nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01794-4

Greater variability in judgements of the value of novel ideas

Received: 17 January 2023

Accepted: 22 November 2023

Published online: 12 January 2024

Check for updates

Wayne Johnson[1] & Devon Proudfoot[2]

Understanding the factors that hinder support for creative ideas is important because creative ideas fuel innovation—a goal prioritized across the arts, sciences and business. Here we document one obstacle faced by creative ideas: as ideas become more novel—that is, they depart more from existing norms and standards—disagreement grows about their potential value. Specifically, across multiple contexts, using both experimental methods (four studies, total $n = 1{,}801$) and analyses of archival data, we find that there is more variability in judgements of the value of more novel (versus less novel) ideas. We also find that people interpret greater variability in others' judgements about an idea's value as a signal of risk, reducing their willingness to invest in the idea. Our findings show that consensus about an idea's worth diminishes the newer it is, highlighting one reason creative ideas may fail to gain traction in the social world.

Society requires innovation to move forward. Innovation, in turn, requires new ideas. However, for new ideas to successfully drive innovation, people must support and invest in them[1–3]. In the current research, we examined a reason why new ideas may fail to gain momentum in the social world. As ideas become more novel—that is, they depart more from existing norms and standards—disagreement grows about their potential value. Specifically, using both experimental and archival data, we tested whether there is more variability in judgements of the value of more novel (versus less novel) ideas. Critically, we also tested whether people interpret greater variability in others' judgements about an idea's value as a signal of risk, reducing their willingness to invest in the idea.

Previous research has identified several psychological and contextual factors explaining why, despite society's purported desire for creativity, creative ideas, once generated, often become stalled[1,2,4–10]. While creative ideas are defined as those that are both novel and useful[11], research on the psychology of idea evaluation points to a bias against novelty as a factor contributing to why creative ideas generate resistance. For instance, even in innovation-focused contexts, evaluators have been shown to favour more familiar ideas over less familiar ones[12–14]. Research by others[7] suggests that evaluators may respond negatively to creative ideas because of an aversion to the uncertainty associated with novelty.

We proposed that, in addition to the already established tendency for novel ideas to often elicit negative reactions, evaluators may also have more varied judgements of the value of ideas as they become more novel, with this variability itself contributing to why novel ideas often fail to gain traction. We conceptualized value as a subjective assessment of an idea's worth within a given context[3,11,15] and novelty as the extent to which an idea is different from current templates and standards[16]. We limited the scope of our examination to situations in which idea evaluations were independent (that is, many independent judges rated how valuable a series of ideas varying in novelty were). This approach allowed us to isolate the impact of novelty on value judgement variability absent of potential peer pressure or contamination effects[17].

Previous studies point to the possibility that judgements of novel ideas might be particularly prone to variability between evaluators. First, there is evidence that creativity judgements may be inconsistent across different groups of judges[18–21]. For instance, a study[18] compared ratings of the creativity of artwork across professionals and student judges, finding that creativity ratings differed between these two groups. Second, there is evidence that groups often face challenges in reaching internal consensus about novel ideas[6,22,23]. For instance, a qualitative study[6] found that consensus pressures often lead groups to reject novel ideas rather than move forward with them. Although these studies did not directly examine whether judgements of value

[1]Management Department Eccles School of Business, University of Utah, Salt Lake City, UT, USA. [2]Human Resource Studies Department, ILR School, Cornell University, Ithaca, NY, USA. e-mail: devon.proudfoot@cornell.edu

varied across individual evaluators more as a function of idea novelty, these findings provide initial evidence that evaluators often disagree when assessing creative ideas.

We theorized that judgements of value might converge less as ideas become more novel because, for more novel ideas, evaluators have fewer common templates upon which to base their judgements. For less novel ideas—ideas relatively similar to existing templates—evaluators may more easily engage in a matching process, comparing the idea to familiar ideas that exist in the mind as schemas[24], We suggest that, as schemas tend to be similar among observers[25,26], there may be relative convergence in judgements of the value of ideas that appear similar to common templates. However, for more novel ideas, this matching process may become more difficult, as evaluators have fewer common templates to use to interpret the value of the focal idea[13,27]. In such situations, evaluators are more likely to be influenced by idiosyncratic knowledge, preferences and situational factors that cause evaluators' judgements to differ from each other, producing greater judgemental variability[28]. This line of reasoning is consistent with the notion that there is an inverse relationship between how novel an idea is and the extent to which one can predict its usefulness[29–32]. Our theorizing is also supported by studies showing that when assessing creative ideas, evaluators' judgements are often prone to errors and may be shaped by peripheral contextual factors and idiosyncratic individual differences, such as idea pitcher personality and evaluator construal level[3,5,33–39].

Notably, people often seek others' feedback about ideas before deciding whether to devote resources to developing them[2,3,40]. We propose that, independent of the content of the idea itself and, importantly, the average judged value of the idea by others, the degree of variability in other people's assessments of the potential value of an idea may itself be a cue that individuals use when deciding whether to invest in an idea. Specifically, we proposed that greater variability in evaluations may create an obstacle for creative ideas because people may interpret higher variability in others' assessments of an idea's value—as we predict for more novel ideas—as a sign that investing in the idea is risky, thus reducing people's willingness to support the idea[41].

Results

Overview

Five studies tested our predictions. Studies 1–4 examined the relationship between idea novelty and variability in value judgements across different contexts. Study 5 examined how variability in others' assessments of an idea's value affected observers' intent to invest in the idea. All analyses were conducted using SPSS. All effect sizes reported were calculated using SPSS, with the exception of the Cliff's delta effect sizes for study 5 which were generated using R v.4.3.1. All statistical tests reported are two-tailed tests. To capture variability (that is, variability between evaluators in judgements of idea value), we examined the standard deviation of value ratings for each idea evaluated[28]. We tested whether value standard deviation (referred to as $M_{s.d.}$) was larger on average for higher-novelty ideas compared to lower-novelty ideas. Hence, in our studies examining this prediction (studies 1–4), ideas were the unit of analysis. Across all studies, we did not find a positive relationship between mean value ratings and value standard deviation, indicating that our predicted effect is not reducible to scalar variability—the tendency for ratings on a response scale to become more varied as ratings increase in magnitude[42] (Supplementary Table 1).

Study 1: business venture pitches

In study 1, we examined whether there was more variability in value judgements for more novel (versus less novel) business venture pitches. We used descriptions of ventures pitched on the US television show *Shark Tank* (n = 1088). First, we conducted a pretest in which we asked a convenience sample of 1,927 US residents recruited from the online platforms Prolific Academic and Amazon's Mechanical Turk (Mturk) to rate the novelty of each pitch. For more details on the pretest, see Methods. Next, we asked a separate convenience sample of 1,000 US residents recruited from Mturk to rate the value offered by the 250 pitches rated most novel and the 250 pitches rated least novel in the pretest. For each pitch, participants were asked, 'What is the degree of value offered by this idea?' The data met the assumptions of homogeneity of variances and normality. As predicted, there was more variability in value ratings for the higher-novelty pitches ($M_{s.d.}$ = 1.62, s.d. = 0.27) compared to the lower-novelty pitches ($M_{s.d.}$ = 1.51, s.d. = 0.25), $t(498) = 4.46$, $P < 0.001$, $d = 0.40$, 95% confidence interval (CI) for difference [0.06, 0.15]. Also consistent with predictions, there was a positive association between pitch novelty score (generated in our pretest) and value judgement variability, $r(498) = 0.198$, $P < 0.001$, 95% CI [0.11, 0.28].

