

market specializations. They may repeatedly produce similar creative trajectories. In our specific context, which concerns the creation of visual creative work, one such decision could be the systematic tendency of innovators who take more risk to produce more abstract work (compared with work that is a more accurate depiction of visual reality). Such a tendency would be consistent with the sociology literature of professions linking abstraction to prestige (Abbott, 2014). We predict that markets governed by different reward structures will produce detectable differences in the direction of their cultural evolution.

Hypothesis 1. The market with stratified reward structures will produce distinctly different creative work than the market with flat reward structures.

2.1.2. Cumulative advantage and inequality

An extensive body of work examines how recognition is accumulated. In the context of science, for example, this research has repeatedly demonstrated a high degree of social and institutional stratification (Azoulay et al., 2014; Merton, 1968). Most notably, the work by Merton (1968) has shown that stratification in the scientific community is a result of cumulative advantage allowing the rich to get richer over time. Similar results have been documented in art (Fraiberger et al., 2018) and patents (Perc, 2014).

Inequality tends to be extremely high as a result of such stratification. Some individuals accumulate much higher recognition (e.g., citations, exhibitions in reputable galleries) than the average individual. While it is well established that self-organization and preferential attachment can explain the observed patterns on the aggregate level it is still largely unclear which micro foundations cause cumulative advantage to emerge on the aggregate level. It is also unclear why many market structures operate in a way that seems to amplify cumulative advantage rather than dampen it.

Recent insights suggest that in the presence of skewed payoffs, cumulative advantage and high degrees of inequality can be generated by random fluctuations alone (Bol, de Vaan, & van de Rijt, 2018; Perc, 2014; Pluchino, Biondo, & Rapisarda, 2018), even in the absence of individual differences in talent. Conversely, high degrees of inequality are less likely and less extreme when recognition is distributed more evenly (Pluchino et al., 2018) or when individuals can move between institutions more freely (Foley, Smead, Forber, & Riedl, 2020). We consequently expect to see higher levels of payoff inequality in the market with stratified rewards, and when additional mechanisms are in place that prevent innovators from freely moving between institutions.

Hypothesis 2. Market-level inequality will be higher in the market with stratified reward structures.

While the key focus of our work is the study of emergent, market-level outcomes, it is clear that those emergent outcomes must be grounded in individual-level behavior (Felin & Foss, 2005). We theorize that three key individual effects are drivers behind these emergent differences between stratified and flat markets: the role of slack resources, heterogeneity, and strategic behavior of reviewers.[2] Hence, we formalize three directly testable hypotheses of individual behavior for producers and one for reviewers.

2.2. Producers

A stratified market structure leads to higher rates of market exit according to insights from contest theory in economics (Lazear & Oyer, 2012). Two mechanisms that have been identified as leading to a "discouragement" effect are larger prize spreads between winners and losers, and comparative feedback (Fershtman & Gneezy, 2011). Lagging behind in terms of accumulated rewards, not necessarily in terms of talent, can lead to market exit. Consequently, we expect to see higher rates of market exit in the stratified market than in the flat market.

Hypothesis 3. Producers will be more likely to exit the market under stratified reward structures.

The relationship between slack—the presence of accumulated resources from past performance that exceed the target—and innovation has been of major focus in organization and innovation research (Cyert & March, 1963; Levinthal & March, 1981; Nohria & Gulati, 1996; Wiseman & Bromiley, 1996). The relationship is complex, with slack linked to both increased experimentation and higher innovation (Cyert & March, 1963; Levinthal & March, 1981), but also to lower innovation caused by less exploration and more exploitation (Lavie, Stettner, & Tushman, 2010). This ambiguity suggests that there are some missing mechanisms to moderate the effect.

We theorize that the relationship between slack and innovation can depend on the reward structure under which slack resources can be employed. Specifically, the reward structure affects risk-taking behavior by providing high-risk/high-reward opportunities. The ability to pursue high rewards has a positive influence on risk-taking (see Bromiley (1991) and also Cyert and March (1963), who call this effect "problemistic search"). In the flat market, all institutions are undifferentiated by design. However, in the stratified market, there is the possibility to follow a high-risk/high-reward strategy as the top-tier institution promises significantly higher rewards. This possibility of high returns in the top-tier institution can compel producers to take more risk when they have slack resources. Without opportunities to pursue a high-risk/high-return strategy, slack resources alone would not compel producers to take more risks.

Hypothesis 4. Producers with slack resources will produce more novel output under stratified rewards than producers with flat rewards.

Creative cognition is our theoretical base for predicting the influence of reward stratification on producers' information-processing strategies

[2] We do not claim that these are the only drivers; however, they have received considerable attention in the literature and they can be empirically tested in our experimental setup.

during a generative task (Finke, Ward, & Smith, 1992). Creative cognition theory suggests that certain mental and contextual states enhance or hamper an individual's cognitive capacity to act creatively. Specifically, the competitive pressure and larger prize spread in a stratified reward market could reduce some producers' ability to access knowledge and form new associations when generating new ideas. Many people experience competitive contexts as stressful (Buser, Dreber, & Mollerstrom, 2017). However, the degree to which this stress affects individuals is unlikely to be uniform. We theorize that the cognitive capacity to act creatively will be higher among individuals who have a preference for—or who are more familiar with—competitive environments, or who are otherwise better equipped to deal with the stressful situation (Kahneman & Riis, 2005).

First, we theorize that individuals who are more highly skilled at the task have more cognitive capacity to access knowledge, form new associations, and produce more creative work when under competitive pressure (Boudreau, Lakhani, & Menietti, 2016). Second, we theorize that younger producers who have lower risk aversion (Grubb, Tymula, Gilaie-Dotan, Glimcher, & Levy, 2016) should produce more creative work in the stratified market than older innovators. Third, we theorize that intrinsically motivated innovators will be better able to handle the competitive environment, as their motivation depends less on extrinsic rewards (Shalley, Zhou, & Oldham, 2004). Third, we do not expect to see strong gender differences. Given the absence of direct social influence or social status due to the online nature of our experiment, we expect women to feel at ease with competing against males and hence expect no gender differences (Galinsky & Schweitzer, 2015).

Hypothesis 5. Producers who are (a) highly skilled, (b) young, and (c) intrinsically motivated will produce more novel work in stratified markets than producers who are lower skilled, older, and extrinsically motivated. In addition, the novelty of the work produced by women will be no different than that produced by men in the stratified market.

2.3. Reviewers

Market structures will affect producers of creative work also when they act as peer reviewers. Specifically, we theorize that the increased competition for rewards in the stratified market will induce reviewers to behave more strategically in their reviews. Strategic behavior in contests has been derived theoretically (Teplitskiy, Acuna, Elamrani-Raoult, Körding, & Evans, 2018) and studied empirically (Balietti et al., 2016; Harbring & Irlenbusch, 2008). It can include the sabotage of competitors' work during the peer review process so that it appears to be lower quality. This self-interested behavior then increases the peer reviewer's own chance of winning the competition. A larger prize spread between the top and bottom prizes is expected to attract more sabotage (Harbring & Irlenbusch, 2011).

