Balancing on the Creative High-Wire:
Forecasting the Success of Novel Ideas in Organizations

Justin M. Berg
Stanford Graduate School of Business
Knight Management Center - 655 Knight Way
Stanford, CA 94305-7298
jmberg@stanford.edu

Keywords: Creativity, Innovation, Role Design, Idea Evaluation

Balancing on the Creative High-Wire:
Forecasting the Success of Novel Ideas in Organizations

Abstract

Betting on the most promising new ideas is key to creativity and innovation in organizations, but predicting the success of novel ideas can be difficult. To select the best ideas, creators and managers must excel at "creative forecasting"—i.e., predicting the outcomes of new ideas. In this paper, I build hypotheses on the conditions for accurate creative forecasting, focusing on the effects of creator and manager roles. To test these hypotheses, I use a field study of 339 professionals in the circus arts industry, and a lab experiment. In the field study, creators and managers forecasted the success of new circus acts with audiences, and the accuracy of these forecasts was assessed using data from 13,248 audience members. Results suggest that creators were more accurate than managers when forecasting about others' novel ideas (but not their own ideas). This advantage over managers was undermined when creators had poor ideas that were successful in the marketplace anyway. Results from the lab experiment provide causal evidence for the proposed mechanism driving creators' advantage over managers: the emphasis placed on divergent thinking in the creator role. I discuss the implications of these results for theory and research on creativity and innovation in organizations.

As the world becomes more interconnected and competitive, creativity and innovation are increasingly important. Within organizations, scholars have defined creativity as the production of novel and useful ideas, and innovation as the successful implementation of creative ideas (Amabile, 1988). In this way, novel ideas are the lifeblood of successful innovations. However, even the most capable organizations cannot invest in every novel idea they generate. Instead, managers must select which novel ideas should be implemented—and which should be rejected—based on their predictions of how successful they expect the ideas to be (Sternberg & Lubart, 1991). This can be difficult because novel ideas are by definition unique from existing ideas (Csikszentmihalyi, 1999), and thus their potential cannot be known for certain until they are fully implemented and tested on the market with their intended audience (Fleming, 2001).

Indeed, history is full of cases in which managers unwisely rejected novel ideas that proved to be highly valuable. From Pontiac's popular car the Fiero (Pinchot, 1985), to Hewlett-Packard's successful electronic display (Nemeth, 1997), to hit movies and television shows such as Star Wars, The Truman Show, and Seinfeld (Elsbach & Kramer, 2003), managers have rejected novel ideas that others subsequently developed into highly successful innovations (Mainemelis, 2010). To avoid these costly mistakes and select the best ideas, managers must excel at "creative forecasting"—i.e., predicting the outcomes of new ideas (Byrne, Shipman, & Mumford, 2010; Dailey & Mumford, 2006; Moreau, Lehmann, & Markman, 2001; Mumford, Devin, & Ginamarie, 2002; Wilson et al., 1954). While creative forecasting may focus on a variety of outcomes, the ultimate success of new products and services typically depends on how they are received by potential customers in the marketplace (Kornish & Ulrich, 2014). Thus, in this paper, I focus on creative forecasting in which the goal is predicting the success of new ideas with their intended audience outside the organization (Glăveanu, 2013).

Although creativity and innovation are intertwined processes in organizations, scholars have typically studied them separately (Anderson, Potočnik, & Zhou, 2014). As a result, creative forecasting has been rarely studied, as it lies in the relatively unexplored space in between generating creative ideas and implementing them. Much of the prior research related to creative forecasting has focused on individuals' accuracy in assessing the novelty—but not the ultimate market success—of new ideas (e.g., Basadur, Runco, & Vega, 2000; Licuanan, Dailey, & Mumford, 2007; Runco & Smith, 1992; Silvia, 2008). Some research has addressed individuals' accuracy in forecasting the market success of new ideas, focusing on the influence of domain expertise. This work has revealed the important insight that expertise may be of no value in creative forecasting (Kornish & Ulrich, 2014), and can even be a liability, as domain knowledge may lead individuals to undervalue novel ideas (Moreau, Lehmann, & Markman, 2001) and overvalue more familiar ideas (Dailey & Mumford, 2006). This implies that as individuals gain knowledge in their domains, their ability to accurately forecast the success of new ideas is likely to decrease, or at best remain the same. However, little theory and research has addressed how this trap may be avoided, leaving unanswered questions about what may help individuals improve—or at least maintain—their accuracy in creative forecasting as they gain knowledge and experience in their domains.

Drawing on theories of the creative process (e.g., Amabile, 1996; Basadur, Runco, & Vega, 2000; Guilford, 1967), I propose that the roles in which individuals are embedded in organizations may be key in shaping whether domain knowledge is a blessing or a curse in creative forecasting. In particular, I focus on the two roles that are typically involved with innovation in organizations: *creators*, who are expected to generate novel and useful ideas, and *managers*, who are in charge of selecting which ideas to implement (Drazin, Glynn, &

Kazanjian, 1999; Elsbach & Kramer, 2003; Mintzberg, 1971; Mollick, 2012). Both creators and managers are likely to be experts in their domains, but we know little about how their respective roles may shape the way they use their domain knowledge in creative forecasting.

In this paper, I propose hypotheses on drivers of (in)accuracy in creative forecasting, focusing on the effects of creator and manager roles. To test these hypotheses, I conducted two studies. Study 1 was a field study in the circus arts industry, in which creators and managers were asked to forecast the success of new circus acts with audiences, and the accuracy of these forecasts was assessed using a sample of audience members. Study 2 was a lab experiment in which participants were randomly assigned to creator and/or manager roles before engaging in a creative forecasting task.

These studies offer important contributions to theory and research on creativity and innovation in organizations. First, the studies highlight how creative forecasting is a critical bridge linking creativity and innovation, contributing to the small but growing literature on key processes that happen in between generating creative ideas and implementing them (e.g., Baer, 2012; Lingo & O'Mahony, 2010; Perry-Smith & Mannucci, 2015). Second, although aspects of the innovation process are often divided into separate roles in organizations, scholars have paid relatively little attention to the effects of role designs in creativity and innovation (cf. Elsbach & Kramer, 2003; Mollick, 2012). This research sheds light on the importance of roles in creative forecasting, drawing attention to the effects of roles in creativity and innovation more broadly. Third, this research advances theory on why creative success is difficult to sustain over time: while past theories have focused on declines in the generation of new ideas following success (Audia & Goncalo, 2007; Bayus, 2013; Simonton, 1997, 1999), this research explains why success may also undermine the evaluation of new ideas over time.

ROLES AND CREATIVE FORECASTING: CREATORS VS. MANAGERS

To explain how creative ideas become innovations, scholars have construed innovation as an evolutionary process consisting of three main stages: variation, selection, and retention (e.g., Aldrich, 1999; Campbell, 1960; Ford, 1996; Miner, 1994; Nelson & Winter, 1982; Simonton, 1999; Staw, 1990). In many organizations and industries, these three stages are divided into the following prototypical roles: creators are expected to generate novel and useful ideas (variation), managers decide which of these ideas to implement (selection), and the external audience determines the ultimate success of any implemented ideas (retention) (Burgelman, 1991). The audience's response to implemented ideas is then fed back to creators and managers, and the relatively salient ideas become common knowledge in the domain (Csikszentmihalyi, 1999).

In this prototypical structure, managers act as gatekeepers in between creators and the audience for which their ideas are intended (Mainemelis, 2010). For example, researchers have studied Hollywood executives evaluating whether screenplays should be selected for production based on pitches from screenwriters (Elsbach & Kramer, 2003). Mollick (2012) highlighted the distinction between managers and creators (or "suits and innovators" in his language) in the video game industry, where managers evaluate and select the ideas generated by game creators. Many organizations purposefully construct separate R&D units with just creators so they can specialize in generating novel ideas, but then managers ultimately decide whether these ideas are released to the audience or not (Benner & Tushman, 2003).

As Mintzberg (1971) described, one of the ten roles managers are expected to play is the role of entrepreneur, which involves evaluating and selecting which new ideas the organization should pursue. Managers delegate the work of generating new ideas to creators, but maintain the authority to select which ideas are retained or discarded. The result is that managers frequently

evaluate ideas generated by creators, and must decide whether to keep or drop each idea from the set of ideas they oversee. As Mintzberg (1971, p. B-105) explains, "…the manager may be likened to a juggler. At any one point, he maintains a number of balls in the air. Periodically, one comes down, receives a short burst of energy, and goes up again. Meanwhile, an inventory of new balls waits on the sidelines and, at random intervals, old balls are discarded and new ones added." Thus, managers spend much of their time evaluating creators' ideas—both individually and compared to one another—as they decide which ideas to keep versus drop.

In selecting between ideas, managers are susceptible to two types of errors: false negatives (rejecting relatively good ideas based on an underestimation of their potential success) and false positives (selecting relatively bad ideas based on an overestimation of their potential success). Both errors can be costly—false negatives neglect opportunities to utilize relatively good ideas, and false positives waste resources on relatively bad ideas. To effectively manage innovation, managers must avoid both types of errors simultaneously, as avoiding just one type may squander precious opportunities or resources (Csaszar, 2012; Tetlock & Mellers, 2011).

In the sections that follow, I propose five hypotheses on how creator and manager roles may impact creative forecasting. Hypotheses 1 and 2 comprise the main argument of the paper— i.e., creators have an advantage over managers in forecasting about others' novel ideas. Then, Hypotheses 3-5 focus on the boundary conditions of creators' advantage over managers. See Figure 1 for a visual summary of the hypotheses.

[Insert Figure 1 about here]

Creators' Advantage Over Managers in Forecasting About Others' Ideas

In organizations, the goal of dividing creator and manager roles is often to ensure that creators have freedom to generate variation, and the hope is that managers will then select the

best ideas from this variation (Aldrich, 1999; Miner, 1994). However, due to the nature of their roles, creators may have an advantage over managers in forecasting the success of others' ideas. Although many managers get promoted into management positions based on their success as a creator, once they become managers, they often spend more time evaluating others' ideas than generating their own novel ideas. By specializing in evaluation and selection, managers are typically cut off from the process of generating novel ideas that creators frequently experience. This may be problematic, as theories of the creative process suggest that idea generation and evaluation tend to be intimately intertwined (Amabile, 1996; Basadur, Runco, & Vega, 2000; Lubart, 2001, Silvia, 2008; Ward, Smith, & Finke, 1999).

When individuals try to generate novel ideas, they engage in divergent thinking, which involves searching for novel associations, combinations, or perspectives that may be useful (Guilford, 1967). When individuals evaluate ideas, they engage in convergent thinking, which involves applying criteria, standards, and logics based on their prior knowledge and experience (Cropley, 2006). Theories of the creative process suggest that to develop creative ideas, creators iterate between divergent and convergent thinking, as they generate possible ideas with divergent thinking and evaluate them with convergent thinking (Amabile, 1996; Guilford, 1967; Lubart, 2001). In contrast, because managers typically evaluate ideas after creators have already generated them, managers are likely to do more convergent thinking than divergent thinking (Cropley, 2006; Guilford, 1967). Thus, while both creators and managers may spend plenty of time engaging in convergent thinking, creators often spend more time engaging in divergent thinking than managers.

The notion that creators typically do more divergent thinking than managers may be important because past research has found a positive relationship between divergent and

convergent thinking—i.e., individuals who are good at divergent thinking (idea generation) tend to be good at convergent thinking (idea evaluation) and vice-versa (Basadur, Runco, & Vega, 2000; Runco, 1991; Runco & Smith, 1992; Silvia, 2008). To explain this relationship, Runco (1991) suggests that divergent thinkers create more opportunities to practice convergent thinking, and thus become better convergent thinkers as a result. Building on this logic, the creator role may be better designed for creative forecasting than the manager role, as the creator role may promote a more optimal balance of divergent and convergent thinking. Creators spend much of their time focused on diverging from conventional ideas in the domain to discover novel ways in which ideas may be successful. Indeed, past research suggests that creators produce final ideas that are more creative when they withhold judgment of their initial ideas until they are fairly developed (Finke, 1996; Ward, Smith, & Finke, 1999). This divergent thinking may help creators stay open minded about how novel ideas may be valuable as they then engage in convergent thinking regarding their own or others' ideas. Thus, the very act of divergent thinking may improve creators' convergent thinking by making it more open and flexible.

In contrast, because managers typically enter the creative process at the evaluation stage, they are likely to spend much of their time engaging in convergent thinking without first engaging in divergent thinking. Without divergent thinking paired with it, managers may be more anchored by their domain knowledge as they engage in convergent thinking (Basadur, Runco, & Vega, 2000; Dane, 2010). Even if managers were previously creators, the benefit of the divergent thinking they did as a creator may disappear as they focus on convergent thinking alone, as this involves applying their past knowledge and experience rather than diverging from it. Thus, managers may rely too heavily on conventional models of success in forecasting the success of novel ideas (Dailey & Mumford, 2006; Moreau, Lehmann, & Markman, 2001; Ward,

1994). In turn, managers may undervalue novel ideas and overvalue conventional ideas that better match the ideas that have been successful with the audience in the past (Licuanan, Dailey, & Mumford, 2007; Mueller, Melwani, & Goncalo, 2012; Rietzschel, Nijstad, & Stroebe, 2010).

In sum, while divergent thinking may help creators identify signals of value in novel ideas, convergent thinking (without divergent thinking paired with it) may undermine managers' ability to recognize value in novel ideas. Thus, a key distinction between the two roles is that creators often spend more time engaging in divergent thinking than managers, which may help creators be more open and flexible when they engage in convergent thinking. In turn, when forecasting about others' novel ideas, creators may be less likely than managers to commit false negatives—i.e., undervalue good ideas. In addition to being more open to novel ways in which ideas may be successful, creators should still be able to leverage their domain knowledge on what has succeeded and failed in the past when forecasting about more conventional ideas. Therefore, creators may be more accurate overall than managers in forecasting the success of others' ideas. Putting these arguments together, I propose:

> **Hypothesis 1:** Creators are more accurate than managers at forecasting the success of others' ideas.

> **Hypothesis 2:** Novelty moderates creators' advantage over managers in forecasting accuracy, such that creators are less likely than managers to commit false negatives regarding novel ideas.

Boundary Conditions: The Impact of Creators' Own Ideas

On average, creators may have an advantage over managers in forecasting about others' ideas, but this advantage may have important boundary conditions. The remaining three hypotheses (H3-H5) focus on how creators' own ideas may produce conditions that undermine their advantage over managers. First, H3 proposes that creators are likely to overvalue their own ideas, giving managers the advantage in forecasting about creators' own ideas. Since H3 is in

line with a number of existing theories, it is intended to highlight and empirically test an important boundary condition according to the literature, but not advance theory. However, H4 and H5 are intended to advance theory on creative forecasting by shedding light on how the success and quality of creators' own ideas may influence their accuracy in forecasting about others' ideas.