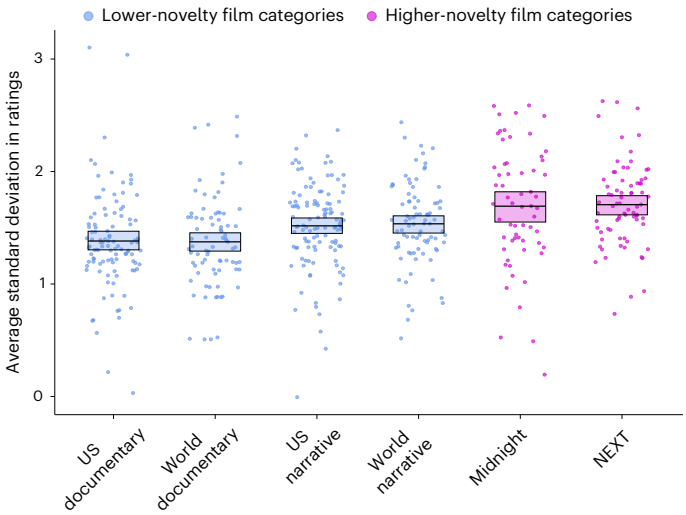

Fig. 1 | Average standard deviation in film ratings a function of film category (study 2). Films were the unit of analysis. The y axis represents the mean standard deviation in ratings for each film. n US documentary = 107, n World documentary = 80, n US narrative = 116, n World narrative = 85, n Midnight = 61, n NEXT = 74. Error bars represent bootstrapped 95% CI.

Study 2: films

In study 2, we examined archival audience evaluations of films (n = 523) premiering at Sundance Film Festival from 2015 to 2022. Sundance offered a useful context in which to test the predicted relationship between novelty and judgmental variability in a natural environment because films premiering at Sundance are organized into many categories, with certain films selected by festival programmers to premiere in categories specifically highlighting more novel projects. We used this feature of the festival's structure to examine whether there was greater variability in audience evaluations of films premiering in more novel categories at Sundance (that is, the festival categories NEXT and Midnight) compared to films premiering in less novel categories at Sundance (that is, the festival categories US documentary, World documentary, US narrative or World narrative). For complete details on how films are sorted into these categories by Sundance programmers and how we determined the relative novelty of each category, see Methods. Archival audience ratings for each film, which we used as a proxy for value, were captured using a single-item measure and were collected from https://cannes-ratings.tk/Sundance, which presents auto-aggregated web-based evaluations from audience members at Sundance, typically a mix of industry insiders, film makers, journalists and film fans[43]. Ratings were aggregated in real-time during the festival.

The data violated both normality and homoscedasticity assumptions. We therefore analysed the data using ordinary least squares (OLS) regression with Huber–White robust standard errors. As detailed in

https://doi.org/10.1038/s41562-023-01794-4

Table 1 | Linear regressions predicting ratings standard deviation (study 2)

	Model 1				Model 2				Model 3 (*n* ratings>19)			
NEXT (reference category)	*b*(s.e.)	*t*(517)	*P*	95% CI		*t*(516)	*P*	95% CI	*b*(s.e.)	*t*(319)	*P*	95% CI
US documentary	−0.323 (0.06)	−5.34	<0.001	[−0.44, −0.20]	−0.336 (0.06)	−5.51	<0.001	[−0.46, −0.22]	−0.359 (0.05)	−7.56	<0.001	[−0.45, −0.27]
World documentary	−0.330 (0.06)	−5.21	<0.001	[−0.45, −0.21]	−0.350 (0.06)	−5.47	<0.001	[−0.48, −0.22]	-0.318 (0.05)	−5.92	<0.001	[−0.42, −0.21]
US narrative	−0.188 (0.06)	−3.33	0.001	[−0.30, −0.08]	−0.167 (0.06)	−3.00	0.003	[−0.28, −0.06]	−0.202 (0.05)	−4.11	<0.001	[−0.30, −0.11]
World narrative	−0.170 (0.06)	−2.87	0.004	[−0.29, −0.05]	−0.174 (0.06)	−2.93	0.004	[−0.29, −0.06]	−0.143 (0.06)	−2.59	0.010	[−0.25, −0.03]
Midnight	−0.014 (0.08)	−0.17	0.867	[−0.17, 0.15]	−0.005 (0.08)	−0.06	0.949	[−0.16, 0.15]	−0.030 (0.08)	−0.38	0.706	[−0.19, −0.13]
Number of ratings					−0.15[a] (0.00)	−4.29	<0.001					
Midnight (reference category)												
US documentary	−0.310 (0.08)	−3.84	<0.001	[−0.47, −0.15]	−0.330 (0.08)	−4.12	<0.001	[−0.49, −0.17]	−0.328 (0.08)	−4.27	<0.001	[−0.48, −0.18]
World documentary	−0.317 (0.08)	−3.83	<0.001	[−0.48, −0.15]	−0.345 (0.08)	−4.17	<0.001	[−0.51, −0.18]	−0.288 (0.08)	−3.56	<0.001	[−0.45, −0.13]
US narrative	−0.175 (0.08)	−2.25	0.025	[−0.33, −0.02]	−0.162 (0.08)	−2.13	0.033	[−0.31, −0.01]	−0.172 (0.08)	−2.21	0.028	[−0.33, −0.02]
World narrative	−0.157 (0.08)	−1.97	0.050	[−0.31, 0.00]	−0.169 (0.08)	−2.14	0.033	[−0.32, −0.01]	−0.113 (0.08)	−1.38	0.169	[−0.27, 0.05]
NEXT	0.014 (0.08)	0.17	0.867	[−0.15, 0.17]	0.005 (0.08)	0.64	0.949	[−0.15, 0.16]	0.030 (0.08)	0.38	0.706	[−0.13, 0.19]
Number of ratings					−0.152[a] (0.00)	−4.29						
R^2	0.084				0.106				0.144			

[a]Standardized regression coefficient displayed. Unstandardized regression coefficients are displayed unless otherwise noted with Huber–White robust standard errors in parentheses.

Fig. 1 and Table 1 model 1, consistent with predictions, there was more variability in archival audience ratings of films within NEXT, a category at Sundance featuring higher-novelty films, compared to films in each of the lower-novelty categories (US documentary, World documentary, US narrative and World narrative). There was also more variability in archival audience ratings of films within Midnight, a category at Sundance featuring higher-novelty films, compared to films in each of the lower-novelty categories (US documentary, World documentary, US narrative and World narrative), although the comparison between Midnight and World narrative did not reach statistical significance ($P = 0.050$). There were similar levels of variability in ratings of films within NEXT and Midnight, the two higher-novelty categories. Results were robust after controlling for number of ratings and when analysing only films with at least 20 ratings (Table 1 models 2 and 3), although the comparison between Midnight and World narrative was non-significant ($P = 0.169$) when analysing only films with at least 20 ratings.

Study 3: experimental evidence

Studies 1 and 2 showed our predicted effect in evaluations of real-world ideas but did not control for covariation between idea novelty and other factors that could impact variability in value judgements. Study 3 addressed this limitation by experimentally manipulating idea novelty, providing a causal test of the influence of novelty on value judgement variability. We explored evaluations of abstract art, predicting more variability in judgements of the value of a set of paintings when they were construed as more novel within a given context than when identical paintings were construed as less novel within a given context.

We recruited 200 US residents from the online platform Lucid Theorem and asked them to imagine a particular context—an alien society on a faraway planet. Participants read that they would assess the subjective value of a series of paintings specifically within the context of this alien society. Participants read that in this context, a particular style of abstract art was highly acclaimed by critics. To manipulate novelty within this context, participants were first shown a prototype of the style of painting that was highly acclaimed and were then asked to judge the value of a subset of 80 additional paintings within the context of the alien society (that is, 'In this alien society, how much potential value does the artwork above have?'). Half of the additional paintings shown to each participant were by the same artist as the prototype. These paintings were of a similar style to the prototype and therefore lower novelty within the context. The other half of the paintings shown to each participant were by a second artist. These paintings were of a different style than the prototype and therefore higher novelty within the context.

We also varied, across participants, which of the two artists' paintings was used as the prototype painting. As such, while the exact same paintings were rated by all participants, half of which were by one artist (artist A) and half of which were by another artist (artist B), we experimentally varied whether a given painting was high versus low novelty within the alien society by randomly assigning participants to view only one of the two artists' work as the prototype (either a painting by artist A or a painting by artist B). By asking participants to each rate paintings by both artist A and artist B, regardless of which artist's work they saw as the prototype, we were able to examine whether there was more variability in judgements of identical paintings when they were construed as higher novelty within a given context compared to when they were construed as lower novelty within a given context.