Hypothesis 6. Reviewers in markets with stratified reward structures review more strategically.

3. Experimental design and method

This paper sets out to experimentally test the complex relationship between reward structures, creative output, and social stratification. Our experiment to test our hypotheses has three key components:

1 a real-effort, high-dimensional task for the generation of creative products whose compositional features can be precisely measured;
2 a quality evaluation mechanism that is based on human cognition and therefore involves a degree of uncertainty—rather than being algorithmic and deterministic;
3 a market that offers sufficiently complex dynamics to enable feedback loops between producers and reviewers, emergent behaviors, evolving style norms, and unintended consequences.

The majority of past experimental research into creative production is not adequate for our purposes because those studies lack the market dynamics and social feedback loops we require (see the Appendix for more details).

To reap the rewards of successful innovation, creators must not only focus on how technically sound or novel an idea is, they must also consider its presentation, reaching the right audience, carefully listening to the feedback from evaluators, etc. We believe that this interplay is a key component of real-world, creative production markets; therefore, we based the design of our experiment on a variation of the Art Exhibition Game (Balietti et al., 2016).

3.1. The Art Exhibition Game experiment

The Art Exhibition Game is a minimalistic and well-controlled experimental paradigm that reproduces some of the relevant features of scholarly peer review in a real-effort, interactive, creative game. Its main focus is the tradeoff between the exploration of new risky solutions and the exploitation of known successful ones while being exposed to peer evaluations. In addition to the behavioral data collected during the experiment, we also collected demographic data, social preferences, affective state using pre- and post-experiment surveys (see Appendix).

3.1.1. Game dynamics

The game is played in groups of nine players. Each player is both a producer and a reviewer, and the game is repeated for 12 rounds of creation and evaluation. Each round is composed of the following four stages:

Creation. Participants create parametric images using an interactive interface (Fig. 2 left). The interface allows participants to create modified versions of Chernoff faces (Chernoff, 1973). However, they are not confined by the interface to create images that resemble faces. They are free to experiment with different shapes and sizes, enabling the creation of other objects, and even completely abstract art. The total number of possible creative outcomes is 5.2×10^{43}.

Submission. Participants choose a virtual exhibition to submit their work. Exhibitions may follow different criteria for accepting a submission and for rewarding successful producers. Participants can also observe the history of previous creative work accepted by each exhibition and use this information to guide their submission decision.

Peer review. Participants as reviewers evaluate the quality of three randomly selected submissions on a scale of 0–10. Each submission receives three reviews. Reviewers are informed about the exhibition they are reviewing for, so that they can also judge whether the work fits into a given exhibition.

Dissemination. The review scores are averaged; the best submissions are put on display and generate a monetary reward for their creators. All successful submissions are shown along with their average review score, grouped by institution, and ordered by rank. Rejected submissions are never shown.

Every new round continues from where each participant has left off. This allows an incremental exploration strategy to be pursued and for previously rejected work to be submitted to other exhibitions. Exploratory leaps, e.g., paradigm shifts, are possible by directly copying any of the previously displayed images (one's own or those of others). The history of past work accepted in each institution is visible on the right of the creation interface (Fig. 2 right).

3.1.2. Treatment conditions

We implemented a between-subjects experimental design in which we varied the market structure to which participants were assigned: flat vs. stratified. We fixed the number of available institutions in both markets and varied the reward and acceptance criteria. The flat market structure comprised three institutions with identical acceptance criteria and reward function: each participant could accept up to two creative works and be rewarded with 250 tokens. The stratified market structure also comprised three exhibitions; however, they had a tiered set of

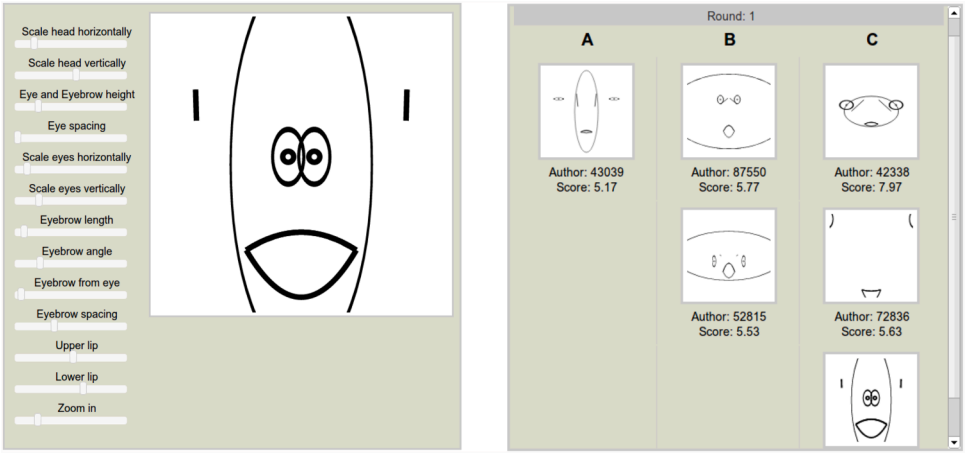

Fig. 2. Game interface to create new images. The interface allows the user to move a set of sliders on the left-hand side to modify the image at the center in real time. On the right-hand side of the screen, participants would see the history of previous rounds, divided by exhibitions and ranked by score. By clicking on an image, participants can enlarge it and eventually decide to use it as a starting point of a new creation.

outlets. The top tier accepted at most one creative work per round and awarded a large reward of 500 tokens the creator; the median tier accepted up to two creative works that were rewarded with 250 tokens each; finally, the bottom tier accepted up to four works and rewarded creators with 125 tokens each.

In both treatments, the slots available for display in all institutions were assigned via a rank–order tournament. That is, relative performance was the determining factor rather than absolute quality criteria. After the peer review stage, all works were ranked by their average review score from highest to lowest within the exhibition to which they were submitted. Beginning from the top and going down the rankings, works were accepted as long as there still were available institutional slots. This means that unsuccessful work at one institution could be successful at a different institution. Likewise, work that received a high score may have been rejected because other work submitted to the same institution received even higher scores. All institutions in our setup are fully competitive and award bonuses to participants through tournaments instead of threshold-based mechanisms.

3.2. Procedures and subject pool

We recruited subjects for our experiment from Amazon Mechanical Turk. AMT is a platform where Human Intelligence Tasks (HITs) can be posted for evaluation by registered users in exchange for a monetary compensation (Mason & Suri, 2012). AMT has been extensively used in (and validated for) academic research; for example, in opinion surveys, group-behavior experiments, and cognitive tasks (Horton, Rand, & Zeckhauser, 2011; Mason & Suri, 2012). Data quality can be an issue when recruiting participants on AMT. Those issues are most severe for very simple tasks such as classifying images, that can attract non-human workers ("bots"). Participation in our behavioral experiment was difficult to automate. We carefully checked our data and did not find suspicious data such as players who only performed random clicks or otherwise appeared to not participate in good faith.