Forecasting about creators' own ideas. Past research suggests that individuals are better at evaluating the novelty of others' ideas than their own ideas (Runco & Smith, 1992), hinting that creators' advantage over managers may not extend to forecasting about their own ideas. Moreover, many existing theories suggest that creators are likely to overestimate the success of their own ideas. From a cognitive standpoint, creators may become entrenched in the content of their own ideas, making it difficult for them to accurately predict how audiences will experience their ideas (Dane, 2010). New ideas often come from recombining knowledge from multiple domains (Burt, 2004; Fleming, Mingo, & Chen, 2007; Weick, 1979). When creators construct ideas using content from multiple domains with which they are familiar, the audience in the focal domain may not be familiar with the content from other domains. Creators may fall for false consensus biases and overestimate the extent to which the focal audience thinks and feels the same as them about their ideas (Ross, Greene, & House, 1977), driving them to commit false positives regarding their own ideas.

Motivational factors may also drive creators to overvalue their own ideas. Theories of self-enhancement and motivated reasoning suggest that creators are likely to focus on self-serving justifications for why their ideas are likely to succeed and ignore reasoning to the contrary (e.g., Audia & Brion, 2007; KC, Staats, & Gino, 2013; Kunda, 1990; Pfeffer & Fong, 2005). Also, creators may come to identify strongly with their ideas (Beggan, 1992; Pierce,

Kostova, & Dirks, 2001). This sense of ownership may drive them to dismiss negative feedback (Baer & Brown, 2012) and escalate their commitment (Staw & Ross, 1989) to their ideas. In sum, creators are subject to a number of cognitive and motivational biases that may lead them to commit false positives regarding their own ideas.

In contrast, when managers forecast about creators' ideas, they may not be as susceptible to the cognitive and motivational biases that drive creators to overvalue their own ideas. Managers' relative distance from the idea may help them be more accurate in taking the audience's perspective (Dailey & Mumford, 2006), and more likely to consider reasons why the idea may fail, rather than focusing only on why it will succeed (Baer & Brown, 2012; Kunda, 1990; Staw & Ross, 1989). Thus, compared to the creators who generated the ideas in question, managers may be more accurate in creative forecasting.

Hypothesis 3: Creators tend to commit false positives regarding their own ideas (H3a), giving managers the advantage in forecasting the success of creators' own ideas (H3b).

Past success and idea quality. Along with a tendency to overvalue their own ideas, creators' experiences with their own ideas are likely to influence their accuracy in forecasting about others' ideas. In particular, the success of creators' ideas may present a liability for them over time (Audia & Goncalo, 2007; Bayus, 2013). When creators generate ideas that achieve success in the marketplace, they are likely to make self-serving attributions, ascribing their success to internal factors such as skill and talent, rather than external factors like luck and chance (Greenwald, 1980; Miller & Ross, 1975). As a result, they are likely to value the aspects of their ideas that they believe made them successful (Audia, Locke, & Smith, 2000). However, it may be difficult for creators to learn from their success (Madsen & Desai, 2010), as it may not be clear which features of their ideas made them successful (Sitkin, 1992), or which features will predict market success in the future (Levinthal & March, 1993). Thus, when creators have

successful ideas, much of what they subsequently value in ideas may serve as noise—and less may serve as signal—in forecasting the success of other ideas going forward.

When is past success better or worse for creators' subsequent forecasting accuracy? The quality of their ideas in the first place may be key. I follow other scholars who define quality as the audience appeal that is intrinsic to ideas themselves, rather than driven by external factors such as social influence, marketing, or opportunity structures (Kornish & Ulrich, 2014; Reeves & Bednar, 1994; Salganik, Dodds, & Watts, 2006). Although the quality of ideas may be one of the most salient and useful predictors of success in the marketplace, many unpredictable factors may also contribute to the market success of ideas (Fleming, 2001; Merton, 1968). Because quality and market success are often not highly correlated, some low-quality ideas may succeed and some high-quality ideas may fail—even though higher quality ideas are more likely to succeed on average (Salganik, Dodds, & Watts, 2006).

Thus, if creators' ideas are low quality but succeed in the marketplace anyway, they may misattribute their success to the quality of their ideas, rather than luck (Miller & Ross, 1975; Ross & Sicoly, 1979). This misattribution may render creators "unskilled and unaware of it" (Kruger & Dunning, 1999), as the misleading feedback from the marketplace may prevent them from learning that their idea was actually low in quality. In turn, creators may overvalue features of their low-quality idea going forward. Because these features are likely not indicative of quality according to the audience (and may actually be indicative of low quality), creators' ability to discern the quality of subsequent ideas may be reduced. This may undermine their forecasting accuracy, as quality is often a useful (albeit imperfect) predictor of market success (Kornish & Ulrich, 2014; Salganik, Dodds, & Watts, 2006). However, when creators have low-quality ideas that fail in the marketplace, they may be more aware of their lack of skill (Kruger &

Dunning, 1999). In response to this negative feedback about their ideas, they are unlikely to overvalue features of their low-quality ideas, helping them maintain their advantage over managers in forecasting accuracy. In sum, I hypothesize the following:

> **Hypothesis 4:** When creators have experienced past success with their own ideas, their accuracy in forecasting the success of others' ideas is reduced.

> **Hypothesis 5:** The quality of creators' ideas moderates the negative relationship between past success and forecasting accuracy. Specifically, low-quality ideas that succeed reduce creators' accuracy more than low-quality ideas that fail.

STUDY 1

Context: The Case of the Circus Arts

I tested the hypotheses (H1-H5) using a field study in the performing arts domain—the circus arts industry in particular. The circus industry includes hundreds of circus companies that employ thousands of individuals worldwide, including circus artists, managers, and supporting staff. The largest circus company is Cirque du Soleil, which has been credited with reinventing and reviving the circus industry (Kim & Mauborgne, 2004). Cirque du Soleil was founded in 1984, and has grown from a small group of street performers to a global corporation with about 4,000 employees, including 1,300 circus artists (Cirque du Soleil, 2015), and estimated annual revenue of $910 million (Forbes, 2014). The rise of Cirque du Soleil has brought about a new era of circus arts, focusing on sophisticated storytelling and the artistry of human performers, rather than the animals and antics of traditional circuses (Leslie & Rantisi, 2011).

I selected the circus industry as an ideal context for studying creative forecasting for three main reasons. First, creativity and innovation are the lifeblood of the circus industry. The principal focus in the industry is putting acts on stage that are highly novel to audiences. At the same time, audience members are likely to have some familiarity with circus arts shows. For instance, more than 150 million people have seen a Cirque du Soleil show since 1984, and

approximately 15 million in 2014 alone (Cirque du Soleil, 2015). Thus, although audiences expect to see novelty in circus shows, they may not see all circus acts as novel, making creative forecasting both important and difficult in this industry.

Second, like many industries, creators (i.e., circus artists) and managers (i.e., circus directors, producers, and agents) are typically divided into separate roles within the circus industry, enabling comparisons of creator and manager roles. Although the industry term is circus "artist," to capture the broader role of idea generator, I use "creator" hereafter. In addition to performing in existing shows, circus creators are typically expected to generate novel acts to showcase their talents, and they present their creations to managers through formal and informal auditions. Managers then decide whether to include creators' acts in future shows. Mintzberg's (1971) aforementioned juggling analogy is not only relevant to the circus industry in a literal sense, but also figuratively, as circus managers often juggle a number of possible acts as they decide which acts to put in future shows. Thus, creators focus on generating their own novel ideas, while managers focus on evaluating creators' ideas, providing the key theoretical distinction between creator and manager roles underlying H1 and H2 (i.e., creators spend more time engaging in divergent thinking than managers). Third, circus creators are trained in circus schools with well-established, legitimized standards for what constitutes appropriate technique and expertise for circus creators (National Circus School, 2014). Because most managers start as creators and thus attend circus school, creators and managers are likely to share much of the same domain knowledge, which may help isolate the effects of their roles in creative forecasting.

Participants

The participants included a total of 489 individuals: 339 circus professionals, and 150 laypeople. The 339 circus professionals spanned 43 different countries and included 177

creators, 120 managers, and 42 hybrids, who occupied roles with both creator and manager duties. To qualify as a creator, manager, or hybrid, participants had to have at least one year in their current role as a creator, manager, or hybrid (see Table 1 for participants' average years of circus experience by role). Although hybrids were not directly part of the hypotheses, they were included in the analyses as a supplementary comparison group because they could help shed light on the key theoretical distinction between creator and manager roles—i.e., the relative focus on idea generation (divergent thinking) versus idea evaluation (convergent thinking). Because hybrids are selected into manager roles but stay involved in idea generation as well, they presumably engage in more divergent thinking than managers, but should have virtually the same power, status, and accountability pressures as managers. Thus, comparing hybrids to pure managers and creators helped rule out selection effects and other potential confounds, helping to isolate the effect of divergent versus convergent thinking in the role.

In addition to these 339 circus professionals, a sample of 150 laypeople was also recruited to provide a comparison group that did not have any professional circus experience (50.0 percent female, M_{age} = 33.86, SD_{age} = 10.95). By providing a baseline level of forecasting accuracy, the laypeople helped shed light on whether the creator role gives creators an advantage and/or the manager role puts managers at a disadvantage. These laypeople were recruited and paid $8.00 for completing the survey via Amazon MTurk.

Data Collection Procedures

Participants were asked to watch online videos of circus acts and forecast how successful the videos would be with the audience. For help with recruiting participants, I partnered with two insiders to the circus arts industry: one was a veteran creator, and the other was a former Cirque du Soleil international creative director and experienced creator. We collected data in three main

stages. First, to collect the videos of circus acts, we hosted a circus video contest, inviting circus creators to submit online videos of their circus acts. To ensure the number of forecasts would not be spread too thin between videos, we conducted a pretest to select a subset of 100 videos to be the subject of the forecasting survey (see Appendix A for a description of the pretest). Second, we invited the creators who submitted videos—as well as an additional sample of circus professionals and laypeople—to complete the forecasting survey, in which they were asked to forecast the audience success of ten videos randomly selected from the pool of 100 videos.[1] We recruited circus professionals to participate through advertising on social media and partnerships with circus companies. As aforementioned, a total of 489 participants completed the forecasting survey (339 circus professionals, and 150 laypeople). Third, to test the accuracy of participants' forecasts, I collected data on the videos from a large sample of audience members. The forecasting and audience surveys are each described in more detail below.

[Insert Tables 1 and 2 about here]

Measures

      Forecasts. The forecasting survey asked participants to forecast the audience reaction to ten videos that were randomly selected from the pool of 100 videos chosen in the pretest. By having participants make forecasts about ten different videos, each participant ended up rating a mix of relatively novel and relatively conventional videos. This enabled scoring of how accurate participants were regarding relatively novel ideas (to test H2). Creators and hybrids who submitted videos also forecasted about their own videos at the end of the pretest survey. Given

---

[1] For the participants who submitted videos to the contest, the forecasting survey was framed as the second round of the contest. The pretest survey was presented as the first round. We designed the contest prizes to encourage participation without biasing the results by offering counteracting rewards for the top videos and judges—contestants were informed that the top video and the top judge would each receive $1,000, and that $500 would be given to each of the next four videos and next four judges. Contestants were not told whether their video qualified for the second round (i.e., forecasting survey), nor how videos or judges would be scored.

that the typical goal of these online videos is to spread appreciation for creators' work, the extent to which the audience likes, shares, and seeks out the performer(s) in a show are central and face-valid dimensions of success. Thus, after participants watched each video, they made forecasts using the following three items (on a 0-100 percent scale): "What percentage of the general public would like this video?"; "What percentage of the general public would want to share this video with other people?"; and "What percentage of the general public would go see the performer(s) from this video in a show?" Following past research on creative success, these three items were intended to capture success in a way that accounted for context (Glăveanu, 2012), but also in a way that could be measured and empirically tested (e.g., Fleming, Mingo, & Chen, 2007).

Audience success. To assess the accuracy of participants' forecasts, I administered a survey to measure the success of the videos with the audience. A total of 13,248 audience members from the general public completed the audience survey. The survey asked each audience member to watch and rate one video that was randomly selected from the pool of 161 videos included in the study (the 100 videos included in the forecasting survey, plus 61 videos that were not selected for the forecasting survey in the pretest but were rated by the creators who submitted them). Each video was watched and rated by an average of 82.29 audience members ($SD = 1.13$).[2] These audience members were recruited and compensated through Amazon MTurk, which has been shown to be a useful context for collecting survey data from large and fairly representative samples of the general public (Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014). Audience members were paid $1.00 for completing the survey.

---

[2] Data collection was concluded once the relative rank stabilized among the 161 videos in terms of audience success. Starting with the first 8,000 respondents (approximately 50 per video), I calculated the video rankings after each video was rated approximately 10 additional times. From 60 to 70 views per video, no video moved more than three ranks. And from 70 to 80 views, no video moved more than one rank, and thus data collection was concluded.

They were 47.83 percent female, averaged 31.81 years of age ($SD = 10.64$), and were located in 117 different countries (81.14 percent from the U.S.). Audience members had attended an average of 1.96 circus arts shows ($SD = 2.94$), and 69.93 percent had attended at least one circus arts show.

The audience survey was designed to capture three dimensions of success, which map onto the three forecasting items: liking, sharing, and financial. First, to capture liking, audience members were asked to rate the video they watched on three items adapted from Oliver and Bartsch's (2010) measure of audience appreciation for entertainment experiences: "I [liked, appreciated, enjoyed] this video" ($\alpha = .98$). These items were rated using a Likert-type scale (1 = "Strongly Disagree" to 7 = "Strongly Agree").

Second, to capture sharing, audience members were given the opportunity to actually share the video they watched via Facebook, Twitter, Google+, or email. On average, each video was shared by 49.82 percent of the audience members who viewed it ($SD = 6.85$ percent), and sharing frequency per video ranged from 29.27 to 67.47 percent.

Third, to capture financial success, audience members were told that they earned a ten-cent bonus and could give as much of this bonus as they would like to support the work of the performer(s) in the video they watched. This was designed to provide a behaviorally and psychologically real proxy for audience members' desire to go see the performer(s) in a show, which was the focus of the third forecasting item.[3] Audience members gave an average of 4.75 cents to the performer(s) in each video ($SD = 0.48$). An approximately equal percentage of audience members either kept all ten cents (34.06 percent) or gave all ten cents away (33.44 percent), and 18.48 percent split it 5-5 with the performer(s). The remaining 14.02 percent were

---

[3] The financial measure was highly correlated with the item "I would go see the performer(s) from this video in a show" in the audience sample, $r = .50$, $p < .001$. As a robustness check, I ran all the reported analyses with this item in place of the financial measure, and the pattern of results remained the same.

spread fairly evenly across the other possible divisions. Table 2 contains means and correlations for all the measures in the audience survey.