The data met the assumptions of homogeneity of variances and normality. As predicted, there was more variability in value judgements for the high novelty paintings ($n = 80$ painting stimuli; $M_{s.d.} = 2.07$, s.d. = 0.22) than the low novelty paintings ($n = 80$ painting stimuli; $M_{s.d.} = 1.79$, s.d. = 0.26), $t(158) = 7.45$, $P < 0.001$, $d = 1.18$, 95% CI for difference [0.21, 0.36]. This effect was robust in the condition in which artist A's work was the prototype ($n = 40$ painting stimuli), $F(1, 156) = 17.22$, $P < 0.001$, η_p^2 (partial eta squared) = 0.01, 95% CI for difference [0.11, 0.30] and in the condition in which artist B's work was the prototype ($n = 40$ painting stimuli), $F(1, 156) = 57.06$, $P < 0.001$, $\eta_p^2 = 0.27$, 95% CI for difference [0.27, 0.46]), with a larger effect in the condition in which artist B's work was the prototype, $F(1, 156) = 5.80$, $P = 0.017$, $\eta_p^2 = 0.04$.

Study 4: mechanistic evidence

Studies 1–3 showed that, as ideas become more novel, there is more variability in how evaluators judge their value. Study 4 examined a boundary condition: whether the positive relationship between novelty

https://doi.org/10.1038/s41562-023-01794-4

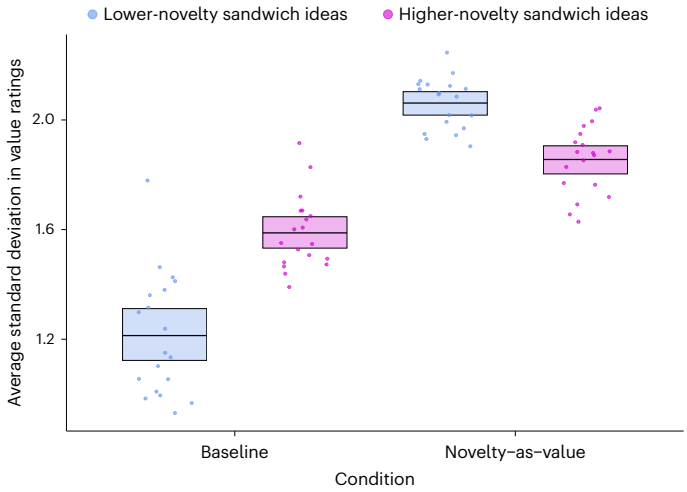

Fig. 2 | Value standard deviation as a function of condition (study 4). Sandwich ideas were the unit of analysis. The *y* axis represents the mean standard deviation in ratings for each sandwich idea. *n* baseline condition, lower-novelty sandwich ideas = 19; *n* baseline condition, higher-novelty sandwich ideas = 19; *n* novelty-as-value condition, lower-novelty sandwich ideas = 19; *n* novelty-as-value condition, higher-novelty sandwich ideas = 19. Error bars represent bootstrapped 95% CI.

and variability might diminish when the value of an idea was closely tied to it how novel it was. Explicitly construing value as dependent on novelty might provide observers with clear criteria on which to judge the value of unfamiliar ideas, reducing variability in value judgements for higher-novelty ideas. Construing value in terms of novelty might also limit the applicability of existing templates for assessing idea value, as the evaluative context is unfamiliar, rendering value judgements more difficult and susceptible to idiosyncratic variability regardless of idea novelty[28]. Both lines of reasoning led us to expect that the effect observed thus far—greater variability in value judgements of more novel ideas—would be attenuated when value was explicitly defined in terms of novelty.

We examined this prediction in the context of ideas for sandwiches. First, we conducted a pretest in which we recruited 101 US residents from Prolific with experience working in the hospitality industry, providing a sample with some domain knowledge. We asked participants to rate the novelty of 40 ideas for sandwiches. We then selected the 19 sandwich ideas that were judged to be higher novelty (significantly higher than the scale midpoint) and the 19 sandwich ideas that were judged to be lower novelty (significantly lower than the scale midpoint) to use in our main study. For pretest details, see Methods.

Next, we recruited a separate sample of 200 US residents with experience working in the hospitality industry to rate the value offered by the 38 sandwich ideas identified in the pretest, half of which were higher novelty and half of which were lower novelty. Participants were randomly assigned to either a baseline condition or a novelty-as-value condition in which we defined value as dependent on novelty. Specifically, in the baseline condition, for each sandwich idea, participants were asked, 'How successful would this sandwich be as a menu item?' In the novelty-as-value condition, for each sandwich idea, participants were asked, 'At a restaurant specializing in sandwiches no one has tried before, how successful would this sandwich be as a menu item?'.

In our preregistered analysis plan, we indicated that we would analyse the data using analysis of variance to test for an interaction between condition (baseline versus novelty-as-value) × idea novelty (low versus high) on average standard deviation in value ratings, followed by mean comparisons to test our predictions. However, the data met the normality assumption but did not meet the assumption of equality of variances, so we analysed the data using OLS regression with Huber–White robust standard errors to account for heteroscedasticity. We report all results using our original, preregistered analysis plan in the Supplementary Information, which yielded results in line with our predictions.

Consistent with our preregistered predictions, there was an idea novelty (low versus high) × condition (baseline versus novelty-as-value) interaction, $b = -0.58$, s.e. = 0.07, $t(72) = -8.13$, $P < 0.001$, 95% CI [−0.72, −0.44] (Fig. 2). In the baseline condition ($n = 38$ ideas), there was more variability in value judgements for the higher-novelty ideas ($n = 19$ ideas; $M_{s.d.} = 1.59$, s.d. = 0.13) compared to the lower-novelty ideas ($n = 19$ ideas), $b = 0.37$, s.e. = 0.06, $t(72) = 6.12$, $P < 0.001$, 95% CI [0.25, 0.50]. This effect reversed in the novelty-as-value condition ($n = 38$ ideas), with the lower-novelty ideas showing greater value variability ($n = 19$ ideas) than the higher-novelty ideas ($n = 19$ ideas), $b = -0.21$, s.e. = 0.04, $t(72) = -5.63$, $P < 0.001$, 95% CI [−0.28, −0.13]. Although not preregistered, we also observed that variability was greater in the novelty-as-value condition than in the baseline condition for both the low novelty ideas, $b = 0.85$, s.e. = 0.06, $t(72) = 14.90$, $P < 0.001$, 95% CI [0.73, 0.96] and the high novelty ideas, $b = 0.27$, s.e. = 0.04, $t(72) = 6.22$, $P < 0.001$, 95% CI [0.18, 0.35].

We found consistent results when we tested for an interaction between idea novelty score (generated in our pretest) and condition (baseline versus novelty as value) on value standard deviation (Supplementary Information). Our predictions were also supported when we analysed only responses from participants who indicated they had experience working in restaurants specifically ($n = 142$ participants or 71.0% of the sample, Supplementary Information).

Study 5: impact on intent to invest

Studies 1–4 demonstrated more variability in judgements of the value of higher-novelty (versus lower-novelty) ideas. In study 5, we examined one consequence of this effect. We theorized that more variability in value judgements for higher-novelty ideas might present a barrier for these ideas because variability might signal risk, diminishing evaluators' willingness to support newer ideas.

We recruited 401 US residents with investment experience from Prolific. Participants were asked to consider whether they would invest in the following product idea: 'A device that directs sound from tablet/laptop speakers to the user, works by funneling the sound from the speakers back in the direction of the user.' Participants were then asked to imagine they had consulted experts, colleagues and others they trusted for advice, asking them for feedback on how valuable they thought the investment opportunity was. Participants then read that each person had given the idea a value rating ranging from 1 star (worst rating) to 5 stars (best rating). Participants were then randomly assigned to one of two conditions (high variability versus low variability). Participants in the high-variability condition saw a graph showing a higher variability of ratings; participants in the low-variability condition saw a graph showing a lower variability of ratings (Fig. 3). The mean rating across conditions was identical (3.0). Participants then completed three items capturing their intent to invest in the idea. Next, participants rated, in a randomized order, how risky the idea was and how novel the idea was. For details on exact items used, see Methods. Finally, as a manipulation check, participants indicated how similar the idea ratings were.