We recruited 398 participants, and 329 completed all stages of the experiment. The remainder dropped out. The analysis in this paper is restricted to groups that had 0 dropouts, i.e., 10 groups for the flat market structure and 9 groups for the stratified structure. There was a total of 171 participants in these groups. These attrition numbers are in line or even slightly better than the interactive online experiment findings by Arechar, Gächter, and Molleman (2017). Overall, the experimental task lasted between 30 and 40 minutes including waiting times, with an average compensation of 3.8 USD. Our experimental protocol was approved by Northeastern University and all participants gave informed consent. After the experiment, we collected unbiased quality appraisal of all images created in the experiment outside the incentive structure of the markets and independent of feedback loops. We recruited 558 independent evaluators through the online labor market Amazon Mechanical Turk (see Appendix for details). Both the behavioral experiment and unbiased quality appraisal were implemented with the nodeGame framework (Balietti, 2017).

3.3. Statistical analysis

We relied on several different analysis methods. Given randomized assignment of individuals to treatments, our primary hypothesis tests relied on simple t-tests and ANOVA to compare market-level outcomes across the two treatments. For several key tests, we also report p-values based on non-parametric permutation tests, which avoid making assumptions about the distribution of data. This is because observations in group-behavior experiments can be correlated, with the risk of p-values being inflated. For the same reason, we mainly used linear mixed models (LMM) with the participants and sessions as the random effects in our analysis of individual-level effects. Where appropriate, t-test and permutation tests were used in conjunction with results from computer-simulated null models: models in which a certain baseline behavior is assumed so that we can detect deviation from those baselines.

We selected our sample size based on a power analysis that used the effect size $d = 0.38$ in previous Art Exhibition Game research (Balietti et al., 2016). We targeted 11 replications for each of the two treatment conditions, i.e., 108 participants per group for a power of 0.79 to detect a true effect with a t-test. We stopped at 10 and 9 replications per treatment condition before looking at any data, due to an increase in the data collection costs caused by the dropouts. This amount of data still left us with a power greater than 0.80 to detect a true effect with a chi-squared test.

4. Results

Reward stratification significantly affects outcomes for consumers, producers, and markets. In the following sections, we report results on each of those levels, highlighting interdependencies and heterogeneous effects where they exist. Given that our analyses focus on different key outcomes—innovation, inequality, market exit, and strategic behavior—across both individual and market levels, we present the most relevant aspects of our analyses in this section and occasionally refer to additional analyses in the Appendix.

4.1. Market outcomes

4.1.1. The emergence of quality standards as a result of filtering

We find that markets do define the quality of their creative output. While the average independent quality ratings of published images are lower in the stratified market (4.62 vs. 4.93; t-test $p < 0.001$), we find significant variations across the tiered institutions in that market (ANOVA $df = 2$, $F = 42$, $p < 0.001$). The top-tier outlet of the stratified market also delivers the highest quality across both treatment conditions (LMM $\beta = 0.61$; $p < 0.05$; Fig. 3A; Table 11 in the Appendix). In order to be published in the top-tier institution, creative work needs a quality score that is 11% higher than in the median tier, and 24% higher than in the bottom tier (LMM $\beta_m = -.63$, $\beta_b = -1.42$; $p < 0.001$; Fig. 3B; Table 6 in the Appendix). We explored four possible explanations for the emergence of different quality standards between different outlets: (1) producers submit their best work to the top tier; (2) reviewers assume that works submitted to the top tier are inherently better quality (regardless of whether this is actually true); (3) more producers submit to the top tier thus increasing the set of submissions to choose from; and (4) the top-tier institution increases the quality of its output by limiting the number of published submissions (see Appendix for details). We tested several different product quality metrics (review scores, independent quality ratings, and separate measures of innovation and exploration) but found no significant quality differences along any dimension). In sum we found that market participants do not target the top-tier institution with higher-quality work, and we found no evidence that reviewers give higher review scores to work in the top tier. Instead, we found the top-tier institution boosts the quality of its publications by both attracting many submissions and very selectively filtering its output. This process is so effective that the quality of the top tier is significantly higher than the average production in the flat market.

The quality stratification in the tiered market directly affects consumers of creative work with budget constraints. Despite the slightly higher overall quality in the flat market, a budget-constrained consumer (e.g., in terms of time to read or admire creative work) is actually better off in the stratified market which identifies the highest-quality work thanks to its systems of tiers. To analyze those differences, we performed computer simulations of consumer behavior in both markets and found that the stratified market delivers higher-quality output to consumers who can consume up to three units of work per round (Fig. 4; see Appendix Section A.7.1 for details on simulation method).

This finding points to an inherent tradeoff in the stratified market: while it helps consumers with constrained budgets to identify the best work, consumers with a large budget who consume more works than the top and median tiers publish actually end up with lower quality than if they choose randomly among published work in the flat market. This is due to market inefficiencies resulting from miscoordination: some high-quality works that should have been published—runner ups in the top tier—were not published at all, while some overall lower-quality works were published in the bottom tier instead.

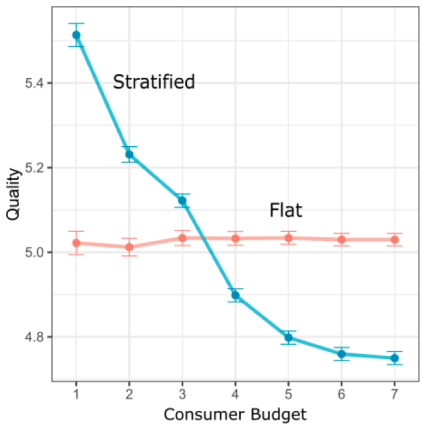

Fig. 4. Budget-constrained consumption. Average quality of published images for a budget-constrained consumer. Error bars 95% CI.

4.1.2. Innovation and emergence of style norms

Looking beyond quality, we see differences in innovation and distinct styles of the creative products in the two different markets. First, we find a higher degree of innovation in the stratified market (Fig. 5A). While innovation increases over rounds in both markets, it is significantly higher in the stratified market overall (t-test $p < 0.001$; non-parametric permutation test $p < 0.0001$). This result is not statistically significant and somewhat more nuanced when looking at regression results, which might be due to heterogeneity in preferences in the population of producers for a given market structure, as we show in the next section. Higher innovation is the result of a specialization in abstract styles in the stratified market. We investigate two drivers behind this trend toward abstractness—heterogeneity and slack—in Section 4.2.