Novelty. To assess the novelty of the videos, I used the consensual assessment technique (Amabile, 1996). Three audience members who did not participate in the audience survey independently rated the 161 videos included in the study. Using a Likert-type scale (1 = "Strongly Disagree" to 7 = "Strongly Agree"), they rated each video on the item "This video was novel", with novelty defined as the degree to which the content of the video was unique from existing ideas, particularly in the circus arts domain but also in general (Csikszentmihalyi, 1999; Ford, 1996). Following suggested practice (Amabile, 1996), the order of the videos was randomized for each rater, and the raters were asked to watch 15 videos before they started rating to establish a means of comparison. The ratings met standard cutoffs for both interrater reliability (ICC2 = .78) and agreement (average deviation = 0.64) (LeBreton & Senter, 2008).

Past success. To measure the past success of creators' videos, I recorded the number of views each video had prior to the study (157 of the 161 videos had publicly available view counts), which ranged from one to 397,877 views ($M = 12,313$, $SD = 54,619$). I also recorded the number of days each video had been online prior to the study, to serve as a control in the analyses for H4 and H5 ($M = 325.51$ days, $SD = 472.63$).

Quality. To measure the underlying quality of creators' videos, I followed other scholars who have defined quality as audience appeal that is intrinsic to the ideas themselves (Kornish & Ulrich, 2014; Reeves & Bednar, 1994; Salganik, Dodds, & Watts, 2006). In particular, I drew on Salganik, Dodds, and Watts (2006), who measured the quality of online content (in their case songs) as the frequency with which audience members downloaded the songs when no other information or social encouragement was provided. The logic of this approach was that if

audience members downloaded the songs absent of any external influences, their decision to download was based on the quality of the content itself. Thus, by giving each song an equal opportunity to be downloaded, they isolated the audience appeal that was intrinsic to the content.

Adapting this approach in a way that paralleled the measurement of past success as prior views, I measured quality as the percentage of respondents who shared each video in the audience sample. Since the videos were embedded in the survey, audience members did not know the number of times the video they watched had been viewed or shared by others (the names of the videos were excluded to minimize the risk of audience members looking them up). Thus, each video was given an approximately equal opportunity to be shared absent of external influences, isolating the intrinsic quality of the videos themselves in driving the percentage of audience members who shared each video.

Accuracy Scoring

Drawing on Moore and Healy (2008), the forecasting and audience surveys were designed to enable scoring of two types of forecasting accuracy: estimation (accuracy in predicting the absolute success of any one idea) and placement (accuracy in ranking how successful a set of ideas will be compared to one another). Although the hypotheses make the same predictions regarding both estimation and placement, both were included in the analyses because each type is important in a different way. As managers evaluate the set of available ideas and decide which to implement, their estimates of the absolute success of each idea may influence the number of ideas they are willing to implement. However, regardless of how many ideas managers decide to implement, placement is important because managers need to rank ideas correctly in order to select the best of the available ideas.

Estimation accuracy. The most salient measure of absolute success for online videos is the number of times they have been viewed, and sharing is the main driver of how videos gain views. Thus, the sharing dimension was designed to enable scoring of estimation accuracy, as this provided a concrete, behavioral measure of absolute success. In the forecasting survey, participants were asked to predict the percentage of the audience who would share each video, and then the actual percentage of audience members who shared each video was captured in the audience survey. I scored estimation accuracy by subtracting the actual percentage from participants' predicted percentage for each video they rated.

Placement accuracy. Following the theoretical focus on false positives (ranking an idea too high relative to other ideas) and false negatives (ranking an idea too low relative to other ideas), I scored participants' placement accuracy based on the deviations of their predicted rankings from the audience's actual rankings for the set of videos they evaluated. This way the results reflected how many ideas participants incorrectly ranked above or below each idea in their set. Because ranking ten videos on multiple dimensions would have been too cumbersome for participants, ratings were collected from participants and then converted to rankings.

I calculated participants' predicted rankings by taking the average of their ratings on the three aforementioned forecasting items for each video they rated ($\alpha = .88$). All the videos they rated were then ranked: the highest-rated video was ranked first, and the lowest-rated video was ranked last (the distances between ranks did not significantly differ by role). The actual audience rankings were calculated by averaging the standardized scores for the liking, sharing, and financial dimensions of success captured in the audience survey, yielding a composite measure of success for each of the 161 videos ($\alpha = .79$; the three dimensions were highly correlated with one another—see Table 2).

For the specific set of videos rated by each participant, the predicted rank was subtracted from the actual rank for each video rated. This produced an inaccuracy score for each video, where zero represented a correct ranking, negative values represented false negatives (ranking too low), and positive values represented false positives (ranking too high). Thus, inaccuracy scores represented the extent to which participants' predicted rank deviated from the correct rank, and scores closer to zero implied greater accuracy than scores further away from zero. The inaccuracy scores were used to shed light on the relative likelihood of false positives versus false negatives in testing H2 and H3, as these hypotheses address the directionality of errors. However, because the inaccuracy scores added up to zero for each participant (except when participants predicted ties between videos[4]), to measure overall accuracy for the other hypotheses, I calculated overall accuracy scores by subtracting the absolute value of each inaccuracy score from ten. This way the overall accuracy scores ranged from one to ten for each video, and a higher mean score for a given participant was indicative of greater overall accuracy (participants' mean scores could range from five to ten). For creators who submitted videos and completed the forecasting survey, two sets of inaccuracy and overall accuracy scores were calculated: one without their own video (for H1-H2 and H4-H5) and one with their own video (for H3).

Inaccuracy and overall accuracy scores were calculated in this way using the composite forecasts and measure of success (i.e., the liking, sharing, and financial dimensions averaged together), as well as for each of the three dimensions separately. The results were consistent across the composite measures and each of the three dimensions, so for simplicity, the composite

---

[4] Although there were no ties between videos in the actual audience rankings, some participants (41.72 percent) had ties between videos in their predicted rankings. To account for this, any ties were given the same rank—e.g., if two videos were tied for third place, both videos were ranked third, and then the next video in the rankings was ranked fifth. This ensured that participants were penalized for predicting ties, which were incorrect according to the audience data.

results are reported below. Most participants (71.98 percent) were scored on all ten videos that they rated. Because some videos initially included in the forecasting survey were taken offline before the audience survey was conducted, some participants were scored on fewer than ten videos. However, all participants were scored on at least six videos (in addition to their own video if they submitted one), and the average was 9.62 videos per participant ($SD = 0.77$). When relevant in the analyses, I controlled for the number of videos on which participants were scored.

Results

Creators' advantage over managers (H1 and H2). To test H1 (creators are more accurate than managers at forecasting the success of others' ideas), I used repeated-measures ANOVA—see Models 1 and 2 in Table 3. For estimation accuracy (Model 1), the main effect for role was significant, $F(3, 4390.40) = 5.05$, $p < .01$. All roles underestimated the percentage of the audience who would share the videos they rated (see Figure 2). However, consistent with H1, Fisher LSD tests showed that creators ($M = -3.92$) underestimated significantly less than managers ($M = -7.85$), $p < .001$, but not significantly different from hybrids ($M = -5.51$), $p = .31$, or laypeople ($M = -4.41$), $p = .64$. In addition, laypeople underestimated significantly less than managers, $p = .001$. No other comparisons between roles were significant. In support of H1, these results suggest that managers underestimated the percentage of the audience who would share videos more than creators (and laypeople).

For placement accuracy (Model 2), the main effect for role was also significant, $F(3, 3970.22) = 3.44$, $p < .05$. As predicted, Fisher LSD tests showed that creators ($M = 7.59$) were significantly more accurate than managers ($M = 7.39$), $p = .01$, and laypeople ($M = 7.36$), $p < .01$, but not significantly different from hybrids ($M = 7.51$), $p = .49$. Managers, hybrids, and laypeople did not significantly differ from one another (see Figure 3). In support of H1, these

results suggest that creators were more accurate than managers in terms of both estimation and placement (and creators' advantage remained significant with a variety of additional controls—see Footnote B in Table 3).

It is worth noting that most managers were previously creators (61.67 percent), but creators' advantage over managers was still significant after controlling for creator and manager experience. Using laypeople as the baseline, the estimation results suggest that the manager role may have dragged estimates down, while the creator role may have kept estimates relatively up. For placement accuracy, the results suggest that the creator role may have produced a boost in accuracy, but when creators became managers, the manger role may have undermined this boost (see Figures 2 and 3). Also, hybrids scored in between creators and managers, which is consistent with the notion that the balance of idea generation versus evaluation—and thus divergent versus convergent thinking—in the role is a key driver of creators' advantage over managers. See Appendix B for supplementary tests of H1.

[Insert Table 3 and Figures 2 and 3 about here]

To test H2 (creators are less likely than managers to commit false negatives regarding novel ideas), I used Models 3 and 4 in Table 3. For estimation accuracy (Model 3), the interaction between role and novelty was significant, $F(3, 1924.90) = 12.86, p < .001$. As predicted, Fisher LSD tests showed that when videos were novel (one standard deviation above the mean in novelty), creators ($M = -1.37$) underestimated the sharing percentage less than managers ($M = -11.47$), $p < .001$, and laypeople ($M = -4.84$), $p < .05$, but not significantly different from hybrids ($M = -3.76$), $p = .27$. Managers also underestimated significantly more than both hybrids ($p < .01$) and laypeople ($p < .001$). When ideas were conventional (one standard deviation below the mean in novelty), creators ($M = -6.55$), managers ($M = -4.19$),

hybrids (M = -7.49), and laypeople (M = -3.98) did not significantly differ from one another. In support of H2, these results suggest that creators' advantage over managers (and laypeople) in estimation accuracy was greater when ideas were relatively novel.

For placement accuracy (Model 4), the role-novelty interaction was also significant, $F(3, 1802.95) = 5.83$, $p < .01$. As predicted, Fisher LSD tests showed that when videos were novel, creators (M = -0.26) were significantly less inclined to commit false negatives than managers (M = -0.68), $p < .05$, and laypeople (M = -0.58), $p < .05$, but not significantly different from hybrids (M = -0.13), $p = .60$. In support of H2, these results suggest that creators' advantage over managers and laypeople in placement accuracy was partially thanks to creators avoiding false negatives regarding novel ideas (and conversely, avoiding false positives regarding more conventional ideas[5]). In sum, H1 and H2 were supported in terms of estimation and placement.

Creators overvaluing their own ideas (H3). To test H3a (creators tend to commit false positives regarding their own ideas) and H3b (giving managers the advantage regarding creators' own ideas), I used data from the 119 creators who submitted videos and completed the forecasting survey.[6] Consistent with H3a, creators overestimated the percentage of the audience who would share their own video by 3.61 percent (SD = 28.51), and underestimated the percentage of the audience who would share others' videos by 3.46 percent (SD = 24.83), which was a significant difference, $t(118) = 3.04$, $p < .01$, $d = .26$. In terms of placement accuracy, as predicted, creators placed their own videos significantly too high relative to others' videos (M =

---

[5] When ideas were conventional, all roles tended to commit false positives. However, creators (M = 0.35) were significantly less inclined to commit false positives than managers (M = 0.83), $p < .01$, but creators and managers did not significantly differ from laypeople (M = 0.51) or hybrids (M = 0.32). Taken together, the results suggest that managers were more likely than creators to incorrectly place conventional ideas above better novel ideas and better average ideas (ideas closer to the mean in novelty), whereas laypeople were more likely than creators to incorrectly place average ideas—but not conventional ideas—above better novel ideas.

[6] The results for H3a and H3b were all still significant when only the 76 creators who were scored on ten or fewer videos including their own were in the analyses (43 of the 119 creators were scored on 11 videos including their own). Also, when the 15 hybrids who submitted videos and completed the forecasting survey were included with creators, the results remained significant for H3a and H3b.

1.99, $SD = 3.43$), $t(118) = 6.33$, $p < .001$, $d = 1.16$. In support of H3a, these results suggest that creators overvalued their own ideas in terms of both estimation and placement.

To test H3b, I used data from of the 64 creators who completed the forecasting survey and had their videos included in the forecasting survey. In terms of estimation, unlike the broader sample of 119 creators, these 64 creators underestimated the percentage of the audience who would share their own videos ($M = -2.52$, $SD = 26.99$), and this was not significantly different from managers' underestimation for the same videos ($M = -7.61$, $SD = 11.98$), $t(63) = 1.45$, $p = .15$. These results do not support H3b in terms of estimation accuracy, and suggest that the creators who did not have videos qualify for the forecasting survey drove the overestimation effect in the test of H3a. However, in terms of placement accuracy, as predicted in H3b, creators were significantly less accurate regarding their own videos ($M = 1.67$, $SD = 3.19$) than managers were for the same videos ($M = 0.29$, $SD = 2.50$), $t(63) = 3.86$, $p < .001$, $d = .48$. Thus, due to creators' tendency to rank their videos too high compared to others' videos, managers were significantly more accurate. These results suggest that managers had the advantage regarding creators' own ideas in terms of placement, but not estimation, providing partial support for H3b.

Past success and idea quality (H4 and H5). To test H4 (past success reduces creators' accuracy in forecasting about others' ideas) and H5 (bad ideas that succeed reduce accuracy more than bad ideas that fail), I used data from the 132 creators and hybrids who submitted videos with publicly available view counts and completed the forecasting survey. The ANOVA models testing H4 and H5 are in Table 4. Regarding H4, the main effect for past success was not significant for estimation accuracy (Model 5), but was significant for placement accuracy (Model 6), $b = .10$, $t(198.44) = 2.20$, $p < .05$. However, with additional controls, past success was significant in the predicted direction for estimation accuracy, $b = -2.70$, $t(299.99) = -2.71$, $p <$

.01, but no longer significant for placement accuracy (see Footnotes C and D in Table 4).

Consistent with H4, these results suggest that more past success was associated with less

accuracy for both estimation and placement, but these relationships were relatively tenuous.

Regarding H5, the interaction between past success and quality was not significant for

estimation accuracy (Model 7), but was significant for placement accuracy (Model 8), $b = -.16$,

$t(175.09) = -2.67$, $p < .01$. As predicted, the simple slope was positive and significantly different

from zero when quality was low, $b = .33$, $t(126) = 3.69$, $p < .001$, but not significantly different

from zero when quality was high, $b = .01$, $t(126) = .16$, $p = .88$ (see Figure 4). These results

suggest that in terms of placement, creators with bad videos that succeeded anyway were less

accurate than creators with equally bad videos that failed (and creators with good videos were

not adversely affected by their past success). These results support H5 regarding placement, but

not estimation. This is consistent with the theory underlying H5, as the argument is that creators

who succeed with bad ideas get misleading feedback on what constitutes high-quality ideas in

the domain, which should affect their ability to discern between ideas more than their ability to

estimate the absolute success of any one idea.

In sum, all the hypotheses (H1-H5) were supported in terms of placement accuracy, and

H1-H3a and H4 were also supported in terms of estimation accuracy (although the support for

H4 was relatively weak for both estimation and placement).