In our preregistered analysis plan, we indicated that we would analyse the data using independent samples *t*-tests to examine the effect of condition (low variability versus high variability) on our dependent measures. However, the data met the equality of variances assumption but did not meet the assumption of normality, so we analysed the data using non-parametric Mann–Whitney *U*-tests instead of *t*-tests. We report all results using our original, preregistered analysis plan in the Supplementary Information, which yielded results in line with our predictions.

https://doi.org/10.1038/s41562-023-01794-4

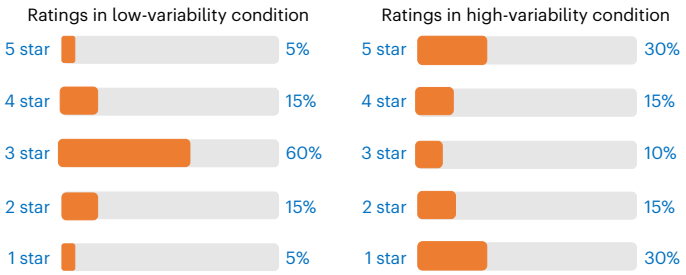

Fig. 3 | Manipulation stimuli (study 5). Participants were randomly assigned to high-variability or low-variability conditions. Those in the high-variability condition show a higher variability of ratings, while those in the low-variability condition show a lower variability of ratings.

Our manipulation check confirmed that perceived similarity of ratings was higher in the low-variability condition (n = 200 participants) (M = 4.29, s.d. = 1.34) than in the high-variability condition (n = 201 participants) (M = 3.97, s.d. = 1.49), $z(1{,}398) = -2.29$, $P = 0.022$, Cliff's delta = 0.13, 95% CI for the delta estimate = [0.02, 0.24]. Consistent with predictions, participants in the high-variability condition indicated lower intent to invest in the idea (M = 2.51, s.d. = 1.01) compared to participants in the low-variability condition (M = 2.96, s.d. = 1.13), $z(1{,}399) = -4.13$, $P < 0.001$, Cliff's delta = 0.24, 95% CI for the delta estimate = [0.13, 0.34]). Participants also judged the investment to be riskier in the high-variability condition (M = 5.61, s.d. = 1.09) compared to the low-variability condition (M = 5.24, s.d. = 1.15), $z(1{,}399) = -3.52$, $P < 0.001$, Cliff's delta = −0.20, 95% CI for the delta estimate = [−0.31, −0.09]. The effect of condition on ratings of idea novelty was non-significant, $M_{high} = 3.98$, s.d. = 1.51, $M_{low} = 3.78$, s.d. = 1.46), $z(1{,}399) = -1.37$, $P = 0.169$, Cliff's delta = 0.08, 95% CI for the delta estimate = [−0.19, 0.03], indicating that our manipulation of value judgement variability did not significantly affect perceptions of how novel the idea was. There was an indirect effect of condition on intent to invest through perceived risk, $b = -0.12$, s.e. = 0.05, 95% CI [−0.23, −0.04], indicating that one reason that greater variability in judgements of an idea's value leads to lower intent to invest in the idea is that greater variability signals greater risk (Fig. 4).

Discussion

Innovation is a goal prioritized across the arts, sciences and business[3,12,14]. Despite this general desire for novel products and solutions, our research suggests that when evaluators actually encounter novel ideas—the very ideas society hopes to advance—these ideas may tend to generate disagreement. Specifically, we showed that judgements of the value of highly novel ideas are more varied across evaluators than judgements of less novel ideas. We found this effect in multiple contexts, using archival and experimental methods and in convenience samples and samples of evaluators with domain knowledge. Furthermore, we found that people interpret variability in value judgements as a negative signal, reducing their support for novel ideas. Our results, particularly those of study 4, suggest that evaluators' judgements of novel ideas may vary because evaluators have relatively few common templates against which to evaluate them, making judgements more reliant on idiosyncratic knowledge and preferences. Broadly, our findings are practically important because they highlight that the way human minds, in aggregate, process novelty may be at odds with society's goal to innovate. Our findings also offer practical insight into why creative ideas often generate conflict and may seem socially controversial[8,23,44–46].

Our research contributes to the literature on evaluation of creative ideas. Previous studies point to a bias against novel ideas or a tendency for evaluators to judge highly novel ideas more negatively on average than less novel ones[7,8]. Our findings build on this work by showing that more novel ideas may not only elicit negative reactions but may also produce greater variability in evaluators' reactions, advancing our understanding of the psychological processes at play in creative idea evaluation. In study 5, we also found that greater variability in value judgements (compared to less variability) reduces people's intent to invest in ideas, even when average judged value was held constant. This result points to a second-order effect (that is, people's judgements about variability in others' judgements) that may operate as an independent mechanism, alongside a general negativity toward novel ideas, explaining why creative ideas often lose momentum in the social world[4–6].

Our findings advance the perspective within the creativity literature that assessments of creative ideas may be inconsistent across different sets of judges. Previous studies demonstrated that ratings of the 'creativity' of ideas may vary between groups (for instance, creator ratings versus peer ratings), calling into question the reliability of creativity judgements[18,21,47,48]. Our research, particularly study 3's evidence for the causal role of novelty in impacting variability in value judgements, builds on these findings and offers evidence that judgements may become more inconsistent between individual evaluators as ideas become more creative. Our findings are notable given that we identified our predicted effect in domains often examined in creativity research, such as entrepreneurial pitches, films and visual art[18,19,49–51]. Our findings are also broadly consistent with previous theoretical work in the creativity literature questioning whether judgements of the 'value' of creative ideas might be prone to error and may vary depending on context[52].

Our research also has implications for decision-making more broadly by furthering our understanding of the factors that drive consensus, or its inverse, judgmental variability[28,53]. Recent work has drawn attention to how irrelevant individual and contextual factors, such as a person's mood or the weather, produce unwanted inconsistency across decision-makers[28]. We build on this work by showing that evaluators' judgements may be especially prone to variability when the stimuli being evaluated are new. Our studies demonstrate that consensus is a feature of the familiar—novelty itself produces judgmental variability.

Our research has several notable limitations. Our studies captured idea value using subjective ratings on a Likert-type scale. Future research is needed to test whether our effects extend to other measures of value, such as numerical assessments of an idea's potential value in terms of revenue to an organization over a certain period of time. Most of our studies also captured idea novelty using subjective assessments. This approach was aligned with our theoretical account, as we hypothesized that ideas that subjectively seem more novel should also produce more value judgement variability. Study 3's results also showed that our effects extended to a context in which we objectively manipulated novelty (by varying the extent to which a focal stimulus was similar to/different from a prototype), providing some evidence that our effects are generalizable to situations beyond those in which novelty is subjectively assessed. Still, more research is needed to extend our findings using objective measures of novelty[44,54,55].

The evidence we provide of our mechanistic account is also limited. We theorized that a lack of common templates could explain greater variability in judgements of idea value as ideas become more novel. The empirical evidence we present offers some evidence consistent with this account, showing that our predicted effect is diminished when we offer evaluators a common template (idea novelty) on which to assess idea value (study 4). However, more research is needed to rule in our theoretical mechanism. For instance, future studies could directly measure whether evaluators increasingly call to mind different templates and different evaluative criteria as ideas become more novel.