Second, we find that markets shape what their participants produce in terms of the creative trajectory of the actual works. To quantify the differences, we trained a machine-learning random forest classifier to predict whether a product was created in the flat or stratified market structure. This prediction was based solely on product features. We reached an average predictive accuracy of about 78%. Strikingly, we can predict the market structure in which a work was created with about 69% accuracy using data from only the first three rounds (Fig. 5B). This finding supports our Hypothesis 1. That is, over time, different market structures develop their own unique styles that become stronger as time goes on. In the flat market, market specialization seems to be driven more by images resembling faces and in the stratified market more by

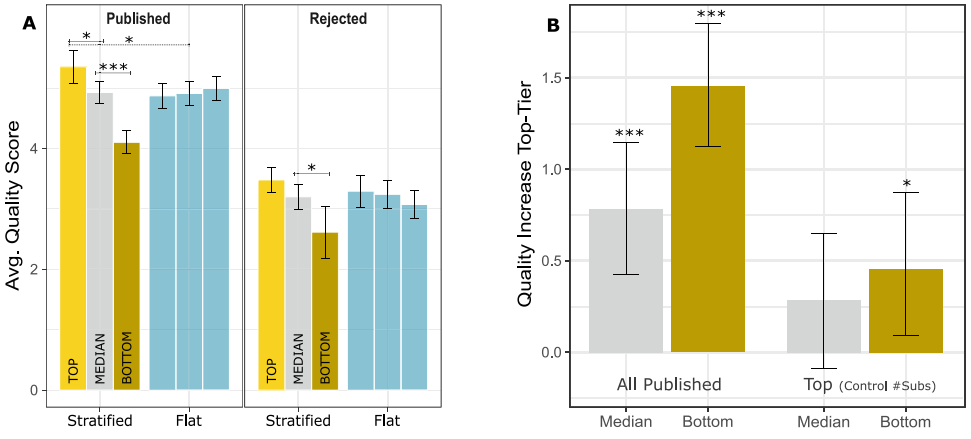

Fig. 3. Outlet Stratification. A) Average quality scores across markets for published and rejected images (Table 11 in the Appendix). Dotted lines test the difference between the top tier and the entire flat market. **B)** Model estimates of the quality differential of images published in the top tier compared to the median and bottom tier: (left) for all images, (right) just the best image, controlled number of submissions (respectively models Pub and Rank1(Sub) in Table 10 in the Appendix). Stars indicate significance level of the regression coefficients: ***=0.001, **=0.01, *=0.05. Error bars 95% CI.

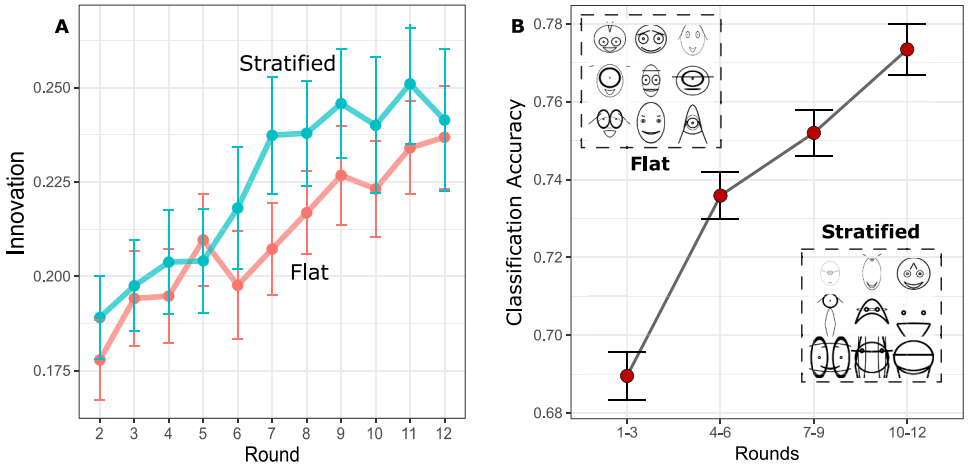

Fig. 5. Market specialization. A) Innovation over rounds in the stratified and flat market structures. **B)** Accuracy of random forest classifier predicting the market structure in which an image was created. The classifier used only image features, no extra information. Error bars 95% CI.

abstract images. The top-tier institution is the preferred venue for more abstract images (LMM $\beta = .23$; $p < 0.05$; Table 5 in the Appendix). Results are marginally significant for more creative images (LMM $\beta = .19$; $p < 0.1$). This suggests that rapid cultural adaptation can lead to persistent differences between social groups.

4.1.3. Miscoordination and market failure

The filtering in the stratified market impacts on the allocative efficiency of the market. We find substantial congestion in the stratified market, as producers fail to coordinate their submissions across the different institutions and good work is not published. In equilibrium, the flat market assigns six awards and the stratified market seven awards. Hence, coordination should be easier in the stratified market, but we find the opposite (Fig. 6). On average, an additional 1.54 producers in the stratified vs. 0.43 in the flat market do not earn any payoff per round, as compared to their respective static equilibria prediction (LMM $\beta = 1.10$; $p < 0.001$; Table 19 in the Appendix). The number of producers without payoffs is significantly higher in the stratified market. As expected, congestion in the stratified market is highest in the top-tier institution (LMM $\beta = -.39$; $p < 0.05$) and minimal in the bottom tier (LMM $\beta = -3.17$; $p < 0.001$). This miscoordination results in the average payoff for participants in the stratified market being significantly lower: 1,856 vs. 1,718 points (Wilcoxon rank sum $p < 0.05$). It is a primary reason for producers to exit the market altogether (see analysis on market exit in Section 4.2).

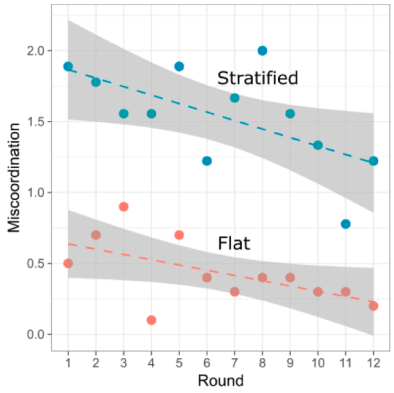

Fig. 6. Market miscoordination over rounds. Miscoordination is measured as the actual number of participants who fail to get their creation accepted in any exhibition minus the theoretical predictions under perfect coordination in either market. Both markets show positive levels of miscoordination; however, the level is significantly higher in the stratified market; the trend is decreasing over rounds for both markets. Shaded area 95% CI.

The miscoordination directly leads to another type of market failure. The stratified market has a significantly higher rate of type II errors (i.e., not publishing higher-quality contributions; GLMM $\beta = .67$; $p < 0.001$; Table 20 in the Appendix), and there is only weak evidence that it lowers the rate of type I errors (i.e., publishing lower-quality contributions). For example, the second-best image submitted to the top tier could have been published in the median tier 44% of the time, and in the bottom tier in 83% of the time. It could have been published in at least one of those two tiers 100% of the time. This directly explains the lower quality work available for consumers with larger budgets in the stratified market.

4.1.4. Inequality

We find higher inequality in earnings in the stratified market where miscoordination, market failures, and sorting combine with the skewed distribution of rewards (Fig. 7A). The flat market has a more symmetric distribution (skewness = 0.03), while the stratified market's distribution is more skewed with a long right tail. This indicates that some stratified market producers earn very large payoffs (skewness = 0.64; KS test $D = 0.29$, $p < 0.05$). We found that the average Gini coefficient—which quantifies inequality—per session is significantly higher in the stratified market: 0.19 vs. 0.13 (t-test $p < 0.05$; non-parametric permutation-test $p < 0.05$). This finding lends support for our Hypothesis 2.