[Insert Table 4 and Figure 4 about here]

STUDY 2

Overview

To test the proposed mechanism driving the main argument—i.e., creators have an

advantage over managers in forecasting about others' novel ideas (H1 and H2)—I conducted a

lab experiment. Due to the cross-sectional nature of Study 1, the proposed causal mechanism could not be directly tested. It is possible, for instance, that dispositional differences between creators and managers drove the effects, rather than the nature of their roles. I designed the experiment to target the proposed theoretical mechanism driving creators' advantage over managers: the notion that the creator role focuses on generating one's own novel ideas (and thus divergent thinking), while the manager role focuses on evaluating others' ideas (and thus convergent thinking). Participants were randomly assigned to roles and then asked to rank a set of product ideas according to how successful they expected each to be in the consumer marketplace. The product ideas were pretested to enable scoring of placement accuracy (estimation accuracy was not measured separately, as the experiment was designed such that estimation accuracy would not provide theoretical insight beyond placement accuracy).

The experiment was designed to build on Study 1 in three key ways. First, the circus arts may be an atypical context, as success may depend on novelty to an unusual extent. Thus, the creative forecasting task in Study 2 focused on a more typical industry: consumer products. Second, the results of Study 1 suggest that when managers become creators, their focus on convergent thinking may undermine the benefit of the divergent thinking they did as creators. However, because none of the participants started as managers and then became creators, Study 1 could not test whether starting with convergent thinking may undermine the benefit of then engaging in divergent thinking. As a result, Study 1 could not address whether managers may improve their accuracy by engaging in more divergent thinking through generating their own novel ideas. To test the effect of starting with divergent versus convergent thinking, Study 2 includes two hybrid conditions (creator-manager and manager-creator), along with conditions for creator, manager, and lay/control.

Third, in Study 1, laypeople had less domain knowledge than creators and managers, and creators may have had less domain knowledge than managers, making it difficult to know whether the results were driven by differences in domain knowledge or roles. In Study 2, by focusing the task on consumer products, all participants were likely to enter the experiment with some domain knowledge. Because participants were then randomly assigned to conditions, the degree of domain knowledge should have been consistent across the conditions. Thus, I was able to test how the role manipulations shaped the way participants used their domain knowledge in creative forecasting, while keeping the amount of domain knowledge constant across roles.

Participants and Procedures

A total of 206 participants were recruited through a subject pool at a large private university in the Northeast U.S (66.0 percent female, M_{age} = 22.19, SD_{age} = 6.21). The experiment was designed to promote both psychological and mundane realism (Berkowitz & Donnerstein, 1982). To foster psychological realism, participants were informed that the goal of the task at hand was to learn how members of the university could help support innovation. To foster mundane realism, the task was framed around the goal of bringing patentable ideas to the consumer marketplace. In particular, the survey began with the following for all participants:

> *"Like many universities, [this university] is interested in supporting innovation in the world at large. We are trying to learn how the university can help faculty, staff, and students bring their most innovative ideas to the marketplace. We would appreciate your help in learning more about this. One of the ways the university can support innovation is by helping faculty, staff, and students obtain patents for their most promising product ideas. To be patentable, ideas must be novel compared to existing products. Ideally, the university and its members would focus on securing patents for novel ideas that end up succeeding in the marketplace. The first part of this survey relates to this challenge."*

Then, participants were randomly assigned to one of five conditions: creator, manager, creator-manager, manager-creator, and lay/control. In the creator condition, participants were asked to spend ten minutes generating three different novel product ideas that were likely to

succeed in the consumer marketplace (and thus engage in divergent and convergent thinking). In the manager condition, participants were asked to spend ten minutes describing three different criteria for evaluating how successful novel product ideas will be in the consumer marketplace (and thus engage in convergent thinking). In the two hybrid conditions, participants spent five minutes on each manipulation: creator-managers experienced the creator manipulation first followed by the manager manipulation, and vice-versa for manager-creators. To help control the amount of time participants spent on their ideas/criteria, the survey would not advance until ten minutes had passed (five minutes for each manipulation in the hybrid conditions). In the lay/control condition, participants were not asked to do either of these tasks—after reading the introductory passage above, they moved to the next survey page. Thus, the lay/control condition provided a baseline level of accuracy absent of additional divergent or convergent thinking.

On the next survey page, participants in all conditions were asked to rank a set of four product ideas according to how successful they predicted each one to be in the consumer marketplace (1 = most successful, 4 = least successful). The order in which the ideas were presented to participants was randomized. The set of four ideas was selected from a pretest of 16 product ideas, which were rated by a separate sample of 403 consumers recruited through MTurk, all from the United States (46.2 percent female, M_{age} = 32.90, SD_{age} = 10.82). The 16 ideas were all based on products that had already been patented in the United States, but were not familiar to a broad audience because they had not been commercialized or widely marketed yet. Descriptions of the products from the actual patents were each summarized to 63 words, and a visual from each patent was included along with the 63-word summary (see Table 5).

[Insert Table 5 about here]

Pretest Measures

The pretest included measures of audience success and novelty. All items were rated on the same Likert-type scale (1 = "Strongly Disagree" to 7 = "Strongly Agree"). Following the logic of the wisdom of crowds (Clemen, 1989; Larrick & Soll, 2006; Surowiecki, 2005), research shows that a crowd of consumers' intent to purchase ideas for new products tends to be an accurate predictor of subsequent market success, even better than predictions by industry experts (Kornish & Ulrich, 2014). Accordingly, audience success was measured by capturing audience members' purchase intent using three items: "I want to own this product," "I am likely to buy this product," and "I would purchase this product if it was sold at a fair price" (α = .96). Novelty was measured using three items adapted from the Creative Product Semantic Scale (O'Quin & Besemer, 1989): "This product is [novel, unique, original]" (α = .86).

The pretest results were used to select the set of four ideas that participants ranked in the experiment. The four ideas were selected such that there was a correct ranking according to the audience (the first-ranked idea pretested significantly higher in success than the second-ranked idea, the second-ranked idea pretested significantly higher in success than the third-ranked idea, and so on). In addition, the first- and third-ranked ideas were relatively novel, while the second- and fourth-ranked ideas were relatively conventional (the two novel ideas did not significantly differ from one another in novelty, nor did the two conventional ideas). By including a relatively good and a relatively bad novel idea (and a relatively worse conventional idea ranked below each one), the design enabled a test of whether creators tend to value all novel ideas regardless of their potential success, or whether they can discern between novel ideas with varying levels of potential success. See Table 6 for the means and results from the pretest.

[Insert Table 6 about here]

Results

To test whether the five conditions varied in terms of participants' rankings of the four ideas, I used contingency table analyses. This revealed that participants' rankings significantly varied by condition for the novel idea that should have been ranked first, χ^2 (12) = 27.15, $p < .01$, and the conventional idea that should have been ranked second, χ^2 (12) = 22.75, $p < .05$, but not for the ideas that should have been ranked third, χ^2 (12) = 15.84, $p = .20$, or fourth, χ^2 (12) = 14.63, $p = .26$. Thus, the conditions varied significantly in their rankings of the two good ideas, but not the two bad ideas.

Next, I compared the five conditions in terms of whether participants ranked the best idea first, and thus ahead of the second-best idea (avoiding a false negative regarding the best idea)— see Figure 5. As expected, creators (77.50 percent) were significantly more likely to correctly rank the best idea than managers (51.16 percent), χ^2 (3) = 9.88, $p < .05$, creator-managers (41.03 percent), χ^2 (3) = 20.56, $p < .001$, and laypeople (56.10 percent), χ^2 (3) = 9.71, $p < .05$, but not manager-creators (65.12 percent), χ^2 (3) = 3.05, $p = .38$. In addition, manager-creators were significantly more likely to correctly rank the best idea than creator-managers, χ^2 (3) = 10.19, $p < .05$, and numerically (but not significantly) more likely to correctly rank the best idea than managers, χ^2 (3) = 2.66, $p = .45$ and laypeople, χ^2 (3) = 3.46, $p = .33$.

[Insert Figure 5 about here]

These results constructively replicate the results from Study 1 on creators' advantage over managers in forecasting about others' novel ideas. By itself, the creator manipulation gave participants a significant advantage over managers and laypeople. The pattern of results suggests that recency effects occurred for the two hybrid conditions—i.e., the manipulation that came second seemed to drive participants' forecasting more than the one that came first. Even after

experiencing the manager manipulation first, the creator manipulation helped participants correctly identify the best idea (and thus avoid a false negative). As in Study 1, managers and laypeople did not significantly differ, suggesting that the manager manipulation by itself did not undermine accuracy. However, when the manager manipulation followed the creator manipulation, the manager manipulation undermined the advantage of the creator manipulation, as creators and manager-creators were both significantly better than creator-managers. Thus, divergent thinking led to a boost in accuracy, but this boost disappeared when individuals then focused on convergent thinking. This is consistent with the result from Study 1 that prior experience as a creator did not make managers significantly more accurate, as their current role as a manager may have undermined any advantage they once had as a creator.

In general, these results provide causal evidence that the nature of creator and manager roles may be key in driving creators' advantage over managers in forecasting accuracy. The additional time creators spend engaging in divergent thinking may account for at least some of their advantage over managers (and laypeople). However, when creators become managers, their focus on convergent thinking may undermine the benefit of the divergent thinking they did as a creator, making them no better at creative forecasting than an average layperson.

GENERAL DISCUSSION

Using a field study in the circus industry and a lab experiment, I examined the effects of creator and manager roles in forecasting the audience success of new ideas. Results suggest that creators were more accurate than managers at forecasting about others' novel ideas. However, creators overvalued their own ideas, and their advantage over managers regarding others' ideas was undermined when they had low-quality ideas that achieved success anyway. These results have key implications for theory and research on creativity and innovation in organizations.

Theoretical Implications

  Creative forecasting as a bridge between creativity and innovation. The literatures on creativity (generating creative ideas) and innovation (successfully implementing creative ideas) have remained largely separate (Anderson, Potočnik, & Zhou, 2014). However, scholars have begun to shed light on the relatively unexplored space in between creativity and innovation. The work done so far has focused on the effects of network position in driving whether and how creators are able to win the social support needed to implement their ideas (e.g., Baer, 2012; Lingo & O'Mahony, 2010; Perry-Smith & Mannucci, 2015). For instance, Baer (2012) proposed that maintaining strong network ties helps creators garner "buy-in" for their ideas. The current studies build on this prior work by adding creative forecasting to the set of key processes that occur in between generating creative ideas and implementing them.

  After creative ideas are generated, employees and managers are likely to engage in creative forecasting as they try to predict how successful the ideas will be with the intended audience, which is often consumers outside of the organization. These creative forecasts are likely to play an important part in whether new ideas are able to garner the social support they need to actually be implemented. The present studies complement previous work by suggesting that in addition to network structures, role structures may also influence the level of social support that creative ideas receive. While prior work addresses the network structures that enable creators to win social support for their ideas, the present work addresses how role structures may shape whether novel ideas win the level of social support they deserve. This extends prior work by bringing the notion of accuracy into the picture. Higher-quality ideas may deserve more social support than lower-quality ideas, but discerning between the two requires accurate creative forecasting by the individuals in creators' networks.

Across both studies, results suggest that individuals in creator roles are more likely to give promising novel ideas the support they deserve than individuals in manager roles. However, since managers usually wield more power than creators, good novel ideas that deserve to be implemented may fail to garner the necessary social support due to managers' inaccurate creative forecasting. In general, these studies suggest that creative forecasting is a key part of the bridge between creativity and innovation in organizations, particularly in determining whether novel ideas receive the level of social support they deserve. In so doing, this research advances the emerging literature on the space between creativity and innovation in organizations.

Roles in creativity and innovation. These studies have important implications for theory and research on role designs in creativity and innovation. Although scholars have paid much attention to describing and understanding each stage of the innovation process, relatively little attention has been paid to the effects of dividing the innovation process into different roles in organizations (cf. Elsbach & Kramer, 2003; Mollick, 2012). The present studies draw attention to potential inefficiencies stemming from the design of the roles that typically handle variation (creators) and selection (managers) in the innovation process. Evolutionary views of creativity and innovation imply that the best ideas will be selected within organizations over time (Aldrich, 1999; Campbell, 1960; Nelson & Winter, 1982; Simonton, 1999; Staw, 1990). However, the results from these studies suggest that by having managers specialize in selection, the role may (ironically) hinder managers' ability to select the best ideas. Typically, creators are only asked to evaluate their own ideas, and managers are the sole gatekeepers in deciding which ideas reach the audience (e.g., Kaplan, 2008). Yet, these studies suggest that creators are bad at forecasting the success of their own ideas, but better than managers at forecasting about others' ideas.

Across the two studies, creators were more accurate than both managers and laypeople in

forecasting about others' ideas, but managers were no better than laypeople. This suggests that the creator role may be better designed to foster accurate creative forecasting than manager and layperson roles, as long as creators are not forecasting about their own ideas. Whereas managers and laypeople may both rely too heavily on conventional models of success, the additional time creators spend on divergent thinking may help them recognize novel ways in which ideas may succeed. The fact that hybrids scored in between creators and managers supports the notion that the balance of divergent versus convergent thinking in the role is key in driving forecasting accuracy. Furthermore, the notion that laypeople were less accurate than creators at forecasting their fellow laypeople's reactions suggests that merely asking individuals to evaluate novel ideas—without first generating their own novel ideas—may foster false negatives.

Taken together, these results present a paradoxical tradeoff: specializing in variation may give creators an advantage over managers in selection, but assigning creators more selection (thus turning them into hybrids) may undermine this advantage. This suggests that role designs may have unintended consequences when it comes to creative forecasting, answering the call for more research on the paradoxical dynamics involved in creativity within organizations (George, 2007). In general, these results shed light on the importance of role designs in creative forecasting, extending prior work on the effects of creator and manager roles in creativity and innovation (Elsbach & Kramer, 2003; Mollick, 2012). In addition, the results hint that roles may have important—and perhaps unexpected—effects in other realms of creativity and innovation. Thus, in future research, scholars may benefit from further exploring the impact of roles in creativity and innovation, such as the effects of role conflict (Rothbard, 2001), role-making (Graen, 1976), or using roles as resources (Baker & Faulkner, 1991).

Sustaining creativity and innovation over time. Abundant empirical evidence across a diverse array of domains suggests that creative success usually spikes and then declines (Audia & Goncalo, 2007; Bayus, 2013; Dennis, 1966; Simonton, 1997; Stephan & Levin, 1993; Stern, 1978). Scholars have proposed theories to explain this puzzling rise and fall in creative success. Simonton's (1997, 1999) blind variation and selective retention model posits that the decline in creative success is due to a decline in the quantity of ideas produced. Other scholars have argued that when creators develop successful ideas, they fixate on them, undermining the novelty—and thus success—of any ideas they subsequently generate (Audia & Goncalo, 2007; Bayus, 2013).

While these theories offer compelling explanations for why high levels of creativity and innovation are rarely sustained, they focus on the generation of ideas, overlooking the role of idea evaluation. The current studies suggest that along with losing their ability to generate creative ideas over time, creators may also lose their ability to accurately evaluate new ideas, as their ideas achieve disproportionate levels of success or they get promoted to managerial roles (perhaps based on their success) that leave them prone to false negatives. This may form a vicious cycle, as creators may continue to generate ideas that are decreasingly creative, but remain blissfully unaware that their creativity has declined (Kruger & Dunning, 1999). Even if creators are able to generate better ideas, they may not be able to recognize them as better than less promising ideas. In this way, idea generation and evaluation may work together in making creativity and innovation difficult to sustain, and considering the dynamics between them may provide a more comprehensive view of how creativity and innovation can be sustained over time.