Several potential boundary conditions could also be explored in future research. Our studies examined judgements of idea value made by convenience samples and samples of evaluators with domain knowledge (that is, study 2 which sampled audience members at Sundance and study 4 which sampled evaluators with experience working in

https://doi.org/10.1038/s41562-023-01794-4

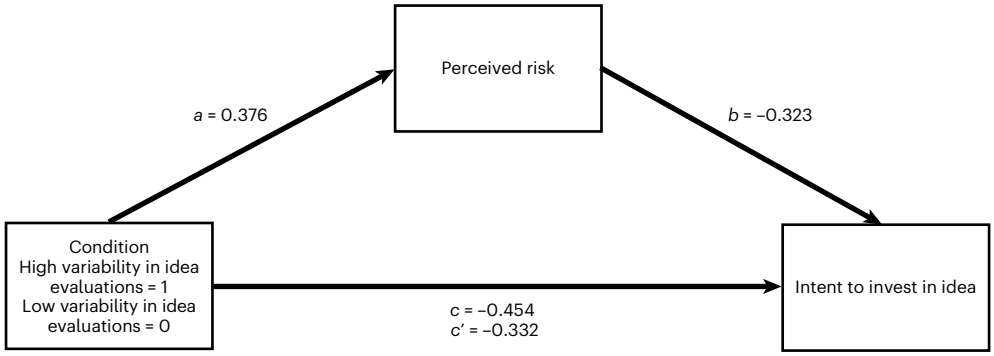

Fig. 4 | Mediational model (study 5). We used the PROCESS macro for SPSS to conduct a simple mediation model to test for the indirect effect of condition (low variability in idea evaluations versus high variability in idea evaluations) on intent to invest in the idea through perceived risk. *n* low variability in idea evaluations = 200; *n* high variability in idea evaluations = 201. Regression coefficients shown are unstandardized. All tests are two-sided. Adjustments were not made for multiple comparisons. *a* estimates the effect of condition on perceived risk: coefficient = 0.376, s.e. = 0.11, $t(399) = 3.35$, $P = 0.001$, 95% CI [0.16, 0.60]; *b* estimates the effect of perceived risk on intent to invest, holding condition constant: coefficient = −0.323, s.e. = 0.05, $t(399) = -7.15$, $P < 0.001$, 95% CI [−0.41, −0.23]; *c* estimates the total effect of condition on intent to invest: coefficient = −0.454, s.e. = 0.11, $t(399) = -4.23$, $P < 0.001$, 95% CI [−0.66, −0.24]; *c′* estimates the effect of condition on intent to invest holding perceived risk constant: coefficient = −0.332, s.e. = 0.10, $t(399) = -3.24$, $P = 0.001$, 95% CI [−0.53, −0.13].

the hospitality industry). Our results thus suggest that our predicted effects extend to samples of evaluators with some degree of expertise. It is possible, however, that we could observe relative consensus in judgements of the value of novel ideas in samples of experts, as experts may be more likely to share the same evaluative criteria[21,48,56]. The consensual assessment technique[16], a commonly used method for scoring the creativity of ideas, in which creativity scores are aggregated across judges, is recommended to be used specifically with experts for this reason. The possibility of moderation of our predicted effects by evaluator expertise is consistent with the results of study 4, in which we found that providing judges with shared evaluative criteria diminished our effect.

Relatedly, study 4's results suggest that in domains in which evaluators view the novelty of a particular idea as synonymous with its value, or value is defined in a highly specific way, the positive relationship between novelty and variability may diminish. Our studies also only examined contexts in which evaluators' value judgements were independent. The novelty–variability relationship may be less likely to emerge when evaluators are aware of others' assessments, as anchoring effects and conformity pressures may drive consensus in judgements[57]. There may also be a curvilinear effect of novelty on variability. While the empirical scope of our studies did not provide enough exceptionally novel ideas to test this possibility, future studies could examine whether, at very high levels of novelty, opinions of value might converge again, as has been argued for assessments of 'unicorn' ideas in the technology sector[58].

Future research could also explore additional consequences of variability in judgements of novel ideas, beyond investment intentions. For instance, future studies could explore whether people who seek others' opinions about their own ideas are less likely to pursue creative ideas because of variability in feedback[2,40]. We also specifically examined the consequences of variability when the distribution of value ratings was centred at the midpoint on the scale. Future research could test how varying levels of variability in different distributions (positive versus negative skewed distributions) impacts evaluators' support for ideas.

Conclusion

New ideas are an important source of innovation. This research identifies one reason why the implementation of creative ideas might become stalled—evaluators disagree more about the value of more novel ideas than less novel ideas. Observing less consensus (that is, more variability) in others' judgements of the value of an idea can subsequently diminish people's interest in supporting creative ideas, as disagreement is seen as a sign of risk. Our findings suggest that, for those seeking new ideas, a certain amount of disagreement about the value of a creative idea may be expected, as we find that disagreement is a byproduct of novelty.

Methods

All studies received ethics approval from the institutional review board at Cornell University (protocol nos. 2102010124 and 1905008829) and comply with all relevant ethical regulations. For all studies except for study 2 (analysis of archival data), study design, sample size, predictions, exclusion criteria and analysis plans were preregistered at AsPredicted.org. Studies 1, 3, 4 and 5 were preregistered on 6 April 2022, 14 August 2022, 29 December 2021 and 20 April 2022, respectively. The data for studies 1, 3, 4 and 5 were collected using Qualtrics. Informed consent was obtained from all study participants in studies 1, 3, 4 and 5. Participants in studies 1, 4 and 5 were compensated for their time with a flat fee (study 1, US$0.20; study 4, US$0.80; study 5, US$0.45). For study 3 participants, the researchers paid Lucid Theorem US$1.00 per participant. Regarding participant compensation, Lucid Theorem's website provides the following information, 'Our respondents are sourced from a variety of supplier types who have control over incentivizing their respondents based on their business rules.... Some suppliers do not incentivize their respondents at all, most provide loyalty reward points or gift cards and some provide cash payments'[59].

For all studies using the same sampling source, participants who took part in one study were excluded from all subsequent studies. Data collection and analysis were not performed blind to the conditions of the experiments. For all studies, anonymized data, code and materials, including all idea stimuli used, are available on the Open Science Framework (OSF) at https://osf.io/h3puf/.

Study 1

Pretest. We gathered descriptions of the 1,088 ventures pitched on seasons 1–12 of *Shark Tank* from https://allsharktankproducts.com (all seasons available at the time of the study). We recruited a sample of 1,927 US residents from Prolific Academic and Mturk ($M_{age} = 38.41$, s.d. = 28.69; 51.9% women, 46.7% men, 1.5% other gender identity) to rate the novelty of each pitch on a 7-point scale (1 = not at all novel, 7 = extremely novel). Novelty was defined as the degree to which the pitch was unusual, unique or unfamiliar[16]. Each participant rated a random subset of 10–20 pitches, providing approximately 30 ratings per

pitch. We selected the 250 pitches rated most novel and the 250 pitches rated least novel to use in our main study. For exact descriptions of all pitches used in the main study, see the OSF page. A sensitivity analysis using G*Power[60] indicated that, at $\alpha = 0.05$, our sample size of 500 pitches provided 80% power to detect a minimum effect of $d = 0.25$.

Main study. The study preregistration is available at https://aspredicted.org/blind.php?x=76Z_274. The study was launched on 21 December 2022. We recruited a sample of 1,000 US residents from Mturk ($M_{age} = 33.25$, s.d. = 9.79; 49% women, 50% men, 1% other gender identity). No participants were excluded from analyses. Participants rated the value of a random subset of 15 of the 500 pitches identified in the pretest, providing approximately 30 ratings per pitch. For each idea, participants were asked, 'What is the degree of value offered by this idea?' rated on a 7-point scale (1 = extremely low value, 7 = extremely high value).

Study 2

Film novelty (independent variable). Film makers can submit their feature-length films for consideration in one of four categories at the Sundance film festival: US documentary, World documentary, US narrative or World narrative. Films considered for these categories are also considered by festival programmers for two additional categories: NEXT and Midnight. NEXT features, 'Pure, bold works distinguished by an innovative approach to storytelling,' while Midnight features, 'An eclectic mix of horror, sci-fi, over-the-top comedy, explicit animation and bizarre stories that defy categorization'[61]. Hence, all feature-length films submitted to Sundance are either rejected or are sorted by programmers into one of two groups: (1) films accepted into the original category submitted to (US documentary, World documentary, US narrative, World narrative), which we labelled lower-novelty categories or (2) films accepted into NEXT or Midnight, which we labelled the higher-novelty categories. We confirmed through a former programmer at Sundance who is a personal contact of one of the authors that this characterization of the relative novelty of films in the six feature film categories is accurate.