To better understand the mechanism generating inequality in earnings, we relied on a computer simulation to generate a null model of the earning distribution we would expect to see if producers were to randomly submit their work to one of the tiers (see Appendix). In sum, we found that the increase in inequality in the markets was driven by a high degree of sorting, i.e., the average number of consecutive submissions to the same institution (see also Fig. 7B). Instead of submitting their work to different institutions in consecutive rounds, producers tend to submit to the same institution repeatedly.

Interestingly, sorting in the flat market null model is negatively correlated with inequality, suggesting a beneficial effect of coordination across institutions. That is, the flat market structure has the tendency to increase equality despite rising sorting in the later rounds. Persistent inequality in the flat market can arise only from differences in talent (or strategic behavior), not from sorting alone. Conversely, the increased sorting over time in the stratified market increases inequality. As the tiered outlets in the stratified market develop their own cultural styles, switching between institutions becomes more difficult. This increases the level of overall sorting, which increases inequality.

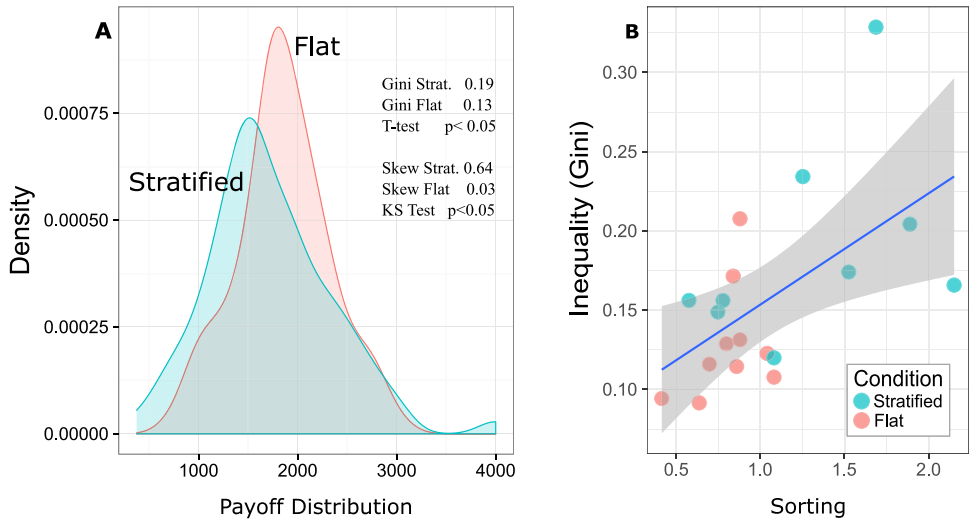

Fig. 7. Market inequality. A) Payoff distributions in the stratified and flat markets. The stratified market generates a significantly more unequal distribution, with lower mean payoff, and a significantly higher average Gini coefficient. **B)** Scatterplot sorting and inequality in experimental sessions. Sorting is the average number of consecutive submissions to the same institution and it is positively correlated with inequality in both markets. Each dot is a session. Shaded area 95% CI.

4.2. *Individual-level effects of producers*

The previous section established several differences in emergent properties such as innovation, style, and reward inequality between a market governed by flat versus stratified rewards. In this section we investigate several individual-level effects to explain the micro foundations that lead to these differences. Specifically, we explore (a) market exit, (b) slack resources, (c) heterogeneity, and (d) strategic behavior. We show key regression results in Table 1, which probes the interaction effects between a dummy indicator for the market condition and different independent variables (Table 1 in the Appendix shows several alternative specifications).

Market exit of producers. The number of participants in our experiment was fixed, but being an online experiment, anyone could quit at any time. If the market structure had no influence on quitting behavior, we would observe about the same rate of dropouts across the two markets. However, this is not the case. The stratified market structure had more than three times the rate of sessions with at least one dropout: 0.5 vs. 0.15 (t-test $p < 0.05$). In the Appendix we investigate the causes of this behavior and find that it is entirely endogenous and driven by payoff differential: participants left behind in the reward distribution are those who are more likely to exit. Overall, this finding supports our Hypothesis 2.

Previous success impacts innovation. We find that slack resources led to a 9-10% lower average level of innovation in both markets (LMM $\beta =$ -.02; $p < 0.001$; Model (1) in Table 1). However, the average decrease is only about half as large in the stratified market (significant interaction between slack resources and indicator dummy for market condition $p <$ 0.05). This supports our Hypothesis 4. Furthermore, this finding suggests that while slack resources act as an equalizing force in the flat market—those who were successful in the past are less innovative in the future—they have a much weaker effect in the stratified market. Slack resources allow individuals in the stratified market to remain innovative and gain additional advantage in the future from their past success. They are a critical driver behind high inequality in the stratified market by generating a self-reinforcing, rich-get-richer effect.

Heterogeneous responses to stratified rewards. Next, we turn to the analysis of heterogeneous treatment effects (models (2) through (5) in Table 1). We find that innovators of (a) higher skill (Model (2)), who are younger (Model (3)), and who are socially motivated (Model (4)) do better in the stratified market. We find no significant differences between male and female innovators (Model (5)). The fact that we do not find gender differences confirms our theory that the online nature of our experiment and the absence of direct interaction with male participants creates an environment in which female innovators are just as innovative under increased competition. Together, all models support our Hypothesis 5.

4.3. *Individual-level effects of reviewers*

We first investigated whether the stratified reward structure affected the degree to which reviewers behaved strategically. We define a review as being "strategic" if it delivers an extremely low score to a direct competitor in the same institution, i.e., less than 0.5 out of 10 (Balietti et al., 2016). This systematic low evaluation of direct competitors can be considered sabotage (Riedl, Grad, & Lettl, 2020). Overall, we find about the same level of strategic reviews in both markets, but in the stratified market these reviews are concentrated in the top tier (GLMM $\beta_m = -.56$; $\beta_b = -.56$; $p < 0.05$; Table 13 in the Appendix). This suggests that while reviewers do not generally behave more strategically in the stratified market, they do review strategically when reviewing for the top tier when they have submitted to the top tier in the same round themselves. This evidence provides support for Hypothesis 6.

The strategic review behavior is mirrored in the producer perception of review fairness. Participants in the post-experiment survey reported that the top-tier institution was by far the least fair (Fig. 8; GLMM $\beta =$ -.1.74; $p < 0.01$; Table 12 in the Appendix). This perception was influenced by the higher rejection rate, but also by the higher number of "strategic reviews." Together the two effects underscore that the stratified marked not only is more competitive, but also feels more competitive. This sense of competition and unfairness likely contributes to the higher market exit rate in the stratified market.

5. Discussion

In this discussion, we summarize (see Table 2) and synthesize our findings on how the different levels—individual and market aggregates—work together to affect the quality and direction of creative work produced in a stylized, reward-stratified market.