Limitations and Future Directions

These studies have key limitations that may be addressed in future research. First, both studies focused on ideas that were relatively developed and above a minimum threshold of

quality, but creative forecasting may be different with more nascent or lower-quality ideas. Because forecasting about relatively developed ideas may be easier than forecasting about more nascent ideas, it is possible that these studies provided more conservative tests of the hypotheses than if they had focused on more nascent ideas (and thus creators' advantage over managers may be even larger with less developed ideas). To uncover key moderators and boundary conditions, future research should examine creative forecasting in other contexts, with other types of ideas at different levels of quality and stages of their development.

Second, although Study 2 provided some causal evidence of the proposed mechanism driving creators' advantage over managers in forecasting about others' novel ideas, divergent and convergent thinking were not directly measured in Study 1. Moreover, other differences between creator and manager roles may be key in creative forecasting, such as differences in power (Sligte, de Dreu, & Nijstad, 2011), status (Duguid & Goncalo, 2015), or accountability pressures (Tetlock, 1983). These issues can be addressed in future lab or field experiments in which divergent and convergent thinking are manipulated, along with other key aspects of roles. Third, in both studies, the level of success that ideas could achieve was bounded. In actuality, some innovations may be exponentially more successful than others. The ideas included in the studies likely represented relatively ordinary innovations in the circus arts (Study 1) and consumer products (Study 2) domains. Creative forecasting may be different when the ideas in question represent extraordinary innovations, with potential payoffs that are exponentially larger than more ordinary innovations (Taleb & Tetlock, 2013). Future longitudinal research could explore creative forecasting in entrepreneurial contexts where the best ideas perform exponentially better than average ideas.

These studies also raise promising new research questions related to creative forecasting. First, while these studies focused on forecasting the success of new ideas with consumers, future research could explore creative forecasting with other audiences and outcomes, such as forecasting the staying power (Hargadon & Douglas, 2001), viral potential (Berger & Milkman, 2012), or cultural impact (Rao, Monin, & Durand, 2003) of new ideas. Second, the results suggest that both creators and managers face challenges that are likely to make them worse forecasters over time (or at least not better), highlighting the importance of more research on what they can do to improve their accuracy over time. Third, because we designed the contest incentives in Study 1 to minimize the risk of biasing the results by offering prizes for both the top videos and judges, we could not isolate the effects of competition or different types of incentives on forecasting accuracy. While the results suggest that participation in the contest was not significantly related to forecasting accuracy, future experimental research could explore how competition and extrinsic incentives may influence forecasting accuracy. Lastly, these studies suggest that using manager ratings of employee creativity may undervalue some novel ideas. Future research could compare manager, peer, and audience ratings to explore when manager ratings of creativity are more or less valid.

Practical Implications and Conclusion

This research offers insights that may help creators, managers, and organizations looking to improve their accuracy in creative forecasting. The results suggest that creators should be harsher critics of their own ideas—perhaps not early in their development, but after the ideas have been well developed and creators must decide whether to continue with them or not. And even after their ideas succeed, creators should be careful not to attribute all the success to the nature of the idea itself. Instead, they may benefit from looking for indicators of underlying

quality across ideas, rather than focusing on market success. Managers, on the other hand, may benefit from engaging in more idea generation in their jobs, as this may help them avoid false negatives when evaluating novel ideas. Lastly, organizations may want to take steps to leverage the wisdom creators have of their peers' ideas. Instead of having managers be the sole gatekeepers to innovation, organizations could involve creators in selecting which ideas get implemented. In essence, this research suggests that the best judges of others' ideas in organizations may be those who are busy generating their own ideas.

REFERENCES

Aldrich, H.
1999   "Organizations evolving." Thousand Oaks, CA: Sage.

Amabile, T. M.
1988   "A model of creativity and innovation in organizations." In B. M. Staw, and L. L. Cummings (eds.), Research in organizational behavior: 123-167. Greenwich, CT: JAI Press.

Amabile, T. M.
1996   Creativity in context. Boulder, CO: Westview Press.

Anderson, N., K. Potočnik, and J. Zhou
2014   "Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework." Journal of Management, 40: 1297-1333.

Audia, P. G., and S. Brion
2007   "Reluctant to change: Self-enhancing responses to diverging performance measures." Organizational Behavior and Human Decision Processes, 102: 255-269.

Audia, P. G., and J. A. Goncalo
2007   "Past success and creativity over time: A study of inventors in the hard-disk drive industry." Management Science, 53: 1-15.

Audia, P. G., E. A. Locke, and K. G. Smith
2000   "The paradox of success: An archival and a laboratory study of strategic persistence following radical environmental change." Academy of Management Journal, 43: 837-853.

Baer, M.
2012   "Putting creativity to work: The implementation of creative ideas in organizations." Academy of Management Journal, 55: 1102-1119.

Baer, M., and G. Brown
2012   "Blind in one eye: How psychological ownership of ideas affects the types of suggestions people adopt." Organizational Behavior and Human Decision Processes, 118: 60-71.

Baker, W. E., and R. R. Faulkner
1991   "Role as resource in the Hollywood film industry." American Journal of Sociology, 97: 279-309.

Basadur, M., M. A. Runco, and L. A. Vega
2000   "Understanding how creative thinking skills, attitudes and behaviors work together: A causal process model." Journal of Creative Behavior, 34: 77-100.

Bayus, B. L.
2013   "Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community."

Management Science, 59: 226-244.

Beggan, J. K.
1992   "On the social nature of nonsocial perception: The mere ownership effect." Journal of Personality and Social Psychology, 62: 229.

Benner, M. J., and M. L. Tushman
2003   "Exploitation, exploration, and process management: The productivity dilemma revisited." Academy of Management Review, 28: 238-256.

Berger, J., and K. L. Milkman
2012   "What makes online content viral?" Journal of Marketing Research, 49: 192-205.

Berkowitz, L., and E. Donnerstein
1982   "External validity is more than skin deep: Some answers to criticisms of laboratory experiments." American Psychologist, 37: 245-257.

Brislin, R. W.
1986   "The wording and translation of research instruments." In W. J. Lonner, and J. W. Berry (eds.), Field methods in cross-cultural research. Cross-cultural research and methodology series: 137-164. Thousand Oaks, CA: Sage.

Buhrmester, M., T. Kwang, and S. D. Gosling
2011   "Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?" Perspectives on Psychological Science, 6: 3-5.

Burgelman, R. A.
1991   "Intraorganizational ecology of strategy making and organizational adaptation: Theory and field research." Organization Science, 2: 239-262.

Burt, R. S.
2004   "Structural holes and good ideas." American Journal of Sociology, 110: 349-399.

Byrne, C. L., A. S. Shipman, and M. D. Mumford
2010   "The effects of forecasting on creative problem-solving: An experimental study." Creativity Research Journal, 22: 119-138.

Campbell, D. T.
1960   "Blind variation and selective retentions in creative thought as in other knowledge processes." Psychological Review, 67: 380.

Cirque du Soleil
2015   "Cirque du Soleil at a glance." http://www.cirquedusoleil.com/en/home/about-us/at-a-glance.aspx.

Clemen, R. T.

1989   "Combining forecasts: A review and annotated bibliography." International Journal of Forecasting, 5: 559-583.

Cropley, A.
2006   "In praise of convergent thinking." Creativity Research Journal, 18: 391-404.

Csaszar, F. A.
2012   "Organizational structure as a determinant of performance: Evidence from mutual funds." Strategic Management Journal, 33: 611-632.

Csikszentmihalyi, M.
1999   "Implications of a systems perspective for the study of creativity." In R. J. Sternberg (ed.), Handbook of creativity: 313-335. New York: Cambridge University Press.

Dailey, L., and M. D. Mumford
2006   "Evaluative aspects of creative thought: Errors in appraising the implications of new ideas." Creativity Research Journal, 18: 385-390.

Dane, E.
2010   "Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective." Academy of Management Review, 35: 579-603.

Dennis, W.
1966   "Creative productivity between the ages of 20 and 80 years." Journal of Gerontology, 21: 1-8.

Drazin, R., M. A. Glynn, and R. K. Kazanjian
1999   "Multilevel theorizing about creativity in organizations: A sensemaking perspective." Academy of Management Review, 24: 286-307.

Duguid, M. M., and J. A. Goncalo
2015   "Squeezed in the middle: The middle status trade creativity for focus." Journal of Personality and Social Psychology, 109: 589-603.

Elsbach, K. D., and R. M. Kramer
2003   "Assessing creativity in Hollywood pitch meetings: Evidence for a dual-process model of creativity judgments." Academy of Management Journal, 46: 283-301.

Finke, R. A.
1996   "Imagery, creativity, and emergent structure." Consciousness and Cognition, 5: 381-393.

Fleming, L.
2001   "Recombinant uncertainty in technological search." Management Science, 47: 117-132.

Fleming, L., S. Mingo, and D. Chen
2007   "Collaborative brokerage, generative creativity, and creative success." Administrative Science Quarterly, 52: 443-475.

Forbes
2014   "The world's billionaires: Guy Laliberte." http://www.forbes.com/profile/guy-laliberte.

Ford, C. M.
1996   "A theory of individual creative action in multiple social domains." Academy of Management Review, 21: 1112-1142.

George, J. M.
2007   "Creativity in organizations." Academy of Management Annals, 1: 439 - 477.

Glăveanu, V. P.
2013   "Rewriting the language of creativity: The Five A's framework." Review of General Psychology, 17: 69-81.

Gosling, S. D., P. J. Rentfrow, and W. B. Swann
2003   "A very brief measure of the big-five personality domains." Journal of Research in Personality, 37: 504-528.

Graen, G. B.
1976   "Role-making processes within complex organizations." In M. D. Dunnette (ed.), Handbook of Industrial/Organizational Psychology: 1201-1245. Chicago: Rand McNally.

Greenwald, A. G.
1980   "The totalitarian ego: Fabrication and revision of personal history." American Psychologist, 35: 603.

Guilford, J. P.
1967   The nature of human intelligence. New York: McGraw-Hill.

Hargadon, A. B., and Y. Douglas
2001   "When innovations meet institutions: Edison and the design of the electric light." Administrative Science Quarterly, 46: 476-501.

Hoffman, D. L., P. K. Kopalle, and T. P. Novak
2010   "The "right" consumers for better concepts: Identifying consumers high in emergent nature to develop new product concepts." Journal of Marketing Research, 47: 854-865.

Kaplan, S.
2008   "Framing contests: Strategy making under uncertainty." Organization Science, 19: 729-752.

KC, D., B. R. Staats, and F. Gino
2013   "Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery." Management Science, 59: 2435-2449.

Kim, W. C., and R. Mauborgne

2004   "Blue ocean strategy." Harvard Business Review, 82: 76-84.

Kornish, L. J., and K. T. Ulrich
2014   "The importance of the raw idea in innovation: Testing the sow's ear hypothesis." Journal of Marketing Research (JMR), 51: 14-26.

Kruger, J., and D. Dunning
1999   "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments." Journal of Personality and Social Psychology, 77: 1121-1134.

Kunda, Z.
1990   "The case for motivated reasoning." Psychological Bulletin, 108: 480.

Larrick, R. P., and J. B. Soll
2006   "Intuitions about combining opinions: Misappreciation of the averaging principle." Management Science, 52: 111-127.

LeBreton, J. M., and J. L. Senter
2008   "Answers to 20 questions about interrater reliability and interrater agreement." Organizational Research Methods, 11: 815.

Leslie, D., and N. M. Rantisi
2011   "Creativity and place in the evolution of a cultural industry: The case of Cirque du Soleil." Urban Studies, 48: 1771-1787.

Levinthal, D. A., and J. G. March
1993   "The myopia of learning." Strategic management journal, 14: 95-112.

Licuanan, B. F., L. R. Dailey, and M. D. Mumford
2007   "Idea evaluation: Error in evaluating highly original ideas." Journal of Creative Behavior, 41: 1-27.

Lingo, E. L., and S. O'Mahony
2010   "Nexus work: Brokerage on creative projects." Administrative Science Quarterly, 55: 47-81.

Lubart, T. I.
2001   "Models of the creative process: Past, present and future." Creativity Research Journal, 13: 295-308.

Madsen, P. M., and V. Desai
2010   "Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry." Academy of Management Journal, 53: 451-476.

Mainemelis, C.
2010   "Stealing fire: Creative deviance in the evolution of new ideas." Academy of Management

Review, 35: 558-578.

Merton, R. K.
1968   "The Matthew effect in science." Science, 159: 56.

Miller, D. T., and M. Ross
1975   "Self-serving biases in the attribution of causality: Fact or fiction?" Psychological bulletin, 82: 213.

Miner, A. S.
1994   "Seeking adaptive advantage: Evolutionary theory and managerial action." In J. A. C. Baum, and J. V. Singh (eds.), Evolutionary dynamics of organizations: 76-89. New York: Oxford University Press.

Mintzberg, H.
1971   "Managerial work: Analysis from observation." Management Science, 18: B-97-B-110.

Mollick, E.
2012   "People and process, suits and innovators: The role of individuals in firm performance." Strategic Management Journal, 33: 1001-1015.

Moore, D. A., and P. J. Healy
2008   "The trouble with overconfidence." Psychological Review, 115: 502-517.

Moreau, C. P., D. R. Lehmann, and A. B. Markman
2001   "Entrenched knowledge structures and consumer response to new products." Journal of Marketing Research, 38: 14-29.

Mueller, J. S., S. Melwani, and J. A. Goncalo
2012   "The bias against creativity: Why people desire but reject creative ideas." Psychological Science, 23: 13-17.

Mumford, M. D., D. C. Lonergan, and G. Scott
2002   "Evaluating creative ideas: Processes, standards, and context." Inquiry: Critical Thinking Across the Disciplines, 22: 21-30.

National Circus School
2014   "Circus arts disciplines." http://www.nationalcircusschool.ca/en/artiste/circus-arts-disciplines.

Nelson, R. R., and S. G. Winter
1982   "Organizational capabilities and behavior." An evolutionary theory of economic change: 96-136. Cambridge, MA: Harvard University Press.

Nemeth, C. J.
1997   "Managing innovation: When less is more." California Management Review, 40: 59-74.

O'Quin, K., and S. P. Besemer
1989   "The development, reliability, and validity of the revised creative product semantic scale." Creativity Research Journal, 2: 267-278.

Oliver, M. B., and A. Bartsch
2010   "Appreciation as audience response: Exploring entertainment gratifications beyond hedonism." Human Communication Research, 36: 53-81.

Paolacci, G., and J. Chandler
2014   "Inside the turk: Understanding mechanical turk as a participant pool." Current Directions in Psychological Science, 23: 184-188.

Perry-Smith, J., and P. V. Mannucci
2015   "From creativity to innovation: The social network drivers of the four phases of the idea journey." Academy of Management Review.