Audience ratings standard deviation (outcome measure). We used Sundance archival audience ratings of the films as a proxy for value, which we collected from the website https://cannes-ratings.tk/Sundance on 16 August 2022. This website automatically standardized each film rating collected into a ten-point scale and provided the standard deviation of ratings for each film, which we used as our dependent measure. Of the 539 films listed on a website, 16 films had one rating; thus a standard deviation could not be computed, leaving 523 films in our analyses. As this website was no longer active at the time of publication of this article, screenshots of the website are available at: https://osf.io/h3puf/. In addition, the source data used on the website is available at: https://github.com/rurban/cannes-ratings/blob/master/lib/Sundance/rurban.pm.

Control variable. Number of ratings varied across films. Given that number of ratings impacts ratings standard deviation, we included numbers of ratings as a control in our analyses.

Study 3

The study preregistration is available at https://aspredicted.org/blind.php?x=89W_9GJ.

The study was launched on 14 August 2022. We recruited 200 US residents from Lucid Theorem, an online platform that provides nationally representative samples of US residents ($M_{age} = 44.42$, s.d. = 16.78; 51% women, 49% men)[62]. No participants were excluded from analyses.

Participants were randomly assigned to one of two conditions. Participants in one condition saw a painting by artist A (J. Albers) as the example of highly acclaimed art within the alien society. Participants in the other condition saw a painting by artist B (F. Winter) as the example of highly acclaimed art within the alien society. Participants then judged the value of a random subset of 20 paintings from a group of 80 paintings, half by artist A and half by artist B. See OSF page for all painting stimuli used.

For each painting evaluated, participants were asked, 'In this alien society, how much potential value does the artwork above have?' rated on a 7-point scale (1 = extremely low value, 7 = extremely high value). For participants who viewed artist A's work as the prototype, ratings of artist A's paintings were coded as low novelty and ratings of artist B's paintings were coded as high novelty. For participants who viewed artist B's work as the prototype, ratings of artist B's paintings were coded as low novelty and ratings of artist A's paintings were coded as high novelty. Across conditions, each of the 80 paintings were shown as a low novelty painting and as a high novelty painting; thus, our sample size consisted of 160 painting stimuli. A sensitivity analysis using G*Power indicated that, at $\alpha = 0.05$, this sample size provided 80% power to detect a minimum effect of $d = 0.45$.

Study 4

Pretest. We recruited 101 US residents from Prolific ($M_{age} = 32.22$, s.d. = 10.38; 50% women, 48% men, 2% other). Using study 1's novelty measure, each participant rated the novelty of a random subset of 10 sandwich ideas from a list of 40 sandwich ideas, providing 25 ratings per idea. We selected 19 ideas that were significantly below the scale midpoint (lower-novelty ideas) and 19 ideas that were significantly above the scale midpoint (higher-novelty ideas) to use in our main study.

Main study. The study preregistration is available at https://aspredicted.org/blind.php?x=X84_T8Y. The study was launched on 29 December 2021. We recruited 200 US residents with experience working in the hospitality industry on Prolific (48% women, 51% men, $M_{age} = 36.01$, s.d.$_{age} = 12.22$). No participants were excluded from analyses. Participants were randomly assigned to either a baseline condition or a novelty-as-value condition in which we defined value as dependent on novelty. In both conditions, participants each rated all 38 sandwich ideas identified in the pretest. Participants in the baseline condition were asked, 'How successful would this sandwich be as a menu item?' rated on a 7-point scale (1 = not successful at all to 7 = extremely successful). Participants in the novelty-as-value condition were asked, 'At a restaurant specializing in sandwiches no one has tried before, how successful would this sandwich be as a menu item?' rated on the same 7-point scale. At the end of the survey, participants indicated whether they had experience working in restaurants specifically. Each of the 38 sandwich ideas were rated both in the baseline condition and in the novelty-as-value condition. A sensitivity analysis using G*Power indicated that, at $\alpha = 0.05$, our sample size of 76 ideas provided 80% power to detect a minimum effect of $\eta_p^2 = 0.10$.

Study 5

The study preregistration is available at https://aspredicted.org/blind.php?x=WJ5_5KL.

The study was launched on 21 April 2022. We recruited 401 US residents with investment experience from Prolific ($M_{age} = 42.89$, s.d. = 16.78; 50% women, 50% men; years of investment experience: $M = 9.30$, s.d. = 9.35). No participants were excluded from analyses. A sensitivity analysis using G*Power indicated that, at $\alpha = 0.05$, our sample size of 401 participants provided 80% power to detect a minimum effect of $d = 0.28$.

Participants were asked to consider whether they would invest in the following product idea: 'A device that directs sound from tablet/laptop speakers to the user, works by funneling the sound from the speakers back in the direction of the user'. Participants were then asked to imagine they had consulted experts, colleagues and others they trusted for advice, asking them for feedback on how valuable they thought the

investment opportunity was. Participants then read that each person had given the idea a value rating ranging from 1 star (worst rating) to 5 stars (best rating). Participants were then randomly assigned to one of two conditions (high variability versus low variability). Participants in the high-variability condition saw a graph showing a higher variability of ratings; participants in the low-variability condition saw a graph showing a lower variability of ratings (Fig. 3). The mean rating across conditions was identical (3.0).

Participants completed three items capturing intent to invest: (1) 'I intend to invest in this company', (2) 'Investing in this company will be profitable for me' and (3) 'Investing in this company is a beneficial decision for me', rated on a 7-point scale (1 = strongly disagree, 7 = strongly agree) ($\alpha = 0.92$). Next, participants rated, in a randomized order, how risky and how novel the idea was. Three items captured perceived risk: (1) 'This idea is risky', (2) 'There is a risk the idea will fail', (3) 'There is a risk this idea will not succeed' rated on a 7-point scale (1 = strongly disagree, 7 = strongly agree) ($\alpha = 0.89$). Three items captured perceived novelty: (1) 'This idea is unusual', (2) 'This idea is novel', (3) 'This idea is unique', rated on a 7-point scale (1 = strongly disagree, 7 = strongly agree) ($\alpha = 0.93$). Finally, as a manipulation check, participants indicated how similar the ratings of the idea were on a 7-point scale (1 = the assessments of the idea were very different, 7 = the assessments of the idea were very similar).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

De-identified participant data for all studies are permanently and publicly available on the Open Science Framework at https://osf.io/h3puf/. Source data are provided with this paper.

Code availability

The code to replicate the analyses in the manuscript and Supplementary Information is available permanently and publicly on the Open Science Framework at https://osf.io/h3puf/.