The quality of each tier in a stratified market is consistent with its position in the payoff hierarchy. Producers neither create higher-quality work in the stratified market nor specifically target the top-tier institution with higher-quality work. We find that stratification of institutions is not the direct result of individual-level behavior but emerges on the

Table 1
Multilevel regressions with player and session as random effects for slack resources and heterogeneous innovation levels among producers. Additional regressions are available in the Appendix.

Dependent Variable:	Innovation				
	Slack	Heterogeneous treatment effect			
	(1)	(2)	(3)	(4)	(5)
Stratified	0.02				
	(0.02)				
Slack × Stratified	0.01*				
	(0.00)				
Skill × Stratified		0.23**			
		(0.08)			
Skill × Flat		0.15			
		(0.08)			
Age ≤ 30 × Stratified			0.03***		
			(0.01)		
Age ≤ 30 × Flat			−0.00		
			(0.02)		
Age > 30 × Flat			−0.00		
			(0.02)		
Prosocial × Stratified				0.02*	
				(0.01)	
Prosocial × Flat				−0.00	
				(0.02)	
Individualistic × Flat				−0.02	
				(0.02)	
Female × Stratified					0.00
					(0.01)
Female × Flat					−0.01
					(0.01)
Individualistic	−0.02**	−0.02**	−0.02**		−0.02**
	(0.01)	(0.01)	(0.01)		(0.01)
Age > 30	−0.01*	−0.01*		−0.01*	−0.01*
	(0.01)	(0.01)		(0.01)	(0.01)
Skill	0.19**		0.19**	0.18**	0.19**
	(0.07)		(0.07)	(0.07)	(0.07)
Female	−0.00	−0.00	−0.00	−0.00	
	(0.01)	(0.01)	(0.01)	(0.01)	
Slack	−0.02***	−0.01***	−0.01***	−0.01***	−0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	0.16***	0.17***	0.16***	0.16***	0.17***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Controls	Yes	Yes	Yes	Yes	Yes
Log Likelihood	2982.43	2986.14	2984.94	2981.05	2983.93
Num. obs.	1848	1848	1848	1848	1848

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

aggregate level through the process of attracting many submissions and then filtering the output.

The reward structures affect not only the quality of what is produced but also the style—the trajectory of innovation—through cultural evolution. As producers repeatedly submit their work to the top tier, a distinct style of work emerges. This shows that the level of competition affects not only the absolute level of novelty and quality, but also the direction of innovation. Repeated submissions build on each other and evolve a distinct style over time.

The distinct style of the top tier within the stratified market is characterized by more abstract work. The emergence of more abstract work in the most competitive institutions suggests that producers may attempt to sidestep competition on traditional quality metrics—incremental improvements—and use a high-risk/high-reward strategy instead. In the flat market on the other hand, no differences emerge between outlets as producers have no payoff reason to differentiate.

The targeting and filtering of the top tiers in a stratified market has two positive consequences. First, filtering enables budget-constrained consumers to identify and consume the highest-quality work by efficiently allocating their budget. Second, we see slightly higher levels of innovation (but not necessarily higher levels of quality overall). On the other hand, we find a number of negative consequences for producers and society—including consumers without budget constraints. The increased competition for publication in the top-tier institution leads to miscoordination with more high-quality being rejected (more type II errors), an overall lower quality of published work, lower overall earnings for producers, and high inequality of earnings.

The higher reward inequality in the stratified market emerges as a result of market inefficiency and producers targeting the same institution repeatedly. The emergent style norms conversely make it difficult for producers to move between institutions, which could otherwise help

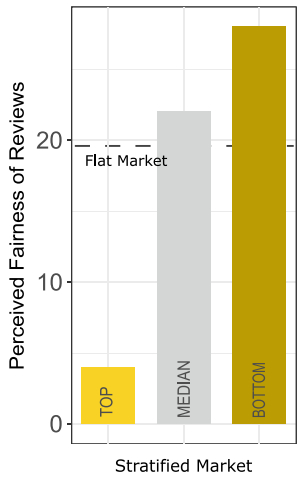

Fig. 8. Perceived fairness of reviews. Response counts to the question "In which exhibition did you feel the review scores were more *fair*?" Dashed line average for flat market. Error bars 95% CI.

alleviate the miscoordination. The sorting of producers to institutions and the miscoordination leads to few (but high payoffs), increasing inequality. An increasing number of producers fall behind, which ultimately causes them to exit the market.

We find that two micro level effects contribute to inequality in the stratified market. First, successful producers who accumulate slack resources remain more innovative than producers without slack resources, which provides them with cumulative advantage. Second, the increased competition amplifies individual-level differences. Some innovators thrive in the competitive environment, while others fall behind. In addition, the harsh competition for the scarce but high payoff of the top-tier outlet fuels strategic behavior during the peer review process. This instills a feeling of unfairness among producers and likely contributes to higher exit rates.

Our results indicate that reward stratification alters both the market and the behavior of participants in complex, interconnected ways and in multiple dimensions. These dimensions have previously been mostly studied in isolation. We aimed to bring notions from the social processes of science (Fortunato et al., 2018), cultural evolution theory (Boyd & Richerson, 1988), and the field of labor economics incentives (Lazear & Oyer, 2012) into our study.

Our results show that competition, peer evaluation, market coordination, and creativity do not act in isolation. Complex, interrelated outcomes emerge endogenously instead. Our large online experiment of multiple, parallel, artificial, cultural markets allows us to make causal claims about the role of reward stratification. It also allows us to link those emergent properties back to the choices, heterogeneity, motivations, and behaviors of individuals that compose the whole market.

Our experimental paradigm is intentionally generic and stylized. Consequently, we believe our insights speak broadly to creative product markets where creative work is evaluated to at least some degree. These markets are characterized by subjective evaluation criteria and culturally embedded evaluators so that complex feedback loops can emerge. Examples are the markets for ideas, business innovation, cultural products (such as art and music), and science.

While our experiment is stylized, three key aspects of our paradigm make it a particularly suitable contribution to science and academic literature: peer review as an evaluation mechanism, uncertainty and subjectivity in evaluation (as opposed to objective and deterministic performance criteria), and feedback loops between producers and reviewers that allow for evolving style norms. Several recent studies on the "science of science" have provided novel insights into patterns of success (see Fortunato et al. (2018) for a review). However, they necessarily rely on observational data, which makes it difficult for these studies to make causal claims about incentive mechanisms. They are also limited to studying a single instance.

Even though our experimental task is not a direct representation of creativity in science and different sectors may have different creativity generation processes, our ability to precisely quantify novelty in multiple "parallel worlds" and to make causal claims offers important complementary insights. The study of cumulative advantage in science has have received significant attention in this journal (e.g., Bourke & Butler, 1999; Bozeman & Youtie, 2017; Braun, 1998). We contribute to this literature by providing a framework that links the emergence of distinct innovation trajectories, sorting, and market failures to the emergence of inequality.

The interrelated nature of our findings encourages a broader reflection. In 1945, Davis & Moore described stratification as a universal necessity. They argued that high-rank positions require the greatest training or talent and that talent is a scarce resource (Davis & Moore, 1945). This is correct in principle, but a natural scarcity of talent no longer seems to be a pressing issue in many modern competitive fields,

Table 2
Summary of hypotheses and findings.