Pfeffer, J., and C. T. Fong
2005   "Building organization theory from first principles: The self-enhancement motive and understanding power and influence." Organization Science, 16: 372-388.

Pierce, J. L., T. Kostova, and K. T. Dirks
2001   "Toward a theory of psychological ownership in organizations." Academy of Management Review, 26: 298-310.

Pinchot, G. I.
1985   Intrapreneuring. New York: Harper & Row.

Rao, H., P. Monin, and R. Durand
2003   "Institutional change in Toque Ville: Nouvelle cuisine as an identity movement in French gastronomy." American Journal of Sociology, 108: 795-843.

Reeves, C. A., and D. A. Bednar
1994   "Defining quality: Alternatives and implications." Academy of Management Review, 19: 419-445.

Rietzschel, E. F., B. A. Nijstad, and W. Stroebe
2010   "The selection of creative ideas after individual idea generation: Choosing between creativity and impact." British Journal of Psychology, 101: 47-68.

Ross, L., D. Greene, and P. House
1977   "The false consensus effect: An egocentric bias in social perception and attribution processes." Journal of Experimental Social Psychology, 13: 279-301.

Ross, M., and F. Sicoly
1979   "Egocentric biases in availability and attribution." Journal of Personality and Social Psychology, 37: 322-336.

Rothbard, N. P.
2001   "Enriching or depleting? The dynamics of engagement in work and family roles."
Administrative Science Quarterly, 46: 655-684.

Runco, M., and W. Smith
1992   "Interpersonal and intrapersonal evaluations of creative ideas." Personality and Individual
Differences, 13: 295-302.

Runco, M. A.
1991   "The evaluative, valuative, and divergent thinking of children." Journal of Creative Behavior, 25:
311-319.

Salganik, M. J., P. S. Dodds, and D. J. Watts
2006   "Experimental study of inequality and unpredictability in an artificial cultural market." Science,
311: 854-856.

Silvia, P. J.
2008   "Discernment and creativity: How well can people identify their most creative ideas?"
Psychology of Aesthetics, Creativity, and the Arts, 2: 139-146.

Simonton, D. K.
1997   "Creative productivity: A predictive and explanatory model of career trajectories and
landmarks." Psychological Review, 104: 66-89.

Simonton, D. K.
1999   "Creativity as blind variation and selective retention: Is the creative process Darwinian?"
Psychological Inquiry, 10: 309.

Sitkin, S. B.
1992   "Learning through failure: The strategy of small losses." Research in Organizational Behavior,
14: 231-266.

Sligte, D. J., C. K. W. de Dreu, and B. A. Nijstad
2011   "Power, stability of power, and creativity." Journal of Experimental Social Psychology, 47: 891-
897.

Staw, B. M.
1990   "An evolutionary approach to creativity and innovation." In M. West, and J. L. Farr (eds.),
Innovation and creativity at work: Psychological and organizational strategies: 287-308. New York:
Wiley.

Staw, B. M., and J. Ross
1989   "Understanding behavior in escalation situations." Science, 246: 216.

Stephan, P. E., and S. G. Levin

1993   "Age and the Nobel Prize revisited." Scientometrics, 28: 387-399.

Stern, N.
1978   "Age and achievement in mathematics: A case-study in the sociology of science." Social Studies of Science, 8: 127-140.

Sternberg, R. J., and T. I. Lubart
1991   "An investment theory of creativity and its development." Human development, 34: 1-31.

Surowiecki, J.
2005   The wisdom of crowds. New York: Anchor.

Taleb, N. N., and P. E. Tetlock
2013   "On the difference between binary prediction and true exposure with implications for forecasting tournaments and decision making research." SSRN 2284964.

Tetlock, P.
1983   "Accountability and complexity of thought." Journal of Personality and Social Psychology, 45: 74-83.

Tetlock, P. E., and B. A. Mellers
2011   "Intelligent management of intelligence agencies: Beyond accountability ping-pong." American Psychologist, 66: 542-554.

Ward, T. B.
1994   "Structured imagination: The role of category structure in exemplar generation." Cognitive Psychology, 27: 1-40.

Ward, T. B., S. M. Smith, and R. A. Finke
1999   "Creative cognition." In R. J. Sternberg (ed.), Handbook of Creativity: 189-212. New York: Cambridge University Press.

Weick, K. E.
1979   The social psychology of organizing, 2 ed. Reading, MA: Addison-Wesley.

Wilson, R. C., J. P. Guilford, P. R. Christensen, and D. J. Lewis
1954   "A factor-analytic study of creative-thinking abilities." Psychometrika, 19: 297-311.

TABLE 1

Study 1: Professional Circus Arts Experience by Current Role

Current Role[a]	Mean Years (*SD*)		
	Current Role	Creator	Manager
Creator	6.52	6.84	1.37
(n = 177, 44.1% female)	(5.91)	(6.24)	(3.70)
Manager[b]	8.40	7.15	9.16
(n = 120, 42.5% female)	(8.33)	(8.84)	(8.07)
Hybrid - Creator/Manager	9.92	11.53	6.80
(n = 42, 55.0% female)	(7.79)	(8.03)	(6.32)
Layperson	n/a	n/a	n/a
(n = 150, 50.0% female)			

Notes:

[a] Of the 177 creators, 137 participated in the contest. Of the 42 hybrids, 15 participated in the contest. Among the circus professionals, approximately 24 percent were from Canada, 17 percent were from Germany, 14 percent were from the United States, 10 percent were from France, and the remaining 35 percent were spread between 38 other countries, each representing less than 4 percent of the sample. To account for the international nature of the sample, participants could complete the forecasting survey in English, French, or German. The English version was translated into French and German, and then back translated (Brislin, 1986) to ensure the meaning remained consistent.

[b] The majority of managers (61.67 percent) were creators for at least a year before becoming a manager.

TABLE 2

Study 1: Means, Standard Deviations, and Correlations for Video Measures

	Mean	SD	1	2	3	4	5	6	7	8	9
1. Audience Liking	5.15	0.50									
2. Audience Sharing[a]	49.82%	6.85%	.66***								
3. Audience Financial[b]	4.75	0.48	.55***	.46***							
4. Novelty	4.78	1.13	.25**	.12	.18*						
5. Past Success[c]	12,313	54,619	.24**	.09	.13	.03					
6. Future Success Rate[c]	8.87	47.22	.12	.17*	.15	.09	-.04				
7. Days Online	325.51	472.63	.08	-.04	.04	.09	.44***	-.11			
8. Length (Minutes)	5.63	2.02	.02	.04	.11	.13	-.10	.00	.06		
9. # of Male Performers	1.01	1.50	.15	.08	.08	-.03	.25**	.30***	.05	.07	
10. # of Female Performers	0.74	0.84	.08	.02	-.03	.19*	.02	.13	-.09	-.11	-.09

Notes:

*p < .05, ** p < .01, *** p < .001

[a] Audience sharing was used as the measure of quality in testing H5.

[b] Number of cents out of ten given to performer(s) in video.

[c] Past success was the number of views each video had prior to the study. Interestingly, quality (i.e., audience sharing) was not significantly correlated with past success. However, to measure future success, I recorded the number of views on each video one year after past success was measured, just before the audience survey launched. To control for the degree of exposure and thus opportunities to be shared each video already had, future success rate was calculated as the number of views gained over the year, divided by the number of initial views (i.e., past success). Quality was significantly correlated with future success rate. This supports Salganik, Dodds, and Watts' (2006) assertion that although quality is a significant driver of success online, much of the variance in success may be driven by other, less predictable factors.

TABLE 3

Study 1: Repeated-Measures ANOVA Models for H1 and H2

Variable[a]	Hypothesis 1		Hypothesis 2	
	Model 1 DV = Estimation	Model 2 DV = Placement (Overall)	Model 3 DV = Estimation	Model 4 DV = Placement (Inaccuracy)
Intercept	-3.92***	7.58***	-3.96***	0.05
	(0.72)	(0.05)	(0.72)	(0.08)
Role (manager vs. creator)[b]	**-3.93***	-0.19***	-3.87***	0.02
	(1.10)	**(0.08)**	(1.10)	(0.13)
Role (hybrid vs. creator)	-1.59	-0.08	-1.67	0.05
	(1.57)	(0.11)	(1.56)	(0.18)
Role (lay vs. creator)	-0.49	-0.22**	-0.45	-0.08
	(1.07)	(0.08)	(1.06)	(0.12)
Number of videos scored	0.28	-0.18***	0.25	0.01
	(0.44)	(0.03)	(0.44)	(0.05)
Novelty			2.60***	-0.30***
			(0.67)	(0.08)
Role (manager vs. creator) X Novelty[b]			**-6.23***	**-0.45***
			(1.03)	**(0.12)**
Role (hybrid vs. creator) X Novelty			-0.73	0.08
			(1.53)	(0.18)
Role (lay vs. creator) X Novelty			-3.02**	-0.24*
			(0.97)	(0.11)

Notes: *$p < .05$, ** $p < .01$, *** $p < .001$;

[a] Both continuous variables were standardized (number of videos scored and novelty). To control for the fact that each participant rated a different set of videos, video was included as a repeated factor clustered by participant in all models. Since this produced 100 different parameter estimates (one for each video), the results are not reported here. However, the Wald Z statistic was significant at $p < .01$ for all 100 videos in all four models.

[b] With only creators and managers in the same four models, the highlighted results remained significant with additional controls, including gender, age, years as pro circus creator, years as pro circus manager, years in current role, nationality, big-five traits (Gosling, Rentfrow, & Swann, 2003), contest participation (yes or no), feedback report (opted to receive one or not), video length, video live audience (yes or no), and number of male and female performers in the video.

TABLE 4

Study 1: Repeated-Measures ANOVA Models for H4 and H5

Variable[a]	Hypothesis 4		Hypothesis 5	
	Model 5 DV = Estimation	Model 6 DV = Placement (Overall)	Model 7 DV = Estimation	Model 8 DV = Placement (Overall)
Intercept	-2.96**	7.58***	-3.04**	7.59***
	(1.00)	(0.06)	(1.00)	(0.06)
Role (hybrid vs. creator)[b]	-0.34	-0.10	0.20	-0.18
	(2.87)	(0.17)	(2.92)	(0.17)
Number of videos scored	0.29	-0.27***	0.17	-0.25***
	(0.59)	(0.05)	(0.60)	(0.05)
Past success[c]	**0.36**	**-0.10***	1.03	-0.17**
	(0.80)	**(0.05)**	(0.91)	(0.05)
Quality			-1.33	0.04
			(0.95)	(0.06)
Past success X Quality[d]			**-1.45**	**0.16***
			(1.05)	**(0.06)**

Notes: *p < .05, ** p < .01, *** p < .001

[a] All continuous variables were standardized. Video was included as a repeated factor clustered by participant in all four models, which produced 99 different parameter estimates (one video included in the broader sample was only rated by one participant in this subset and was thus omitted). The Wald Z statistic was significant at p < .05 for at least 90 of the videos in all four models.

[b] Because 13 of the 132 participants were hybrids (the rest were creators), I controlled for role.

[c] In Model 5, past success was significant with additional controls (b = -2.70, p < .01), including gender, age, years as pro circus creator, years as pro circus manager, years in current role, nationality, big-five traits (Gosling, Rentfrow, & Swann, 2003), contest participation (yes or no), feedback report (opted to receive one or not), video length, video live audience (yes or no), number of male and female performers in video, own video days online, own video length, own video live audience, and number of male and female performers in own video. However, with these extra controls, past success was no longer significant in Model 6.

[d] With the additional controls listed above, the interaction between past success and quality was still not significant in Model 7, but remained significant in Model 8.

TABLE 5

Study 2: Product Ideas Ranked

Note: The US Patent #'s were not shown to participants. The top-ranked idea was also tested in a study by Hoffman, Kopalle, and Novak (2010), who also found it to be both novel and likely to succeed with consumers.

1. Storage Device for Delivery and Pickup of Goods An outdoor storage device that secures goods from theft and exposure to the elements. Provides an electronic notification to user that goods have been delivered and/or picked up. The storage device includes an enclosure for securely enclosing the goods and a computerized communication apparatus for providing notification that goods have been delivered or picked up. (US Patent #5,774,053)	
2. Modular Rack System A kit of parts for constructing a modular rack assembly that can be adjusted in width and depth for different spaces and contents. The kit includes a variety of accessories, such as extension poles, hooks, and shelves that are easily applied and removed from the main rack assembly. The main rack and accessories are selectively reconfigurable for evolving storage needs. (US Patent #8,550,263)	
3. Automatic Bed Maker An electrically-operated bed making apparatus secured to the frame of a bed. This apparatus automatically makes up a bed by causing a framework of arms to be lifted from the foot of the bed to the head by utilizing a pair of rotating wheels on either side of the bed. Consumers can program the operation and movement of the apparatus. (US Patent #5,926,874)	
4. Hands Free Towel Carrying System A towel of a generally rectangular configuration comprised of an absorbent material. Has a loose end, a parallel coupling end, and a pair of side edges. The elastic neck loop is circular and can be stretched around a user's head without messing up the user's hair and glasses and can be formed for enlargement and reduction in size. (US Patent #6,718,554)	

TABLE 6

Study 2: Pretest Means and Results by Condition

Correct Rank	Pretest Novelty [paired *t*-test with idea below]	Pretest Success [paired *t*-test with idea below]	Percentage of Participants with Correct Rank				
			Creator	Manager	Manager-Creator	Creator-Manager	Lay
1. Storage Device for Delivery and Pickup of Goods *(Good novel idea)*	5.04 (1.25) [*t*(402) = 15.33, *p* < .001]	4.04 (1.82) [*t*(402) = 6.53, *p* < .001]	77.5% (31/40)	51.2% (22/43)	65.1% (28/43)	41.0% (16/39)	56.1% (23/41)
2. Modular Rack System *(Moderately good conventional idea)*	3.82 (1.36) [*t*(402) = -15.29, *p* < .001]	3.38 (1.63) [*t*(402) = 3.93, *p* = .001]	55.0% (22/40)	32.6% (14/43)	55.8% (24/43)	33.3% (13/39)	51.2% (21/41)
3. Automatic Bed Maker *(Moderately bad novel idea)*	5.18 (1.38) [*t*(402) = 12.93, *p* < .001]	2.94 (1.93) [*t*(402) = 9.66, *p* < .001]	50.0% (20/40)	34.9% (15/43)	48.8% (21/43)	51.3% (20/39)	39.0% (16/41)
4. Hands Free Towel Carrying System *(Bad conventional idea)*	3.98 (1.62)	2.00 (1.26)	70.0% (28/40)	58.1% (25/43)	74.4% (32/43)	61.5% (24/39)	65.9% (27/41)