References

1. Baer, M. Putting creativity to work: the implementation of creative ideas in organizations. *Acad. Manag. J.* **55**, 1102–1119 (2012).
2. Perry-Smith, J. E. & Mannucci, P. V. From creativity to innovation: the social network drivers of the four phases of the idea journey. *Acad. Manag. Rev.* **42**, 53–79 (2017).
3. Berg, J. M. Balancing on the creative highwire: forecasting the success of novel ideas in organizations. *Adm. Sci. Q.* **61**, 433–468 (2016).
4. Staw B. M. in *Creative Action in Organizations: Ivory Tower Vision and Real World Voices* (eds Ford, C. M. & Gioia, D. A.) 161–166 (Sage, 1995).
5. Zhou, J., Wang, X. M., Bavato, D., Tasselli, S. & Wu, J. Understanding the receiving side of creativity: a multidisciplinary review and implications for management research. *J. Manag.* **45**, 2570–2595 (2019).
6. Harvey, S. & Mueller, J. S. Staying alive: toward a diverging consensus model of overcoming a bias against novelty in groups. *Organ. Sci.* **32**, 293–314 (2021).
7. Mueller, J. S., Melwani, S. & Goncalo, J. A. The bias against creativity: why people desire but reject creative ideas. *Psychol. Sci.* **23**, 13–17 (2012).
8. Mueller, J., Melwani, S., Loewenstein, J. & Deal, J. J. Reframing the decision-makers' dilemma: towards a social context model of creative idea recognition. *Acad. Manag. J.* **61**, 94–110 (2018).
9. Lu, S., Bartol, K. M., Venkataramani, V., Zheng, X. & Liu, X. Pitching novel ideas to the boss: the interactive effects of employees' idea enactment and influence tactics on creativity assessment and implementation. *Acad. Manag. J.* **62**, 579–606 (2019).
10. Acar, O. A., Tarakci, M. & Van Knippenberg, D. Creativity and innovation under constraints: a cross-disciplinary integrative review. *J. Manag.* **45**, 96–121 (2019).
11. Runco, M. A. & Jaeger, G. J. The standard definition of creativity. *Creat. Res. J.* **24**, 92–96 (2012).
12. Boudreau, K. J., Guinan, E. C., Lakhani, K. R. & Riedl, C. Looking across and looking beyond the knowledge frontier: intellectual distance, novelty and resource allocation in science. *Manag. Sci.* **62**, 2765–2783 (2016).
13. Rindova, V. P. & Petkova, A. P. When is a new thing a good thing? Technological change, product form design and perceptions of value for product innovations. *Organ. Sci.* **18**, 217–232 (2007).
14. Criscuolo, P., Dahlander, L., Grohsjean, T. & Salter, A. Evaluating novelty: the role of panels in the selection of R&D projects. *Acad. Manag. J.* **60**, 433–460 (2017).
15. Baldwin, C. Y. & Clark, K. B. *Design Rules: The Power of Modularity* (MIT, 2000).
16. Amabile, T. M. *Creativity in Context* (Westview, 1996).
17. Kerr, N. L. & Tindale, R. S. Group performance and decision making. *Annu. Rev. Psychol.* **55**, 623–655 (2004).
18. Runco, M. A., McCarthy, K. A. & Svenson, E. Judgments of the creativity of artwork from students and professional artists. *J. Psychol.* **128**, 23–31 (1994).
19. Kozbelt, A., Seidel, A., ElBassiouny, A., Mark, Y. & Owen, D. R. Visual selection contributes to artists' advantages in realistic drawing. *Psychol. Aesthet. Creat. Arts* **4**, 93–102 (2010).
20. Hickey, M. An application of Amabile's Consensual Assessment Technique for rating the creativity of children's musical compositions. *J. Res. Music Educ.* **49**, 234–244 (2001).
21. Kaufman, J. C., Baer, J. & Cole, J. C. Expertise, domains and the consensual assessment technique. *J. Creat. Behav.* **43**, 223–233 (2009).
22. Harvey, S. & Kou, C. Y. Collective engagement in creative tasks: the role of evaluation in the creative process in groups. *Adm. Sci. Q.* **58**, 346–386 (2013).
23. Khessina, O. M., Goncalo, J. A. & Krause, V. It's time to sober up: the direct costs, side effects and long-term consequences of creativity and innovation. *Res. Organ. Behav.* **38**, 107–135 (2018).
24. Csikszentmihalyi, M. *Handbook of Creativity* (Cambridge University Press, 1999).
25. Meyer, J. W. & Rowan, B. Institutionalized organizations: formal structure as myth and ceremony. *Am. J. Sociol.* **83**, 340–363 (1977).
26. DiMaggio, P. Culture and cognition. *Annu. Rev. Sociol.* **23**, 263–287 (1997).
27. Moreau, C. P., Lehmann, D. R. & Markman, A. B. Entrenched knowledge structures and consumer response to new products. *J. Mark. Res.* **38**, 14–29 (2001).
28. Kahneman, D., Sibony, O. & Sunstein, C. R. *Noise* (HarperCollins, 2022).
29. Amabile, T. M. A model of creativity and innovation in organizations. *Res. Organ. Behav.* **10**, 123–167 (1988).
30. Miron-Spektor, E. & Erez, M. in *The Oxford Handbook of Organizational Paradox* (eds Lewis, M. et al.) 434–451 (Oxford Univ. Press, 2017).
31. Mueller, J., & Yin, Y. in *Handbook of Research on Creativity and Innovation* (eds Zhou, J. et al.) 267–289 (Edward Elgar Publishing, 2021).
32. Fleming, L. Recombinant uncertainty in technological search. *Manag. Sci.* **47**, 117–132 (2001).
33. Mumford, M. D., Blair, C., Dailey, L., Leritz, L. E. & Osburn, H. K. Errors in creative thought? Cognitive biases in a complex processing activity. *J. Creat. Behav.* **40**, 75–109 (2006).

34. Licuanan, B. F., Dailey, L. R. & Mumford, M. D. Idea evaluation: error in evaluating highly original ideas. *J. Creat. Behav.* **41**, 1–27 (2007).
35. Elsbach, K. D. & Kramer, R. M. Assessing creativity in Hollywood pitch meetings: evidence for a dual-process model of creativity judgments. *Acad. Manag. J.* **46**, 283–301 (2003).
36. Goncalo, J. A., Flynn, F. J. & Kim, S. H. Are two narcissists better than one? The link between narcissism, perceived creativity and creative performance. *Pers. Soc. Psychol. Bull.* **36**, 1484–1495 (2010).
37. Mueller, J. S., Wakslak, C. J. & Krishnan, V. Construing creativity: the how and why of recognizing creative ideas. *J. Exp. Soc. Psychol.* **51**, 81–87 (2014).
38. Proudfoot, D. & Fath, S. Signaling creative genius: how perceived social connectedness influences judgments of creative potential. *Pers. Soc. Psychol. Bull.* **47**, 580–592 (2021).
39. Hunter, S. T., Blocker, L. D., Gutworth, M. B. & Allen, J. Why we support some original ideas but reject others: an application of signaling theory. *J. Creat. Behav.* **57**, 199–220 (2023).
40. Harrison, S. H. & Rouse, E. D. An inductive study of feedback interactions over the course of creative projects. *Acad. Manag. J.* **58**, 375–404 (2015).
41. Dow, J. & da Costa Werlang, S. R. Uncertainty aversion, risk aversion and the optimal choice of portfolio. *Econometrica* **60**, 197–204 (1992).
42. Whalen, J., Gallistel, C. R. & Gelman, R. Nonverbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* **10**, 130–137 (1999).
43. Craig, B. *Sundance—A Festival Virgin's Guide* 3rd edn (Cinemagine Media Publishing, 2016).
44. Goncalo, J. A. & Staw, B. M. Individualism–collectivism and group creativity. *Organ. Behav. Hum. Decis. Process.* **100**, 96–109 (2006).
45. Nemeth, C. J. Differential contributions of majority and minority influence. *Psychol. Rev.* **93**, 23 (1986).
46. Nemeth, C. J. & Staw, B. M. The tradeoffs of social control and innovation in groups and organizations. *Adv. Exp. Soc. Psychol.* **22**, 175–210 (1989).
47. Runco, M. A. & Smith, W. R. Interpersonal and intrapersonal evaluations of creative ideas. *Pers. Ind. Diff.* **13**, 295–302 (1992).
48. Kaufman, J. C., Baer, J., Cole, J. C. & Sexton, J. D. A comparison of expert and nonexpert raters using the Consensual Assessment Technique. *Creat. Res. J.* **20**, 171–178 (2008).
49. Berg, J. M. & Yu, A. Getting the picture too late: handoffs and the effectiveness of idea implementation in creative work. *Acad. Manag. J.* **64**, 1191–1212 (2021).
50. Simonton, D. K. Cinematic success criteria and their predictors: the art and business of the film industry. *Psychol. Mark.* **26**, 400–420 (2009).
51. Davis, B. C., Hmieleski, K. M., Webb, J. W. & Coombs, J. E. Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: the influence of perceived product creativity and entrepreneurial passion. *J. Bus. Ventur.* **32**, 90–106 (2017).
52. Weisberg, R. W. On the usefulness of 'value' in the definition of creativity. *Creat. Res. J.* **27**, 111–124 (2015).
53. Lee, C. Y. & Morewedge, C. K. Noise increases anchoring effects. *Psychol. Sci.* **33**, 60–75 (2022).
54. Ward, T. B., Patterson, M. J., Sifonis, C. M., Dodds, R. A. & Saunders, K. N. The role of graded category structure in imaginative thought. *Mem. Cogn.* **30**, 199–216 (2002).
55. Maio, S., Dumas, D., Organisciak, P. & Runco, M. Is the reliability of objective originality scores confounded by elaboration? *Creat. Res. J.* **32**, 201–205 (2020).
56. Baer, J. & McKool, S. S. in *Handbook of Research on Assessment Technologies, Methods and Applications in Higher Education* (ed. Schreiner, C. S.) 65–77 (IGI Global, 2009).
57. Wilson, T. D., Houston, C. E., Etling, K. M. & Brekke, N. A new look at anchoring effects: basic anchoring and its antecedents. *J. Exp. Psychol. Gen.* **125**, 387 (1996).
58. Lee, A. Welcome to the Unicorn Club: Learning from Billion-Dollar Startups. *TechCrunch* https://techcrunch.com/2013/11/02/welcome-to-the-unicorn-club/ (2013).
59. Sampling Source FAQs. *Lucid Holdings* https://support.lucidhq.com/s/article/Sample-Sourcing-FAQs (2022).
60. Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. G* Power 3: a flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
61. *Sundance Film Festival 2023: Submitting your project* (Sundance.org, 2023); https://dotorg-stg-assets.s3.amazonaws.com/pdf/submissions/2023_Submissions_FAQ.pdf
62. Coppock, A. & McClellan, O. A. Validating the demographic, political, psychological and experimental results obtained from a new source of online survey respondents. *Res. Pol.* **6**, 2053168018822174 (2019).