	Research Question	Finding and Mechanism
Market-Level		
H1	Do distinct quality norms emerge? What does this mean for consumers? Are innovation trajectories different?	• Reward stratification leads to outlet stratification as a result of: (a) attracting many submissions, and (b) filtering (not due to quality-based targeting by producers, nor to quality standards of reviewers). • Stratified outlet quality has benefits for budget-constrained consumers, but disadvantages for consumers with larger budgets. • Higher overall level of innovation in the stratified market and distinct style norms emerge due to feedback loops (more abstract styles in stratified markets)
H2	Do levels of inequality differ?	• Stratified reward structures lead to: (a) miscoordination, (b) lower overall earnings, and (c) more type II errors. • Stratified reward structures lead to higher reward inequality as a result of market inefficiency and sorting.
Producers		
H3	Which market sees more exits?	• Stratified reward and the resulting payoff inequality lead to a higher rate of market exit.
H4	Does past success affect future success?	• Slack resources reduce inequality in the flat market, but amplify it in the stratified market.
H5	Is everyone equally successful?	• Increased competition in the stratified market amplifies individual-level differences. Some producers thrive in the competitive environment, while others fall behind.
Reviewers		
H6	Strategic reviewing	• Reviewers behave more strategically in the top-tier of the stratified market. • Producers perceive reviews in the stratified market as unfair.

including science (Larson et al., 2014). Artificial scarcity as adopted by the top journals or funding agencies is indeed effective at guaranteeing high-quality output, but it can come with two important trade-off costs.

First, stratification does not necessarily increase the overall quality of output compared to a competitive non-tiered system. Instead it acts as a sorting mechanism. While this might be an advantage for consumers who are overwhelmed by the quantity of the output, we know from empirical contest theory results that sorting is only effective when the proportion of high-ability contestants is small; otherwise, reverse sorting might take place (Dechenaux, Kovenock, & Sheremeta, 2015).

Second, stratification creates higher levels of inequality which eventually induce disadvantaged participants to exit the market. Relatively high exit rates are common in highly competitive fields. Science, for example, has been described as a "a primitive village with maximum fecundity and horrible mortality" (Desolla Price, 1965). Back in the 1960s, the scientific population grew by 7% each year, with a 17% "birth rate" and a 10% "death rate." However, now up to 84% of new PhD recipients in biology are expected to exit academia due to the lack of tenure track positions (Larson et al., 2014). Furthermore, "quitting" among those who succeed to stay in academia can come in the form of reduced willingness to compete for funding opportunities (Bol et al., 2018), fueling rich-get-richer dynamics.

Given this high exit rate, we should take opportunities to explore new mechanisms to reduce—but by no means eliminate—competitive pressure: the objective is rebalancing the current high levels of reward stratification towards a more sustainable equilibrium. Controlled experiments like the one in this paper can isolate causal mechanisms but they are not a map of the real world. Therefore, they should be complemented with real-world trials that take into account the specific organizational and institutional settings already in place in a market (Heinze et al., 2009). A number of proposals for alternative funding schemes, blind selection, and lotteries have been formulated in recent years and are waiting to be tested. In particular, partial random funding in science (Gross & Bergstrom, 2019) or lottery systems in the arts (Fraiberger et al., 2018).

Author Contributions

SB and CR designed the study and experiment, SB implemented and ran the experiment, SB performed the data analysis, SB and CR wrote the paper.

Declaration of Competing Interest

None.

Acknowledgments

This research was supported in part by a grant from the U.S. Office of Naval Research (N00014-16-1-3005). The authors are thankful to M. Seeber, B. Klein, and S. Fraiberger for insightful comments and discussion.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.respol.2021.104212.

References

Abbott, A., 2014. The System of Professions: An Essay on the Division of Expert Labor. University of Chicago Press.
Amabile, T.M., Hennessey, B.A., Grossman, B.S., 1986. Social Influences on Creativity: the Effects of Contracted-For Reward. J. Pers. Soc. Psychol. 50 (1), 14–23.
Arechar, A.A., Gächter, S., Molleman, L., 2017. Conducting Interactive Experiments Online. Experimental Economics 1–33.
Azoulay, P., Graff Zivin, J.S., Manso, G., 2011. Incentives and Creativity: Evidence from the Academic Life Sciences. Rand J. Econ. 42 (3), 527–554.
Azoulay, P., Stuart, T., Wang, Y., 2014. Matthew: Effect or fable? Management Science 60 (1), 92–109.
Balietti, S., 2017. nodeGame: Real-Time, Synchronous, Online Experiments in the Browser. Behavior Research Methods 49 (5), 1696–1715.
Balietti, S., Goldstone, R.L., Helbing, D., 2016. Peer Review and Competition in the Art Exhibition Game. Proc. Natl. Acad. Sci. 113 (30), 8414–8419.
Bol, T., de Vaan, M., van de Rijt, A., 2018. The Matthew Effect in Science Funding. Proc. Natl. Acad. Sci. 115 (19), 4887–4890.
Boudreau, K.J., Lakhani, K.R., Menietti, M., 2016. Performance Responses to Competition Across Skill Levels in Rank-Order Tournaments: Field Evidence and Implications for Tournament Design. Rand J. Econ. 47 (1), 140–165.
Bourke, P., Butler, L., 1999. The Efficacy of Different Modes of Funding Research: Perspectives from Australian Data on the Biological Sciences. Research Policy 28 (5), 489–499.
Boyd, R., Richerson, P.J., 1988. Culture and the Evolutionary Process. University of Chicago Press.
Bozeman, B., Youtie, J., 2017. Socio-economic Impacts and Public Value of Government-funded Research: Lessons from Four US National Science Foundation Initiatives. Research Policy 46 (8), 1387–1398.
Braun, D., 1998. The Role of Funding Agencies in the Cognitive Development of Science. Research Policy 27 (8), 807–821.
Bromiley, P., 1991. Testing a Causal Model of Corporate Risk Taking and Performance. Acad. Manag. J. 34 (1), 37–59.
Buser, T., Dreber, A., Mollerstrom, J., 2017. The Impact of Stress on Tournament Entry. Experimental Economics 20 (2), 506–530.
Chernoff, H., 1973. The Use of Faces to Represent Points in K-Dimensional Space Graphically. J. Am. Statist. Assoc. 68 (342), 361–368.
Choi, J.Y., Lee, J.H., Sohn, S.Y., 2009. Impact Analysis for National R&D Funding in Science and Technology Using Quantification Method II. Research Policy 38 (10), 1534–1544.
Cole, J.R., Cole, S., 1974. Social Stratification in Science. University of Chicago Press, Chicago, IL.
Cyert, R., March, J.G., 1963. A Behavioral Theory of the Firm. Englewood Cliffs, NJ.
D'Aveni, R.A., 1994. Hypercompetition: Managing the Dynamics of Strategic Maneuvering. Free Press.
Davis, K., Moore, W.E., 1945. Some Principles of Stratification. Am. Sociol. Rev. 10 (2), 242–249.
Dechenaux, E., Kovenock, D., Sheremeta, R.M., 2015. A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments. Experimental Economics 18 (4), 609–669.
Desolla Price, D., 1965. The Scientific Foundations of Science Policy. Nature 206, 233–238.
Di Lorenzo, F., Almeida, P., 2017. The Role of Relative Performance in Inter-firm Mobility of Inventors. Research Policy 46 (6), 1162–1174.
Felin, T., Foss, N.J., 2005. Strategic Organization: A Field in Search of Micro-foundations. Sage Publications London, Thousand Oaks, CA and New Delhi.
Feller, I., 1996. The determinants of research competitiveness among universities. Competitiveness in Academic Research. American Association for the Advancement of Science, Washington, pp. 35–72.
Fershtman, C., Gneezy, U., 2011. The Tradeoff Between Performance and Quitting in High Power Tournaments. Journal of the European Economic Association 9 (2), 318–336.
Finke, R.A., Ward, T.B., Smith, S.M., 1992. Creative Cognition: Theory, Research, and Applications. MIT press, Cambridge, MA.
Foley, M., Smead, R., Forber, P., Riedl, C., 2020. Avoiding the Bullies: Resilience of Cooperation Among Unequals.
Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., … Barabási, A.L., 2018. Science of Science. Science 359 (6379).
Fraiberger, S.P., Sinatra, R., Resch, M., Riedl, C., Barabási, A.-L., 2018. Quantifying Reputation and Success in Art. Science 362 (6416), 825–829.
Galinsky, A., Schweitzer, M., 2015. Friend & Foe: When to Cooperate, When to Compete, and How to Succeed at Both. Crown Business.
Gross, K., Bergstrom, C.T., 2019. Contest Models Highlight Inefficiencies of Scientific Funding. PLoS Biol. 17 (1), e3000065.
Grubb, M.A., Tymula, A., Gilaie-Dotan, S., Glimcher, P.W., Levy, I., 2016. Neuroanatomy Accounts for Age-related Changes in Risk Preferences. Nat. Commun. 7, 13822.
Harbring, C., Irlenbusch, B., 2008. How Many Winners Are Good to Have?: On Tournaments With Sabotage. Journal of Economic Behavior & Organization 65 (3), 682–702.
Harbring, C., Irlenbusch, B., 2011. Sabotage in Tournaments: Evidence from a Laboratory Experiment. Management Science 57 (4), 611–627.
Heckman, J.J., Moktan, S., 2020. Publishing and Promotion in Economics: The Tyranny of the Top Five. J. Econ. Lit. 58 (2), 419–470.
Heinze, T., Shapira, P., Rogers, J.D., Senker, J.M., 2009. Organizational and Institutional Influences on Creativity in Scientific Research. Research Policy 38 (4), 610–623.
Horton, J.J., Rand, D.G., Zeckhauser, R.J., 2011. The Online Laboratory: Conducting Experiments in a Real Labor Market. Experimental Economics 14 (3), 399–425.
Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. Q. J. Econ. 108 (3), 577–598.
Jones, B.F., Wuchty, S., Uzzi, B., 2008. Multi-university Research Teams: Shifting Impact, Geography, and Stratification in Science. Science 322 (5905), 1259–1262.
Kahneman, D., Riis, J., 2005. Living and Thinking About It: Two Perspectives on Life. The Science of Wellbeing: Integrating Neurobiology, Psychology, and Social Science. Oxford University Press.

Larson, R.C., Ghaffarzadegan, N., Xue, Y., 2014. Too Many PhD Graduates or Too Few Academic Job Openings: The Basic Reproductive Number R0 in Academia. Systems Research and Behavioral Science 31 (6), 745–750.
Lavie, D., Stettner, U., Tushman, M.L., 2010. Exploration and Exploitation Within and Across Organizations. Academy of Management Annals 4 (1), 109–155.
Lazear, E.P., Oyer, P., 2012. Personnel Economics. The Handbook of Organizational Economics. Princeton University Press, pp. 479–519.
Levinthal, D., March, J.G., 1981. A Model of Adaptive Organizational Search. Journal of Economic Behavior & Organization 2 (4), 307–333.
Markovits, D., 2019. The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite. Penguin Press.
Mason, W., Suri, S., 2012. Conducting Behavioral Research on Amazon's Mechanical Turk. Behavior Research Methods 44 (1), 1–23.
Merton, R.K., 1968. The Matthew Effect in Science. The Reward and Communication Systems of Science Are Considered. Science, 159 (810), 56–63.
Nohria, N., Gulati, R., 1996. Is Slack Good or Bad for Innovation? Acad. Manag. J. 39 (5), 1245–1264.
Perc, M., 2014. The Matthew Effect in Empirical Data. J. R. Soc., Interface 11 (98), 20140378.
Piketty, T., Saez, E., 2014. Inequality in the Long Run. Science 344 (6186), 838–843.
Pluchino, A., Biondo, A.P., Rapisarda, A., 2018. Talent vs Luck: The Role of Randomness in Success and Failure. Adv. Complex Syst. 21 (03n04), 1850014.
Riedl, C., Grad, T., Lettl, C., 2020. Strategic Behavior in Contests with Ability Heterogeneous Contestants: Evidence from Field Data. Available at SSRN, 3387056.
Rivera, L.A., 2011. Ivies, Extracurriculars, and Exclusion: Elite Employers' Use of Educational Credentials. Research in Social Stratification and Mobility 29 (1), 71–90.
Rzhetsky, A., Foster, J.G., Foster, I.T., Evans, J.A., 2015. Choosing Experiments to Accelerate Collective Discovery. Proc. Natl. Acad. Sci. 112 (47), 14569–14574.
Seeber, M., Cattaneo, M., Meoli, M., Malighetti, P., 2019. Self-citations as strategic response to the use of metrics for career decisions. Research Policy 48 (2), 478–491.
Shalley, C.E., Zhou, J., Oldham, G.R., 2004. The Effects of Personal and Contextual Characteristics on Creativity: Where Should We Go from Here? J. Manag. 30 (6), 933–958.
Siler, K., Lee, K., Bero, L., 2015. Measuring the Effectiveness of Scientific Gatekeeping. Proceedings of the National Academy of Sciences (PNAS) 112 (2), 360–365.
Teplitskiy, M., Acuna, D., Elamrani-Raoult, A., Körding, K., Evans, J., 2018. The Sociology of Scientific Validity: How Professional Networks Shape Judgement in Peer Review. Research Policy 47 (9), 1825–1841.
Wang, J., Lee, Y.-N., Walsh, J.P., 2018. Funding Model and Creativity in Science: Competitive Versus Block Funding and Status Contingency Effects. Research Policy 47 (6), 1070–1083.
Wiseman, R.M., Bromiley, P., 1996. Toward a Model of Risk in Declining Organizations: An Empirical Examination of Risk, Performance and Decline. Organization Science 7 (5), 524–543.