Note: The two novel ideas did not significantly differ from one another in pretest novelty, nor did the two conventional ideas.
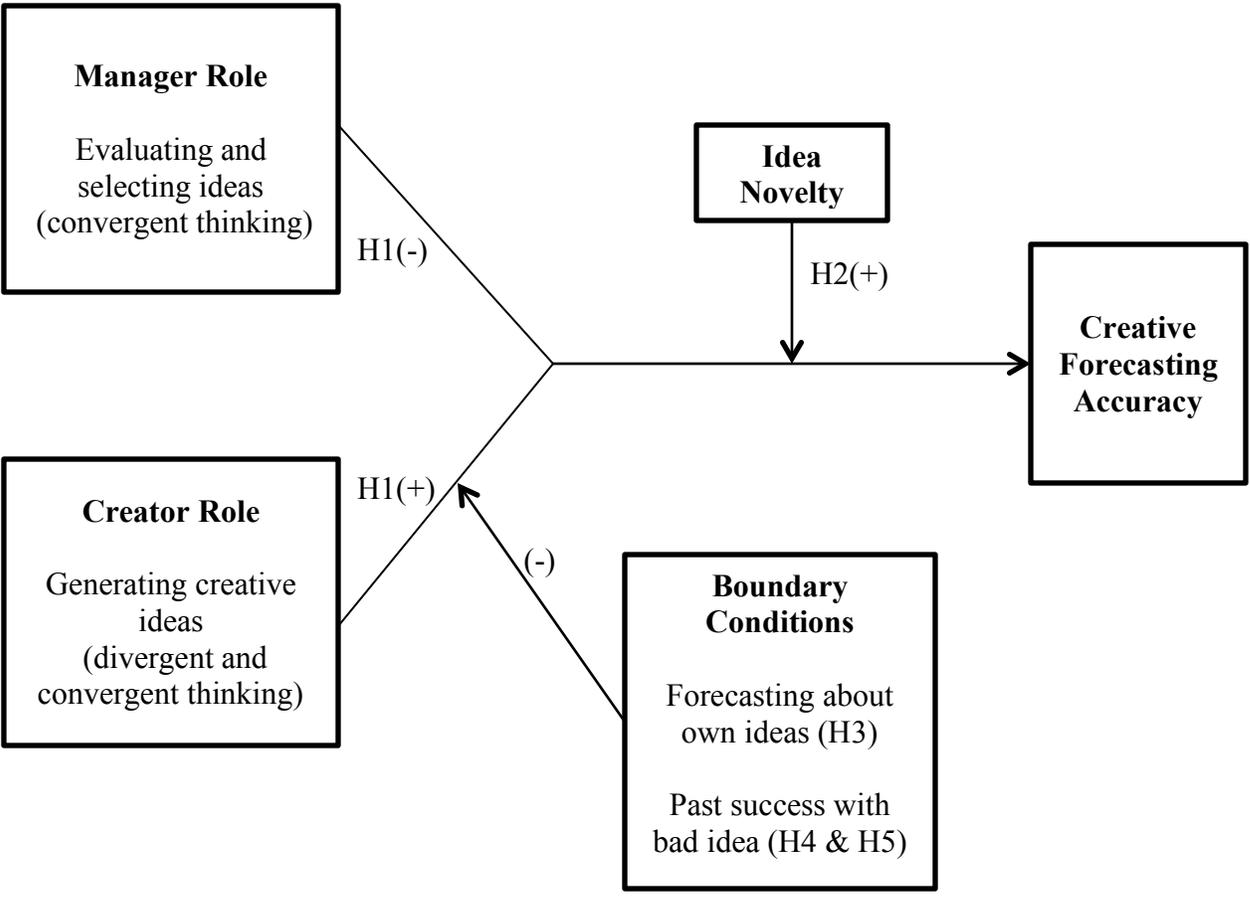
FIGURE 1

Visual Summary of Hypotheses

FIGURE 2

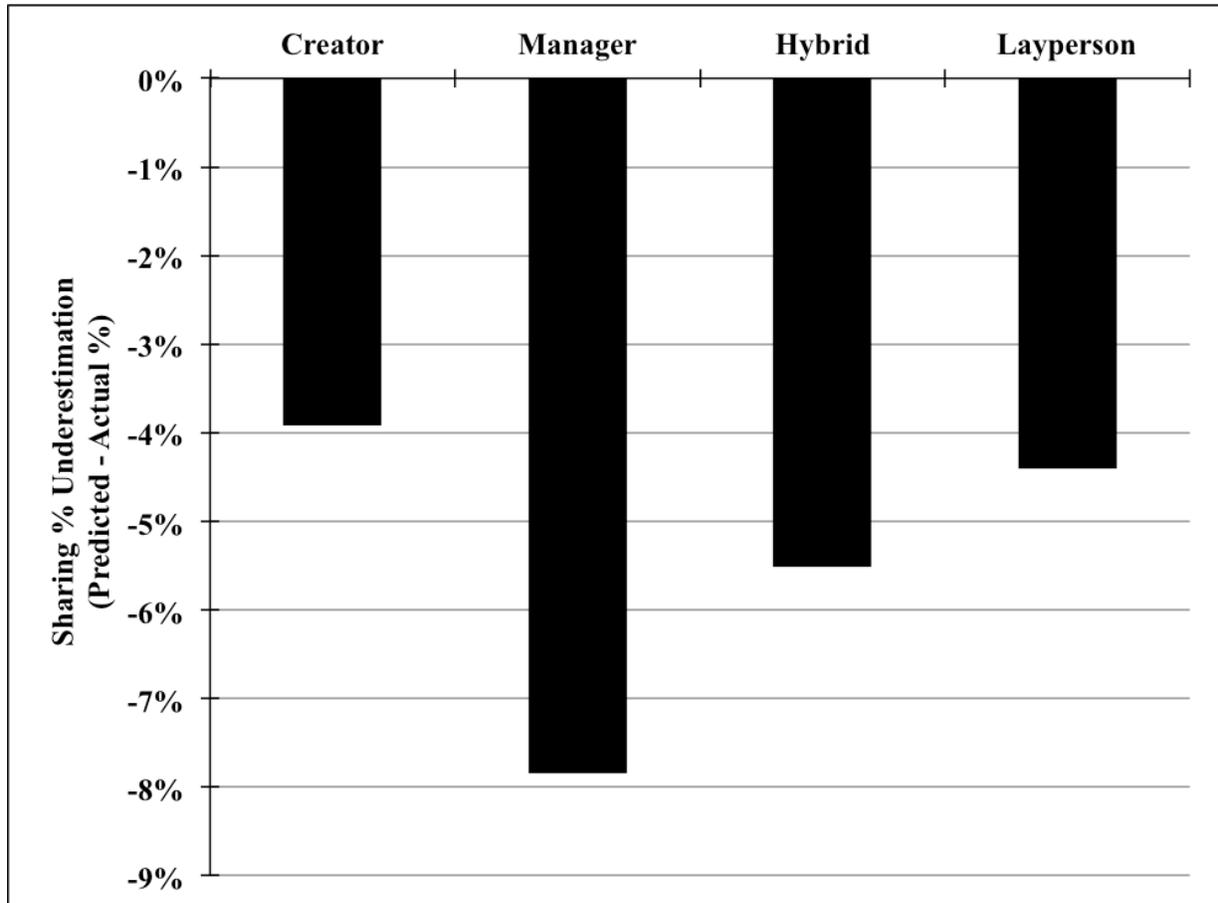**Study 1: Estimation Accuracy by Role (Estimated Marginal Means for H1)**

FIGURE 3

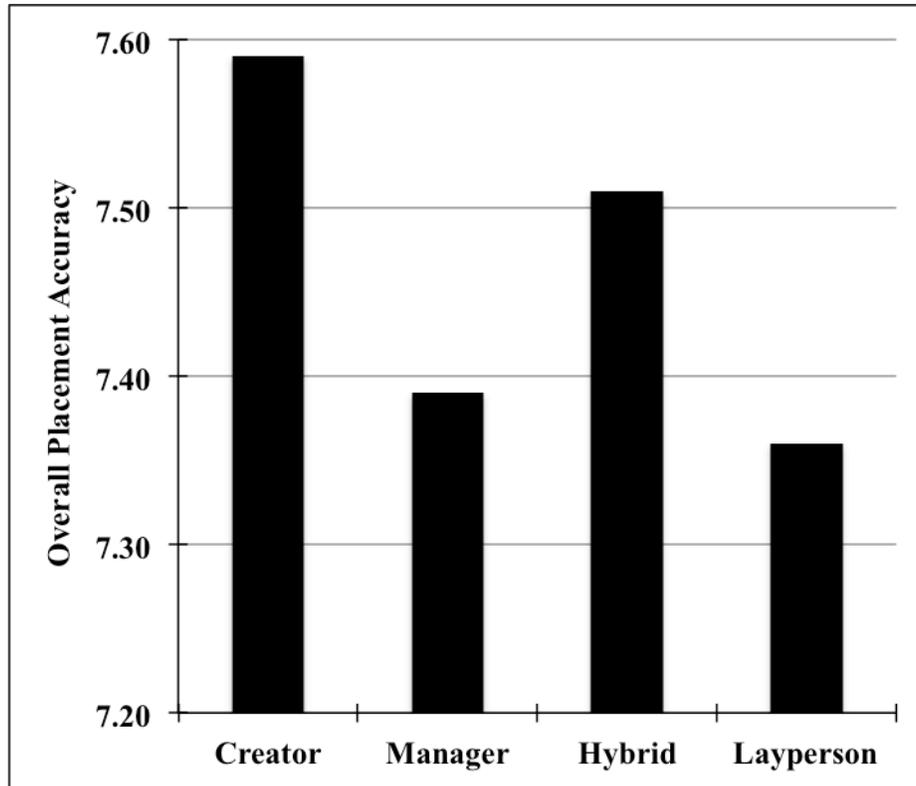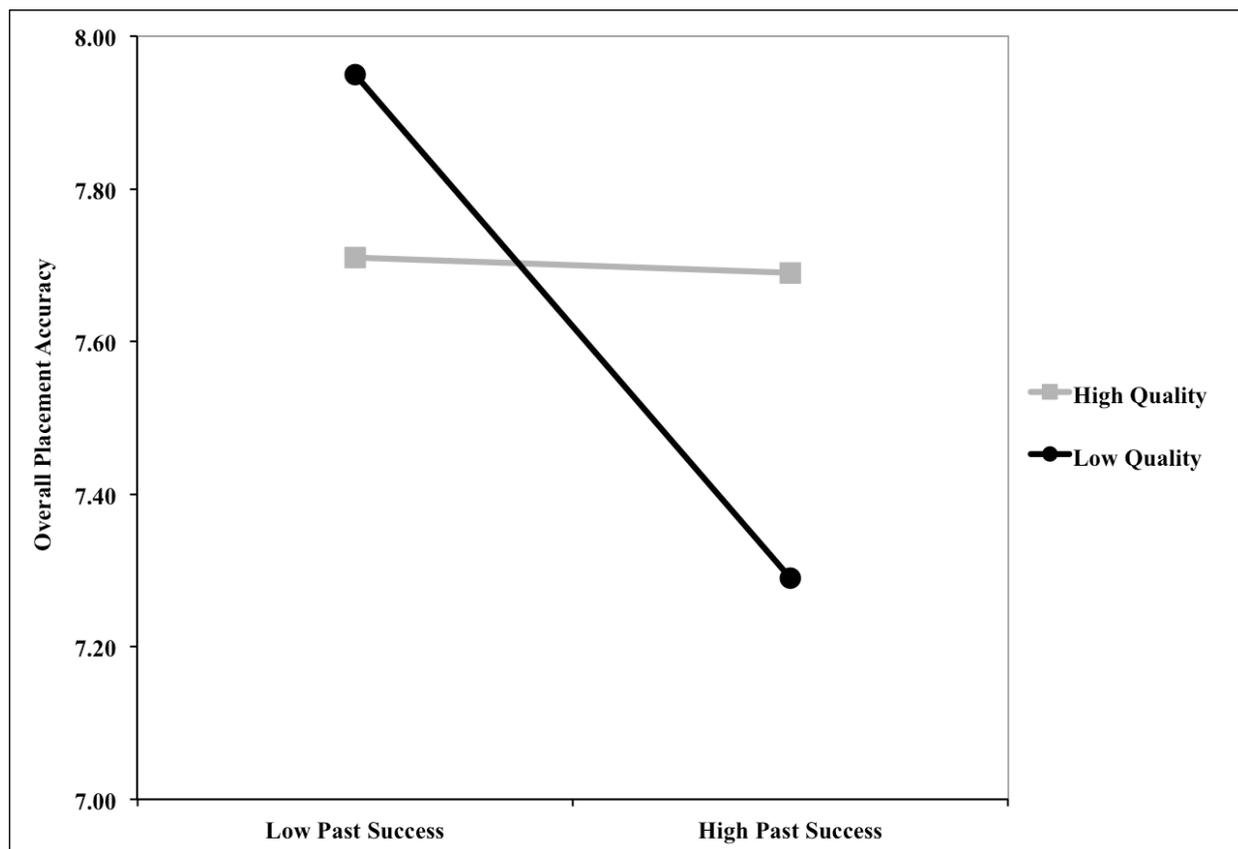**Study 1: Overall Placement Accuracy by Role (Estimated Marginal Means for H1)**
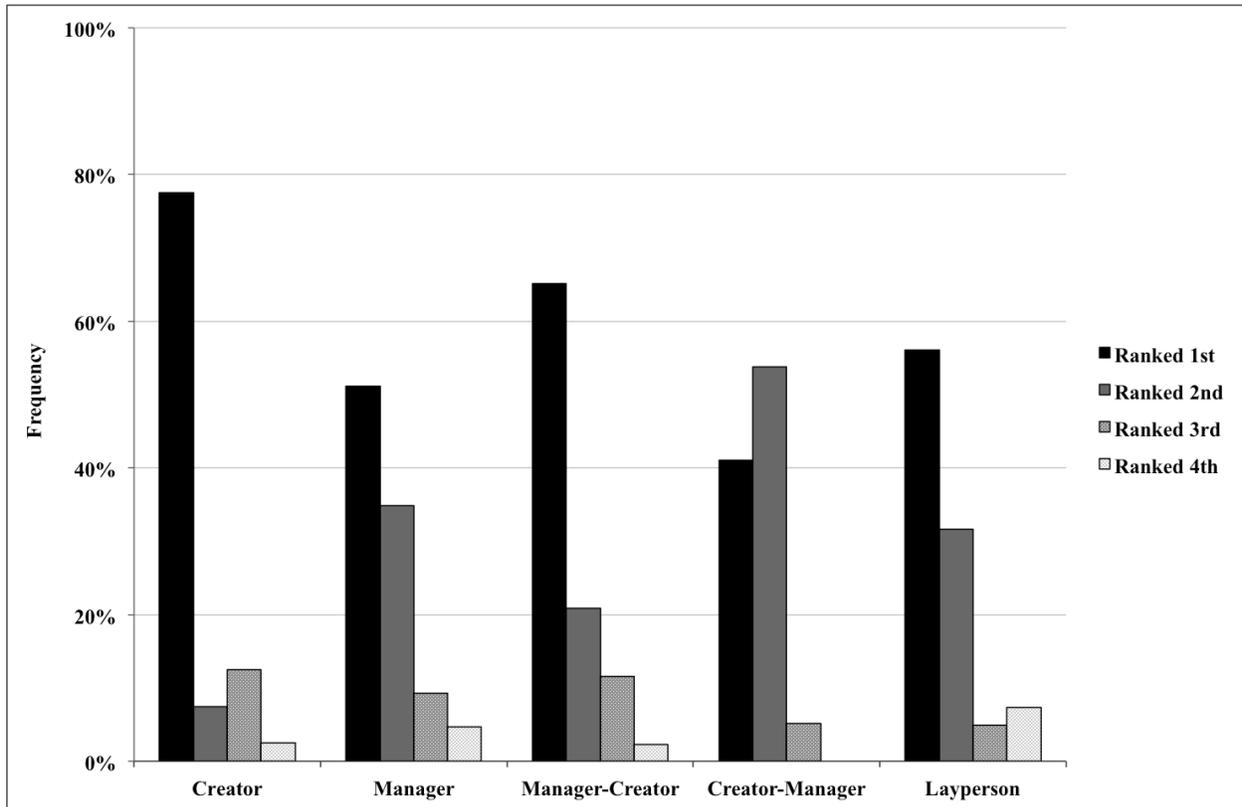
FIGURE 4

Study 1: Past Success X Quality Predicting Overall Placement Accuracy (Simple Slopes for H5)

Note: High and low refer to one standard deviation above and below the mean in past success and quality.
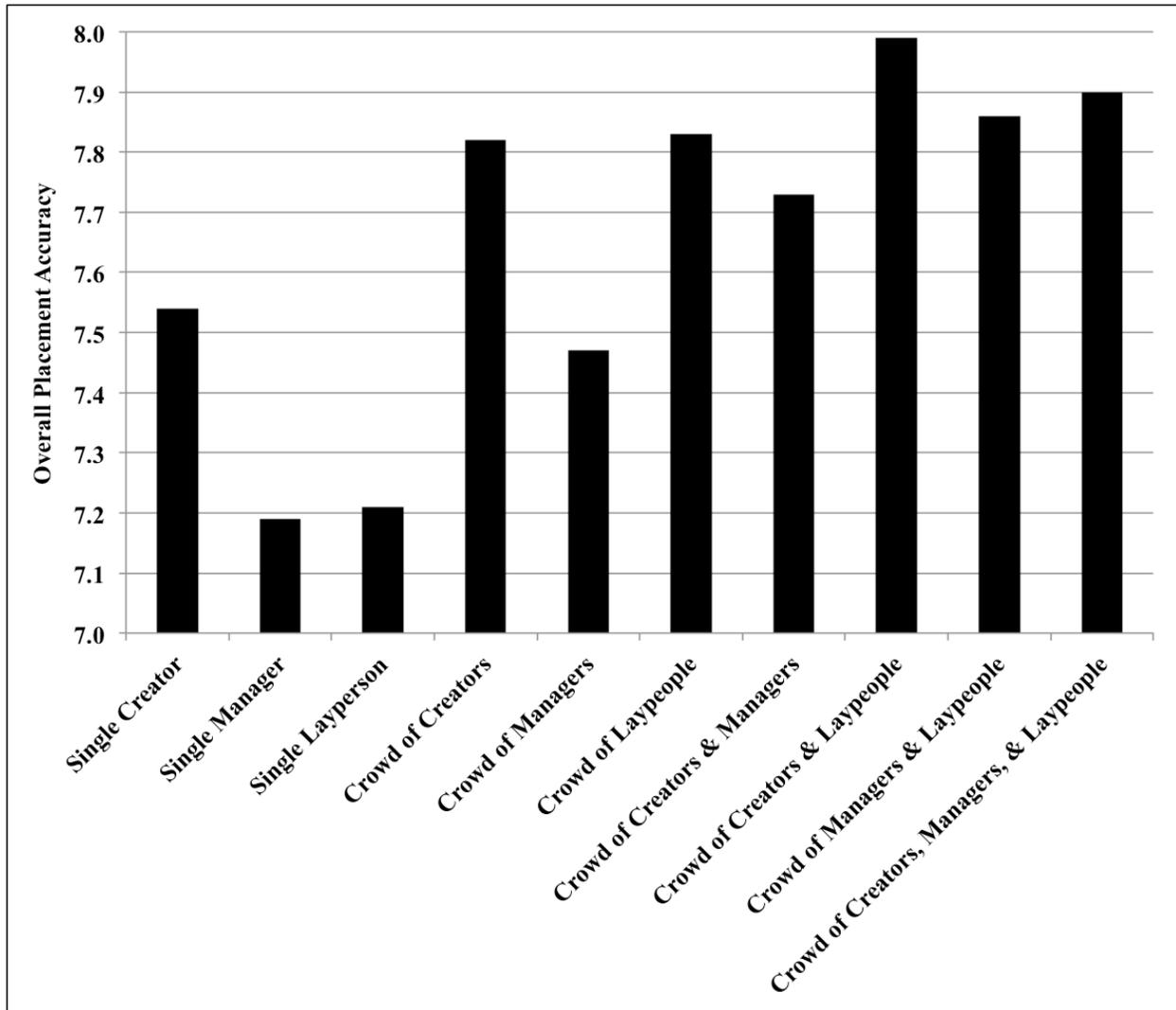
FIGURE 5

Study 2: Rankings of the Best Idea by Condition

Note: This graph shows participants' rankings of the novel idea that should have been ranked first, thus the "Ranked 1st" bars represent correct rankings and any lower rankings represent false negatives regarding the best idea.

FIGURE 6

Study 1: Overall Placement Accuracy by Role and Crowd

APPENDIX A

Study 1: Video Collection and Selection Pretest

We collected the videos of circus acts included in Study 1 by hosting an online circus video contest. Through announcements on social media, we invited circus creators to submit online videos of their acts to the contest. A total of 324 videos were submitted by 266 creators (42 creators submitted multiple videos—no one submitted more than 5 videos). To ensure the number of forecasts would not be spread too thin between videos, we selected a subset of these videos to serve as the performances on which participants made forecasts in the forecasting survey. To select this subset of videos, we administered a pretest survey to the 266 creators who submitted videos. They were each asked to watch and rate ten randomly selected videos—181 creators/hybrids completed the pretest survey (68.05 percent response rate). Of the 181 who completed the pretest survey, 152 completed the forecasting survey (83.98 percent response rate; 57.14 percent effective response rate).

In the pretest, we trimmed the pool of videos to a subset of 100 videos using four selection criteria. First, we omitted videos submitted by creators who did not complete the pretest survey, and thus could not provide informed consent to use their video in the study. Second, since some videos would not display properly in certain countries due to copyright laws on the music, we omitted videos that did not play properly for more than half of the creators who tried to view them. This ensured that every qualifying video was rated by at least three creators in the pretest.

Third, to ensure only videos of stage acts were included, we omitted videos with means lower than 5.00 on the item "This video showed an act that can be performed on stage" (1 = "Strongly Disagree" to 7 = "Strongly Agree")—thus the cutoff of five implied at least moderate

agreement. This omitted acts that could only exist in video form. Fourth, to simulate managers'

typical experience of evaluating and selecting between acts that could realistically be included in

future shows, we set a minimum threshold on the quality of ideas included. To set a quality

threshold but still capture a fairly wide range of quality, we omitted videos that scored below the

mean of 4.56 ($SD = 1.20$) on the item "I liked this video," which creators rated on the same

Likert-type scale. These four criteria narrowed the subset to 109 videos, but nine of these videos

were taken offline by their owners before the study concluded, bringing the final subset included

in the forecasting survey to 100 videos. A total of 61 videos were rated below the mean in liking,

but met the other three criteria. Since these 61 videos were rated by the creators/hybrids who

submitted them, they were included in the audience survey to help test H3-H5 (making a total of

161 videos included in the study).

APPENDIX B

Study 1: Supplementary Analyses

Hypothesis 1. As a further test of H1 in terms of placement accuracy, I tested whether the results for H1 replicate when the raw ratings were used, as opposed to the rankings implied by the ratings. I calculated the correlation between each participant's raw forecasts (the three forecasting items averaged together) and the raw audience results (the z-scores of the three dimensions of audience success averaged together) for the set of videos he or she rated. I then calculated the average correlation for participants in each role. In terms of Pearson correlations, creators ($M = .32$, $SD = .31$) were significantly higher on average than managers ($M = .22$, $SD = .32$), $t(295) = 2.43$, $p < .05$, but not significantly different from laypeople ($M = .28$, $SD = .31$), $t(325) = 1.12$, $p = .26$, or hybrids ($M = .31$, $SD = .34$), $t(217) = 0.16$, $p = .87$. However, in terms of Kendall's tau-b correlations, which may be more appropriate given the small sample size for each participant (10 or fewer videos rated), creators ($M = .23$, $SD = .25$) were significantly higher on average than managers ($M = .15$, $SD = .25$), $t(295) = 2.96$, $p < .01$, and laypeople ($M = .18$, $SD = .23$), $t(325) = 2.96$, $p < .05$, but not significantly different from hybrids ($M = .21$, $SD = .26$), $t(217) = 0.55$, $p = .59$. These results provide supplementary support for H1.

It is possible that creators only have an advantage over managers in avoiding relatively small mistakes, and that the two roles are equal in avoiding big mistakes. To test this possibility, I scored big mistakes as overall accuracy scores more than one standard deviation below the mean ($M = 7.34$, $SD = 2.08$), which meant scores of five or less. I conducted a binary logistic regression model with big mistakes as the dependent variable, role as the independent variable, and number of videos scored as a control. Across all roles, big mistakes were made 17 percent of the time [$b = -1.57$, Wald $\chi^2 (1) = 507.34$, $p < .001$]. Managers were 56.46 percent more likely to

make big mistakes than creators [$b = 0.26$, Wald χ^2 (1) = 6.61, $p < .05$], and laypeople were

54.73 percent more likely to make big mistakes than creators, although this effect was marginal

[$b = 0.19$, Wald χ^2 (1) = 3.74, $p = .06$]. No other comparisons between roles were significant.

Consistent with H1, this suggests that creators had an advantage over managers in avoiding

severe false positives and false negatives, which are likely to have greater practical significance

than more minor mistakes.

I also tested whether creators had an advantage in identifying the single best idea from

the set they evaluated, using the inaccuracy scores for the top-ranked idea (according to the

audience) in each participant's set of ideas. As expected, creators ($M = -2.23$, $SD = 2.36$) ranked

the best idea significantly closer to first than managers ($M = -2.98$, $SD = 2.64$), $t(295) = 2.58$, $p <$

.05, and laypeople ($M = -3.18$, $SD = 2.49$), $t(325) = 3.55$, $p < .001$, but not significantly different

from hybrids ($M = -2.55$, $SD = 2.44$), $t(217) = 0.79$, $p = .43$. Thus, creators had an advantage

over managers in identifying the single best idea from the set they evaluated.

Comparing crowds of creators, managers, and laypeople. To help enrich and test the

robustness of the results, I conducted supplementary analyses on the crowd-level forecasts of

creators, managers, and laypeople. According to theory and research on the "wisdom of crowds"

(Clemen, 1989; Surowiecki, 2005), it is possible that creators no longer have an advantage over

managers and laypeople in placement accuracy when individuals' predictions are averaged

together (the results on estimation accuracy would be the same at the crowd level). For each of

the 100 videos in the forecasting survey, I calculated the crowd average on the three forecasting

items for creators, managers, and laypeople. To compare the overall placement accuracy of each

individual participant to each crowd, I used these crowd averages to formulate crowd rankings

for each subset of videos watched by individual participants. Overall accuracy scores were

calculated for each crowd (creator, manager, and layperson) in the same fashion as individual participants. Thus, the crowds were scored on the same subsets of videos that were randomly assigned to each individual (see Figure 6).

[Insert Figure 6 about here]

Paired-samples t-tests comparing the overall accuracy scores of individuals and crowds showed that all three crowds were significantly more accurate than individuals ($M = 7.56$, $SD = 2.08$), including crowds of creators ($M = 7.82$, $SD = 1.84$), $t(4668) = 13.43$, $p < .001$, managers ($M = 7.47$, $SD = 1.88$), $t(4668) = 3.67$, $p < .001$, and laypeople ($M = 7.83$ $SD = 1.79$), $t(4641) = 12.72$, $p < .001$. However, the manager crowd was significantly less accurate than the creator crowd, $t(4668) = 11.66$, $p < .001$, and the layperson crowd, $t(4641) = 9.81$, $p < .001$, while the creator and layperson crowds were approximately equal, $t(4642) = 0.06$, $p = .95$. Individual creators ($M = 7.54$, $SD = 1.98$) were significantly less accurate than the creator crowd ($M = 7.92$, $SD = 1.82$), $t(1614) = 6.89$, $p < .001$, and layperson crowd ($M = 7.95$, $SD = 1.73$), $t(1602) = 6.19$, $p < .001$, but approximately equal to the manager crowd ($M = 7.54$, $SD = 1.85$), $t(1614) = -0.06$, $p = .95$. Thus, an average creator was as accurate as the entire crowd of managers.

To test the possibility that the manager crowd (n = 120) was at a disadvantage because it contained fewer participants than the creator (n = 177) and layperson (n = 150) crowds, I calculated overall accuracy scores for four combined crowds (using the forecasts of all the participants in the combined crowd): creator-manager ($M = 7.73$, $SD = 1.80$), creator-layperson ($M = 7.99$, $SD = 1.75$), manager-layperson ($M = 7.86$, $SD = 1.78$), and creator-manager-layperson ($M = 7.90$, $SD = 1.74$). The crowd of just creators was significantly more accurate than the creator-manager crowd, $t(4668) = 4.46$, $p < .001$. Also, the manager-layperson crowd was not significantly more accurate than the crowd of just laypeople, $t(4641) = 1.88$, $p = .06$. Thus,

adding managers made the creator crowd worse and did not significantly improve the layperson crowd. The most accurate forecasts were from the creator-layperson crowd, which was significantly more accurate than the creator crowd, $t(4668) = 9.08$, $p < .001$, layperson crowd, $t(4641) = 7.29$, $p < .001$, and the creator-manager-layperson crowd, $t(4668) = 5.78$, $p < .001$.

I also scored the crowds in predicting the 1-100 rankings of the 100 videos included in the forecasting survey. Consistent with the other results, the creator crowd ($M = 77.76$, $SD = 18.48$) was significantly more accurate than the manager crowd ($M = 73.26$, $SD = 17.97$), $t(99) = 2.28$, $p < .05$, but the layperson crowd ($M = 77.41$, $SD = 16.47$) did not significantly differ from the creator or manager crowds. Taken together with the other results, these crowd-level results suggest that creators have an advantage over managers regardless of whether individuals or crowds are making the forecasts, but creators may only have an advantage over laypeople at the individual level—crowds of laypeople may be just as accurate as crowds of creators (though a crowd of both creators and laypeople may be most accurate). In general, these crowd-level results support the value of examining the effects that role designs have on creative forecasting accuracy at the individual level, as some roles may provide better raw material for crowd-level forecasts than others.