Acknowledgements

This research was supported by funds from the ILR School, Cornell University and an Innovation, Entrepreneurship and Technology theme grant from the Johnson College of Business, Cornell University. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We are grateful to E. Mannix and the members of ExPO Lab for their feedback on this research. We thank S. Parry for advice on statistical analyses. We thank S. Owens for providing information about the film categories at Sundance.

Author contributions

W.J. developed the study concept. W.J. and D.P. designed the studies. W.J. and D.P. collected and analysed the data. W.J. drafted the manuscript. D.P. revised the manuscript and prepared the final text for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01794-4.

Correspondence and requests for materials should be addressed to Devon Proudfoot.

Peer review information *Nature Human Behaviour* thanks Samuel Hunter and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Devon Proudfoot

Last updated by author(s): Aug 30, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We collected the data for Study 1, Study 3, Study 4, and Study 5 using Qualtrics.
Data analysis	All in text analyses were conducted using SPSS Version 29.0 with the exception of the sensitivity analyses which were conducted using G*Power Version 3.1.9.7. and the Cliff's delta effect sizes for Study 5 which were generated using R version 4.3.1. Figures 1 and 3 were generated using R Version 4.3.1. The code to replicate the analyses in the manuscript and our Online Supplement is available permanently and publicly on the Open Science Framework at: https://osf.io/h3puf/?view_only=f042b664a538496887a30e9e791fbca9

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

For all studies, data are available at https://osf.io/h3puf/?view_only=f042b664a538496887a30e9e791fbca9.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	In Studies 1, 3, 4, and 5, we asked study participants to self report their gender. We did not ask participants to report their sex in any of our studies. In studies in which participant gender was asked, the sample was roughly gender balanced. We report participant self-reported gender for all studies in which this question was asked. The data posted on OSF (Open Science Framework) for the manuscript includes participant gender data for the studies in which this question was asked. We did not have any a priori hypotheses regarding how participant gender would moderate our effects, so we did not conduct these analyses. We have no a priori reason to believe our predicted effect should apply to only one gender.
Population characteristics	See above
Recruitment	For studies in which we recruited participants (Studies 1, 3, 4 5), participants were recruited from commonly used online panels (Mturk, Prolific, Lucid). Selection bias is relatively limited on these platforms because individuals using these platforms generally complete a broad range of surveys and our studies involved easy tasks (i.e., rating ideas).
Ethics oversight	All studies received ethics approval from the institutional review board at Cornell University (Protocols #2102010124 and #1905008829) and comply with all relevant ethical regulations

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative
Research sample	Study 1- sample of U.S. adults recruited from Mturk (Mage = 33.25, SD = 9.79; 49% women, 50% men; 1% other gender identity). No participants were excluded from analyses. Sample was not representative. Sample was chosen for convenience. Participants rated the value offered by 250 high novelty pitches rated and the 250 low novelty pitches. Study 2-Data were collected from: http://cannes-ratings.herokuapp.com/Sundance2022, which presents auto-aggregated web-based evaluations from audience members at Sundance, typically a mix of industry insiders, film makers, journalists, and film fans. Ratings are aggregated in real-time during the festival. Of the 539 films listed on a website, 16 films had one rating; thus a standard deviation could not be computed, leaving 523 films in our analyses. Sample was chosen to assess the robustness of our predicted effect with participants with domain knowledge rather than a general population sample. We compared average standard deviation in ratings between 4 novel novelty film categories and 2 high novelty film categories. Study 3-representative sample of U.S. adults recruited from Lucid Theorem (Mage = 44.42, SD = 16.78; 51% women, 49% men). No participants were excluded from analysis. Sample was chosen for convenience. Participants judged the value of a subset of 80 paintings, half of which were construed as high novelty within the study context and half of which were construed as low novelty within the study context. Study 4-sample of U.S. residents with experience working in the hospitality industry recruited from Prolific (48% women, 51% men, Mage = 36.01, SDage = 12.22). No participants were excluded from analysis. This sample was non-representative. Sample was chosen to assess the robustness of our predicted effect with participants with domain knowledge rather than a general population sample. Participants were randomly assigned to either a baseline condition or a novelty-as-value condition in which we defined value as dependent on novelty. Participants in both conditions rated the value of a subset of 19 low novelty sandwich ideas and 19 high value sandwich ideas. Study 5-sample of U.S. residents with investment experience recruited from Prolific (Mage = 42.89, SD = 16.78; 50% women, 50% men). No participants were excluded from analysis. This sample was non-representative. Sample was chosen because our study examined investment intentions, thus a sample of participants with investment experience provided an ecologically valid test. Participants were then randomly assigned to one of two conditions (high variability in idea evaluations vs. low variability in idea

	evaluations). Participants rated their intent to invest in the idea, perceived risk, perceived novelty, and similarity in ratings (manipulation check).
Sampling strategy	We expected a small-to-medium sized effect for all studies. Studies 1, 3, 4, 5 were all conducted online--i.e., we posted the survey to an online platform and participants completed the survey via Qualtrics. Study 1 - A sensitivity analysis using G*Power indicated that, at α = .05, our sample size of 500 ideas provided 80% power to detect a minimum effect of d = 0.25. Study 2- We collected all available data points in the archival dataset. Study 3 -A sensitivity analysis using G*Power indicated that, at α = .05, our sample size of 160 paintings provided 80% power to detect a minimum effect of d = .45. Study 4 - A sensitivity analysis using G*Power indicated that, at α = .05, our sample size of 76 ideas provided 80% power to detect a minimum effect of ηp2 = .10. Study 5 - A sensitivity analysis using G*Power indicated that, at α = .05, our sample size of 401 participants provided 80% power to detect a minimum effect of d = 0.28.
Data collection	All data collection was done online using Qualtrics. The researchers were not blind to experimental condition or study hypotheses during data collection.
Timing	Study 1 - data collected from December 21 2022-Dec 31, 2022 Study 2 - data were all collected on August 16, 2022 Study 3 - data were all collected on August 14, 2022 Study 4 - data were all collected on December 29, 2021 Study 5 - data all collected on April 21, 2022
Data exclusions	Study 1- no data was excluded from analyses Study 2- Of the 539 films listed on a website, 14 films had one rating; thus a standard deviation could not be computed, leaving 525 films in our analyses. Study 3- no data was excluded from analyses Study 4- no data was excluded from analyses Study 5- no data was excluded from analyses
Non-participation	No participants dropped out/declined participation
Randomization	For Studies 1, 3, 4, and 5, participants were randomly assigned to groups using the randomization function in Qualtrics. For Study 2 (archival data), participants were not randomly assigned to groups. We only had access to film ratings in aggregate; we did not have access to data on individual participants and thus were not able to include covariates at the participant level.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging