

January 2020

**Sidi, Torgovitsky, Soibelman, Miron-Spektor & Ackerman (in press). You may be more original than you think: Predictable biases in self-assessment of originality. *Acta Psychologica*.**

<https://www.journals.elsevier.com/acta-psychologica>

This article may not exactly replicate the final version published in the journal.  
It is not the copy of record.

**You may be more original than you think:  
Predictable biases in self-assessment of originality**

Yael Sidi [1], Ilan Torgovitsky [2], Daniela Soibelman [2], Ella Miron-Spektor [3],  
and Rakefet Ackerman[2]

*[1]The Open University of Israel; [2] The Technion – Israel Institute of Technology; [3]*

*INSEAD*

Corresponding Author - E-mail: yaelsi@openu.ac.il

### **Abstract**

How accurate are individuals in judging the originality of their own ideas? Most metacognitive research has focused on well-defined tasks, such as learning, memory, and problem solving, providing limited insight into ill-defined tasks. The present study introduces a novel metacognitive self-judgment of originality, defined as assessments of the uniqueness of an idea in a given context. In three experiments, we examined the reliability, potential biases, and factors affecting originality judgments. Using an ideation task, designed to assess the ability to generate multiple divergent ideas, we show that people accurately acknowledge the serial order effect—judging later ideas as more original than earlier ideas. However, they systematically underestimate their ideas' originality. We employed a manipulation for affecting actual originality level, which did not affect originality judgments, and another one designed to affect originality judgments, which did not affect actual originality performance. This double dissociation between judgments and performance calls for future research to expose additional factors underlying originality judgments.

**Keywords:** metacognition; originality judgments; under-confidence; calibration bias; serial order effect

Original thinking—the ability to generate uncommon and unique ideas, has long been considered essential for scientific discoveries, growth, and prosperity in modern life (Amabile, 1983; Guilford, 1967; Torrance, 1966). The psychological processes enabling the generation of original ideas have been extensively studied (see Boot, Baas, van Gaal, Cools, & De Dreu, 2017; Sowden, Pringle, & Gabora, 2015, for reviews). Extensive research has also dealt with self-assessments of creative thinking abilities. In a recent review Kaufman (2019) delineated types of self-assessments and concluded that "participants are not in touch with their abilities" and that creativity self-assessments (CSAs) "should not be considered a proper substitute for more objective or performance-based measures" (p. 190). However, the self-assessments reviewed by Kaufman referred to a global skill or ability to generate creative ideas. In the present study, we compared actual creative performance and corresponding self-assessments at an item-by-item level. Such a detailed analysis characterizes metacognitive research in other types of cognitive tasks (e.g., learning), yet has not been done before in the context of creativity. In particular, as detailed below, we focused on the originality aspect of creativity, involved in the initial stage of finding a solution for a challenge. At this stage, people generate several potential ideas that can later be examined further, depending on context and goals. We asked the following questions: How accurate are individuals in judging the originality of their own ideas? How tightly do originality judgments correspond with actual originality performance? Notably, exposing predictable biases in originality estimation is important for further understanding of the process that affects creative thinking effectiveness.

Metacognitive research suggests that when individuals engage in cognitive tasks, like solving a problem, their cognitive processes are accompanied by meta-level processes, including judging the adequacy of solutions that they consider and guiding

their effort investment (Thompson, Turner, & Pennycook, 2011). However, metacognitive judgments are known to rely on heuristic cues and consequently suffer from predictable biases (see Ackerman, 2019; Bjork, Dunlosky, & Kornell, 2013, for reviews). Thus, effort regulation based on biased judgments is expected to be less effective than when people have reliable judgments—those judgments that accurately reflect performance (e.g., Metcalfe & Finn, 2008).

Traditionally, metacognitive research has mostly focused on memory tasks, investigating the extent to which individuals reliably judge their knowledge (Bjork, et al., 2013). Recent works have extended this body of literature to more complex cognitive processes involved in reasoning (see Ackerman & Thompson, 2017, for a review; Schwartz & Efklides, 2012). Notably, both memory and reasoning tasks typically share the characteristic of being well-defined, with one correct answer for each question. In contrast, idea generation tasks, reflecting divergent thinking, require individuals to generate multiple alternative and distinct solutions to an ill-defined problem (Guilford, 1959).

Divergent thinking tasks have been extensively used to assess creative thinking (e.g., Acar & Runco, 2019; Forthmann, Bürkner, Szardenings, Benedek, & Holling, 2019; Runco & Acar, 2012) and were found to be associated with real-life scenarios involving creativity in organizations (Berg, 2016; Mueller, Melwani, Loewenstein, & Deal, 2018). In these open-ended tasks, people are commonly asked to come up with as many ideas as they can produce. The classic model for analyzing idea generation (Guilford, 1967; Torrance, 1966) delineated four distinct dimensions of creativity: fluency, flexibility, elaboration, and originality, which have been consistently validated across different tests and batteries (Bass, De Dreu, & Nijstad, 2008; Forthmann, et al., 2019; Reiter-Palmon, Forthmann, & Barbot, 2019). *Fluency* refers to the number of

distinct ideas that one generates. Notably, to avoid confusion with the term "fluency" commonly used in metacognitive research as ease of processing, we refer to this dimension of creativity here as *number of ideas*. *Flexibility* refers to the ability to approach a problem from different perspectives and is usually measured as the number of distinct semantic categories that a person accesses (Rietzschel, De Dreu & Nijstad, 2009). *Elaboration* refers to the appropriateness of the idea and the extent to which individuals explain the usefulness and practicality of the solution. Finally, the most defining characteristic of creativity is *originality*, which focuses on the rarity (or infrequency) of an idea in a specific sample, and directly assesses the *idea*, rather than the creator (Bass, et al., 2008; Forthmann, et al., 2019; Reiter-Palmon, et al., 2019; Runco & Chand, 1995).

The few studies that examined how people judge the originality of ideas focused on the evaluation of ideas generated by *others*. For example, Runco and Charles (1993) found a moderate correspondence between originality judgments made by experts and objective originality. Others showed that creators are more accurate than managers when judging others' ideas, but not when judging their own ideas (Berg, 2016; Fuchs, Sting, Schilckel & Alexy, 2019), and that the social context shapes these judgments (Elsbach & Kramer, 2003 ; Mueller, Melwani, Loewenstein, & Deal, 2018). Licuanan, Dailey, and Mumford (2007) demonstrated that people underestimate novel ideas unless given training to focus on the creative aspects of the task. Taken together, these findings suggest that originality judgments are affected by irrelevant information, such as social cues and stereotypes, and can be biased.

The present study investigated the reliability of *originality judgments* regarding *one's own initial list of ideas*. Understanding biases in originality judgments is highly important as originality judgments regarding initial ideas guide decisions of consequent

continuous creative efforts (Perry-Smith & Mannucci, 2017). According to the dual pathway model of creativity (De Dreu, Baas & Nijstad, 2008), prolonged efforts in producing a higher number of ideas in a given task contribute to originality. Such within-category persistence in generating ideas enables to move beyond trivial, highly accessible solutions. The effort invested in searching for additional ideas largely depends on the evaluation of the generated solutions. If, for example, individuals judge their idea as highly original, they may terminate the idea generation process earlier, producing fewer ideas, compared to when they think the idea is not original enough (De Dreu, et al., 2008). This means that if people underestimate the originality of their idea, they will persist longer in the creative process. Investing more effort in searching for additional solutions can increase originality but may leave fewer attentional resources for elaborating and developing the ideas and ensuring their usefulness. By exploring the metacognitive processes inherent in idea generation, we begin to address the timely yet understudied question of what influences originality judgments.

### **Experiment 1**

Experiment 1 investigated the overall reliability of originality judgments in an idea generation task (i.e., Divergent Thinking, Guilford, 1967). This task represents the initial ideation process in real-life creative situations that later leads to idea screening and development (Perry-Smith & Mannucci, 2017). Objects are presented (e.g., a brick) and participants generate as many possible uses for each object as they can. Unlike well-defined tasks, in this task, the quality measure is ill-defined. Ideas are considered original if they are relatively rare in a given sample (Amabile et al., 1983; Roskes, De Dreu & Nijstad, 2012). Following studies that used and validated this task (e.g. Forthmann, et al., 2019; Guilford, 1967; Runco & Acar, 2012), for each idea we computed an originality score using its frequency in the sample (Roskes et al., 2012).

Originality judgments were measured by asking participants to rate the frequency of each of their ideas on a scale ranging from 0 (suggested by nobody else) to 100 (suggested by everybody).

The metacognitive framework allows examining several dimensions of the reliability of these judgments by comparing the rarity of generated ideas to their subjective judgments. First, from a process perspective, studies using the divergent thinking task have documented a *serial order effect*: usually when people engage in this task, their initial ideas are the more common and accessible ones, thus less original than the following ideas within the ideation process (e.g., Beaty & Silvia, 2012; Christensen, Guilford, & Wilson, 1957; Eliav & Miron-Spektor, 2015). Namely, originality scores typically increase along the ideation process (Osburn & Mumford, 2006). In this context, ideas from unconventional contexts are only generated after generation of conventional responses (Gabora, 2018, 2019). Importantly, it is an open question whether this naturally occurring pattern is also reflected in people's originality judgments. Research on creativity implies on some awareness to the creative process. For example, Jaarsveld and van Leeuwen (2005) asked participants to invent a logo for a novel brand of soft drink by producing a sequence of sketches. During the task, participants also provided self-evaluations of their progress. These evaluations were consistent with the stage reached in the designs. This is an encouraging finding, which directs us to expect awareness to this effect.

Overall, metacognitive research has revealed that item-based characteristics (e.g., familiarity, fluency) generate experience-based heuristic cues which dominate over cues generated by cross-item characteristics of the task (Koriat, 1997; e.g., test type, Ackerman & Zalmanov, 2012; study order, Castel, 2008; expected test time, Koriat, Bjork, Sheffer, & Bar, 2004; see Ackerman, 2019, for a review). Thus, the idea

generation serial order itself might be underweighted in originality judgments, when considering its value in absolute terms, relative to actual originality of each idea in a given population. On the other hand, it is reasonable to expect that performing the idea generation task is accompanied by experience-based fluency—perceived difficulty of generating each later idea relative to earlier ones. Fluency is a dominant heuristic cue in metacognitive research—easily generated responses are typically accompanied by higher metacognitive judgments (e.g., confidence) than those generated after lengthy thinking (e.g., Kelley & Lindsay, 1993). In the case of idea generation, fluency is expected to decrease as one progresses in the stream of thought since the ideation challenge increases after producing the more common uses. If fluency affects originality judgments in a similar manner to other metacognitive judgments, we expect originality judgments to reliably reflect the increase in originality as additional ideas are generated.

The second dimension of judgment reliability commonly used in metacognitive literature is *Calibration*, or absolute accuracy, which compares overall judgments with overall performance as a measure of under- or overconfidence. The typical tendency of people across various tasks is to show overconfidence—higher confidence than actual success rates (e.g., Ackerman & Zalmanov, 2012; Kominsky & Keil, 2014; Metcalfe, 1998; Prowse Turner, & Thompson, 2009). If originality judgments are similar to other metacognitive judgments, we should expect to find overestimation in this context as well. However, creativity research suggests that people tend to underestimate the novelty of other people's ideas (Licuanan et al., 2007), leading to a competing prediction. In particular, originality is context specific, as the same idea can be judged as novel in one context but not that original in another, depending on how frequent it is in that context. For example, creating a fresh bread vending machine may seem novel in some countries but is relatively common in remote villages in France. As people are

unaware of the uses generated by relevant others and tend to assume similarity, such underestimation of originality could also manifest for one's own ideas, resulting in underestimation. Finding underestimation in originality judgments would highlight their uniqueness relative to the tendency of overconfidence commonly found in judgments regarding well-defined tasks, as well as the task-dependency of calibration bias (Ackerman & Zalmanov, 2012; Thiede & de Bruin, 2011).

Finally, *Resolution*, or relative accuracy, measures discrimination between more and less original ideas. It is measured as a within-participant correlation between the evaluation of performance and actual performance in -1.0 to +1.0 range. Resolution in memory-related tasks is generally moderate. For example, in a series of experiments examining memorization of word pairs, Koriat, Ma'ayan, and Nussinson (2006) found resolution scores in a range of 0.44 to 0.56 (see Schwartz & Efklides, 2012; Goldsmith, Pansky, & Koriat, 2014, for reviews). We expected moderate resolution to manifest in originality judgments as well.

### **Method**

**Participants.** To determine sample size we used a power analysis (using GPower 3.1.9.4) for a repeated-measures ANOVA. We assumed a medium effect size ( $F = 0.23$ ) and based on a  $\alpha = .05$  and power = 0.95 the desired sample size was 64. Sixty-one undergraduate students from the Technion participated in a laboratory study for course credit (38% females,  $M_{age} = 25.25$ ,  $SD = 3.8$ ).

**Materials and Procedure.** Experimental sessions took place in a small computer lab (8 seats). Participants were presented with the Divergent Thinking Task in which they were asked to list as many possible uses for ten common household items. Each item was presented with both its name and a corresponding picture (Fig. 1). The first item, a brick, was used as a training item, and participants were required to provide two uses

for it. Following the training item, participants produced uses for nine more items (bucket, shoelace, sheet protector, office clip, shoebox, wine cork, hanger, pillowcase, and a drinking straw) in random order. Participants did not receive any feedback during the task. As recommended by Benedek, Mühlmann, Jauk, and Neubauer (2013), to maintain the validity of originality scores, each item was allotted two minutes. Following each given use, participants were asked to assess what percentages of their peers came up with this particular similar use. See Fig. 1. When presented, the marker on the originality judgment scale was located at the midpoint of the scale (50%).

*Fig. 1* Item example with a use of a participant typed in and his/her originality judgment.

## Measures.

*Originality score.* To determine the frequency of each idea in the sample, a trained rater coded all the ideas according to a predefined coding scheme (e.g., all uses which referred to using the object as a “paperweight” got the same single code). To examine agreement among judges, two other trained raters coded 25% of the ideas. Interrater

reliability was strong (Cohen's  $K=.867$ ,  $p < .001$ ). Originality score was calculated for each suggested use as its frequency (%) within the experiment's sample for the particular object (e.g., brick). The higher the score, the less original (more common) the suggested use.

*Originality judgments.* The originality rating each participant provided for each use. The higher the judgment, the less original (more common) the participant judged this use to be<sup>1</sup>.

*Calibration.* Under- or overestimation was calculated as the gap between mean originality judgments and mean originality score across uses for each participant. Higher values mean larger underestimation: the ideas were judged by the participants as less original (more common) than they actually were.

*Resolution.* Discrimination between more and less original ideas was calculated by a within-participant Pearson correlation between originality judgment and originality score, across uses that each participant provided for all items.

## Results

One item (“pillow cover”) was misunderstood by nine participants as a pillow. Their data was removed from the analysis. Data analysis across participants revealed that overall, originality judgments were biased in the direction of underestimation, which was significantly different than zero,  $t_{(60)} = 16.74$ ,  $p < .001$ , see Table 1.

---

<sup>1</sup> In a pilot study ( $N = 21$ ) we used a reversed scale for originality judgments (0% indicating low originality, 100% indicating high originality) to verify that the direction of the scale does not affect judgments. Judgments, actual originality, and their association in the pilot replicated the results reported here.

**Table 1***Descriptive statistics and analysis for Experiment 1*

	Overall	Early ideas	Later ideas	<i>F</i>	<i>MSE</i>	<i>P</i>	$\eta_p^2$
Originality score	23.6 (5.6)	35.6 (9.2)	12.2 (4.4)	445.83	31.283	.000	.899
Originality judgments	46.9 (12.1)	64.5 (13.6)	30.1 (13.8)	445.83	81.203	.000	.881
Calibration	23.3 (10.9)	28.9(10.6)	17.9(13.4)	67.98	54.539	.000	.531

We examined the serial order effect by the location within the sequence of uses each participant generated for each item. As early and late uses depend on each participant's overall stream of ideas, we divided the uses into early versus later ideas for each participant and each item, based on a median split, similar to Kapatsinski, Olejarczuk, and Redford (2017). This method is widely used in metacognitive research (e.g., Koriat & Sorka, 2015), and fits the data which is varied in the number of items generated by each participant<sup>2</sup>. A repeated-measures ANOVA of idea generation order as a within-participant factor on originality score revealed a significant effect, with later ideas being more original than early ideas, replicating the serial order effect. A similar analysis on originality judgments revealed a correspondence between judgments and the serial order effect, with later ideas judged as more original than early ideas. Importantly, even though underestimation decreased with later ideas compared to early ideas,

<sup>2</sup> A multilevel regression model was used to validate the median-split analysis (level 1: object, level 2: participant), using the packages of R software: lme4 and lmerTest (Bates, Mächler, & Bolker, 2015; Kuznetsova, Brockhoff, & Christensen, 2015). First, we predicted originality score by the idea order, as a continuous variable. A negative significant slope indicated originality score scaled with order, replicating the serial order effect,  $t(2879.4) = -13.45$ ,  $p < .001$ . Second, we predicted originality judgments by order. A negative significant slope indicated that originality judgments were sensitive to the serial order effect as well,  $t(2882.3) = -35.61$ ,  $p < .001$ . Finally, we examined the interaction between the two by predicting underestimation by order. A negative significant slope indicated that underestimation reduced with order,  $t(2884.9) = -30.43$ ,  $p < .001$ . Thus, the results of the multilevel regression models replicated the results of the median-split analysis.

underestimation was still large and significant for later ideas. Finally, resolution scores ( $M = .66$ ,  $SD = .13$ ,  $t(62) = 28.325$ ,  $p < .001$ ), indicated that despite the calibration bias, participants discriminated well between more and less original ideas.

In sum, Experiment 1 sheds light on similarities and dissimilarities between judgments for well-defined and ill-defined tasks. Originality judgments have good resolution, similar to other metacognitive judgments. Our findings also demonstrate that originality judgments correspond to the serial order effect, suggesting that they reflect experience-based cues. Finally, originality judgments show a unique pattern of underestimation for self-judgments, as previously shown for the evaluation of ideas generated by others (Licuanan et al., 2007).

### **Experiment 2**

Experiment 2 aimed to investigate the correspondence between factors affecting actual originality level and subjective originality judgments. Research on the validity of idea generation tasks for creativity research has shown that originality scores correlate with the stream of given ideas (Benedek, et al., 2013). Namely, when people generate a larger number of ideas, they gain higher originality scores. However, a question arises regarding the awareness of people to this effect. Namely, if people are motivated or influenced to generate more ideas, would originality judgments reflect an increase in their actual originality?

To answer this question, we sought to affect actual originality by manipulating the number of ideas generated by participants and examining the sensitivity of originality judgments to performance fluctuations. One way to control the number of generated ideas is by limiting them—asking participants to generate a preset number of responses (Hocevar, 1979b). However, this method eliminates experiencing the ideation stream (Benedek, et al., 2013). Thus, we used an implicit limiting method of anchoring.

Particularly, we presented participants with "information" regarding the number of ideas people usually generate in the task—either six or two ideas, representing the medium range of recommended ideas for valid scoring in idea generation tasks (Benedek, et al., 2013). We hypothesized that participants will generate more ideas, and consequently achieve higher originality scores when given a higher anchor (i.e., six ideas) than when given a lower anchor (i.e., two ideas). Our goal was to examine whether originality judgments would correspond to these differences in performance.

This experiment was conducted with an online sample rather than university students. This change in population was intended to examine the generalizability of the main phenomena.

### **Method**

**Participants.** To determine sample size we used a power analysis (using GPower 3.1.9.4) for a repeated-measures ANOVA. We assumed a medium effect size ( $F = 0.24$ ) and based on a  $\alpha = .05$  and power = 0.95 the desired sample size was 104. One hundred and three participants were recruited from Prolific Academic (<http://www.prolific.ac>) in exchange for 1.3£ for an estimated time of 15min. Two participants were excluded from the study following improper use of originality scale, leaving 101 participants (71% females;  $M_{age} = 31$ ,  $SD = 5.6$ ) in the analysis. All participants reported being native English speakers.

**Materials and Procedure.** The experiment was administered online in the Prolific Academic crowdsourcing service by a Qualtrics<sup>®</sup> questionnaire. Participants were randomly assigned to low-anchor or high-anchor groups. Following signing an informed consent form, both groups were presented with an attention screener to identify negligent participants. Participants read a brief passage wherein they were instructed to give a specific answer to the following question, regardless of its phrasing (the required

answer was a number, e.g., "257"). The question that appeared below the passage was "what day of the week is today?" Participants who typed the day of the week instead of the required number were considered inattentive to the task and were directed to the end of the study. Only participants who passed the screener continued to the next phase of the study, in which they were presented with the Divergent Thinking Task (Fig. 1).

Following the training item and the instructions, the high-anchor group received this text on a separate page: "Note that participants usually give six uses on average for each object", while the low-anchor group received the same text, with the number two. This text appeared again for both groups following the instructions for using the judgment scale. After completing the training and instructions, participants completed the task by producing uses for five items (i.e., bucket, shoelace, office clip, cork, and hanger) in random order. They rated their originality for each idea, as in Experiment 1.

**Measures.** The same measures as in Experiment 1 were used in Experiment 2.

### **Results**

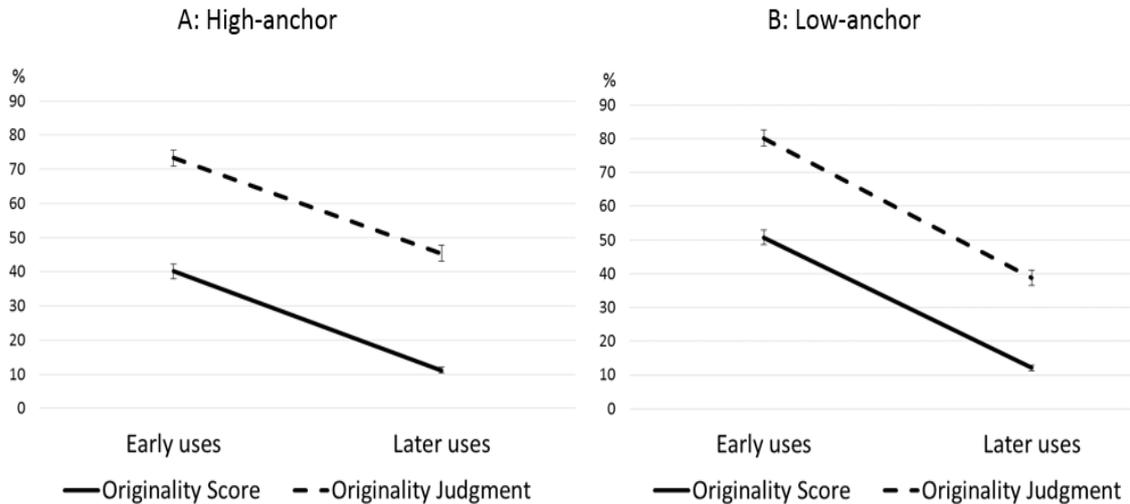
No participant failed the screening phase. To examine the reliability of originality judgments, we conducted three mixed Analyses of Variance (ANOVA) with generation order (early vs. later) as a within-participant factor and anchoring (high-anchor vs. low-anchor) as a between-participants factor for originality score, originality judgments, and calibration (Table 2).

**Table 2***Mixed Analyses of Variance (ANOVA) results in Experiment 2*

		<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta_p^2$
<i>Originality Score</i>	Order	521.46	110.472	<b>&lt;.001</b>	.840
	Anchoring	10.56	157.442	<b>.002</b>	.096
	Interaction	10.72	110.472	<b>.001</b>	.098
<i>Originality Judgment</i>	Order	344.56	175.542	<b>&lt;.001</b>	.777
	Anchoring	.004	370.293	.950	.000
	Interaction	13	175.542	<b>&lt;.001</b>	.116
Calibration	Order	0.42	100.421	.520	.004
	Anchoring	5.844	267.646	<b>.017</b>	.056
	Interaction	1.77	100.421	.186	.018

*Originality Scores*

Confirming the effectiveness of the manipulation, the high-anchor group produced more ideas ( $M = 21.5$ ,  $SD = 7.1$ ) than the low-anchor group ( $M = 15.1$ ,  $SD = 4.6$ ),  $t(101) = 5.490$ ,  $p < .001$ . An ANOVA revealed a main effect for order on originality score, with later uses more original than early uses, replicating the serial order effect as in Experiment 1 (Christensen, et al., 1957). As predicted, there was a main effect of anchoring: The high-anchor group had a higher originality score compared to the low-anchor group (Fig. 2). A significant interaction revealed that this effect stemmed from the differences in the earlier ideas,  $t(99) = 3.495$ ,  $p = .001$ , rather than in the later ideas,  $t < 1$ . This interaction was unexpected. As can be seen in Fig. 2, the later ideas were quite rare, with a mean of 10% of the population generating them. Thus, a possible reason for this interaction is a ceiling effect for later ideas, which are very rare, and thus cannot reflect an additional effect of the anchoring manipulation towards being more original.



*Fig 2.* Originality scores and originality judgments in Experiment 2. Panel A represents the high-anchor group and panel B represents the low-anchor group. NOTE: Lower values represent higher originality (less common ideas). Error bars represent standard error of the means.

### *Originality judgments*

In line with performance, originality judgments were affected by order, with later ideas being judged as more original (Fig. 2). While there was no main effect for anchoring on judgments, a significant interaction revealed that in earlier ideas judgments were lower in the high-anchor group,  $t(99) = 2.069, p = .041$ , while the opposite pattern was evident in later ideas,  $t(99) = 2.020, p = .046$ , though these effects were relatively small. This slope shift reflects partial sensitivity to performance fluctuations.

### *Calibration*

In accordance with the findings in Experiment 1, we also found a consistent underestimation in both groups,  $p$ 's < .001. An ANOVA on underestimation revealed a main effect for anchoring, with larger underestimation for the high-anchor group ( $M = 33.6, SD = 1.6$ ) compared to the low-anchor group ( $M = 28.1, SD = 1.6$ ). No other effect was significant. See Table 2.

### *Resolution*

An independent-sample t-test revealed that while both groups demonstrated reliable resolutions ( $p$ 's < .001), the high-anchor group had lower resolution ( $M = .5$ ,  $SD = .1$ ) compared to the low-anchor group ( $M = .6$ ,  $SD = .2$ ),  $t(99) = 2.256$ ,  $p = .026$ .

Overall, the findings demonstrate that both calibration and resolution were harmed for the high-anchor group compared to the low-anchor group. Finding a larger gap between performance and judgments for the high-anchor group may indicate that participants were unaware of the anchoring effect on their originality level. However, taken together with the finding regarding resolution, these results may reflect a trade-off between originality and the quality of the metacognitive monitoring. Particularly, it may be that investing more effort in generating a larger number of ideas leaves less resources for accurate monitoring of performance.

In sum, Experiment 2 replicated the patterns found in Experiment 1 with a different population. Here again, we found correspondence of originality judgments to the serial order effect, an overall underestimation bias, and good resolution. Introducing participants to a high anchor had the predicted effect of improving originality scores compared to a low anchor. However, judgments only partially corresponded to performance fluctuations, resulting in a larger calibration bias in the direction of underestimation, and lower resolution for the high-anchor group than for the low-anchor group. This is the first indication for differential effects on originality performance and judgments. The next experiment continued this investigation.

### **Experiment 3**

Both Experiment 1 and Experiment 2 demonstrated underestimation in originality judgments, which withstood even when performance was improved. In Experiment 3 we investigated the malleability of originality judgments. Particularly, we aimed to directly influence originality judgments, rather than originality performance.

For this purpose, we used a false-feedback manipulation, giving participants the impression that the two uses they provided during the training phase were either highly original, or unoriginal, compared to the study's sample, regardless of their actual level of originality. We expected this comparative false-feedback to affect originality judgments (Moore & Klein, 2008), with no subsequent effect on performance (Miller & Geraci, 2011).

### **Method**

**Participants.** To determine sample size we used a power analysis (using GPower 3.1.9.4) for a repeated-measures ANOVA. We assumed a medium effect size ( $F = 0.24$ ) and based on a  $\alpha = .05$  and power = 0.95 the desired sample size was 104. One hundred Prolific Academic participants were recruited as in Experiment 2. Four participants were excluded from the data set following failure on the screening attention question, after which they were directed to the end of the study, leaving 96 participants (72% females;  $M_{age} = 30.4$ ,  $SD = 5.1$ ) in the analyses. All participants were native English speakers.

**Materials and Procedure.** The entire procedure was identical to Experiment 2, except for a false-feedback manipulation: For each of the two uses given during the training phase participants received false-feedback indicating that their suggested use was either highly original or not original, regardless of its actual rarity in the sample (Fig. 3).

Correct.  
 Note that this use that you came up with was very rarely given by other participants in this study.

\*You are given this feedback only in the training phase. You will not receive similar feedback regarding your performance during the task itself.

>>

Correct.  
 Note that this use that you came up with was very commonly given by other participants in this study.

\*You are given this feedback only in the training phase. You will not receive similar feedback regarding your performance during the task itself.

>>

*Fig. 3* Screenshots of false-feedback indicating high originality (top) and feedback indicating low originality (bottom) in Experiment 3.

**Measures.** Same measures as in the previous experiments.

## Results

To examine whether the false-feedback affected performance and judgments, we ran three ANOVAs as in Experiment 2 (Table 3).

**Table 3**

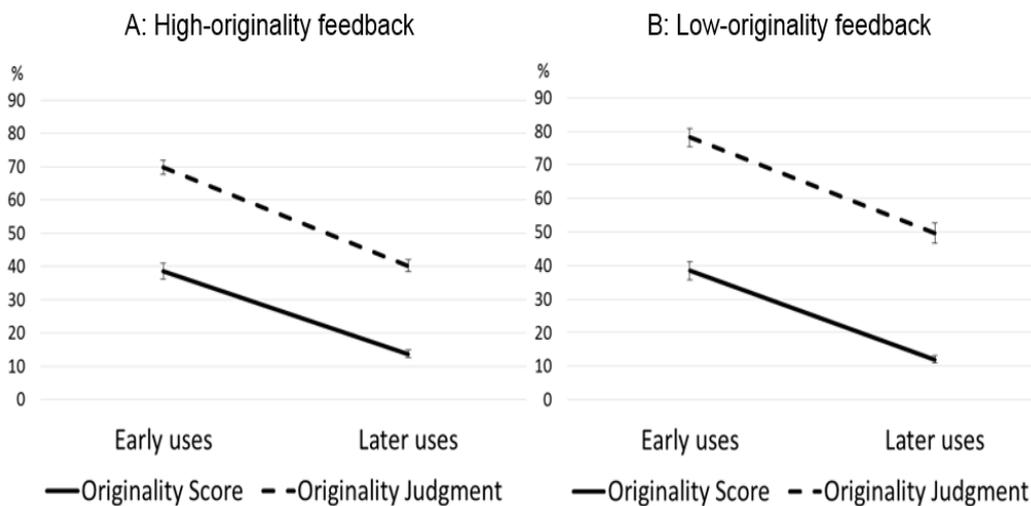
*Mixed ANOVA results in Experiment 3*

		<i>F</i>	<i>MSE</i>	<i>p</i>	$\eta_p^2$
<i>Originality Score</i>	Order	192.68	163.805	< <b>.001</b>	.672
	False-feedback	0.19	205.353	.658	.002
	Interaction	0.19	163.805	.662	.002
<i>Originality Judgment</i>	Order	203.69	197.850	< <b>.001</b>	.684
	False-feedback	10.17	375.871	<b>.002</b>	.098
	Interaction	0.08	197.850	.767	.001
Calibration	Order	3.47	136.986	.066	.036
	False-feedback	11.02	402.606	<b>.001</b>	.105

Interaction	0.49	136.986	.485	.005
-------------	------	---------	------	------

### Originality Scores

Replicating Experiment 1 and Experiment 2, the analysis revealed a main effect for order on originality score, in accordance with the serial order effect. However, the effect of the false-feedback manipulation and the interaction effect (see Table 3) were not significant, indicating that the feedback did not influence performance (Fig. 4). In correspondence, there was no difference in number of ideas generated between the high-originality feedback group ( $M = 18.1$ ,  $SD = 7.01$ ) and the low-originality feedback group ( $M = 18.1$ ,  $SD = 6.1$ ),  $t < 1$ .



*Fig 4.* Originality scores and originality judgments in Experiment 3. Panel A represents the high-originality feedback group, and panel B represents the low-originality feedback group. NOTE: Lower values represent higher originality (less common ideas). Error bars represent standard error of the means.

### Originality judgments

As in Experiment 1 and Experiment 2, originality judgments corresponded to the serial order effect (Fig. 4). While the false-feedback manipulation did not affect originality scores, it had a main effect on originality judgments. Specifically, in line with

our prediction, the high-originality feedback group evaluated their ideas as more original than the low-originality feedback group. This effect was similar for early and later uses.

### *Calibration*

As in Experiment 1 and Experiment 2, both groups consistently underestimated their performance,  $p$ 's  $< .001$ . A main effect of the false-feedback on underestimation indicates that underestimation was more pronounced for the low-originality feedback group ( $M = 38.7$ ,  $SD = 17.2$ ) compared to the high-originality feedback group ( $M = 29.1$ ,  $SD = 10.6$ ), regardless of order.

### *Resolution*

Both groups showed good discrimination between more and less original uses,  $p$ 's  $< .001$ . Comparing the groups, we did not find an effect of the false-feedback manipulation on resolution,  $t(94) = 1.9$ ,  $p = .056$ .

In sum, Experiment 3 replicated the correspondence of originality judgments to the serial order effect as well as the pronounced underestimation and good resolution. The false-feedback manipulation affected judgments as expected, as well as the calibration bias. However, while bias was smaller for the high-originality feedback group, it was not eliminated. The combined findings of Experiment 2 and Experiment 3 demonstrate a double dissociation between performance and judgments in the context of assessing originality.

## **General Discussion**

The present study is the first to apply the metacognitive framework to the ill-defined task of idea generation. Despite its exploratory nature, the study reveals several important factors affecting originality itself, originality judgments, and the correspondence between them. This approach proved decisively that people are in touch with their own creative processes when considered at the item level, unlike when

self-assessing their general creativity abilities (Kaufman, 2019). Nevertheless, originality judgments are based on metacognitive processes which are not strictly tied to their originality performance, but prone to predictable biases.

Our first hypothesis regarded the sensitivity of originality judgments to the serial order effect. Our study confirmed that people acknowledge the rise in originality as the ideation process progresses. Considering the bases for originality judgments, a question which arises is whether this awareness is based on the experience of the task itself, by experience-based heuristic cues like fluency, or on some lay theories people have regarding the ideation process (Koriat, 1997; see Ackerman, 2019 for a review). As mentioned in the introduction, people tend to underweight task characteristics relative to item-based characteristics. However, there are exceptions to this general tendency. For instance, Sidi, Shpigelman, Zalamanov, and Ackerman (2017) found with problem-solving tasks that people's confidence ratings are sensitive to time frame, pressured versus loose, yet are insensitive to the medium, screen versus paper. In fact, these two task-characteristics interacted in their effects on success rates; an interaction that was not reflected in confidence ratings. Further research is required to reveal factors that underlie the reflection of generation order in originality judgments.

Our second focus was the calibration bias of originality judgments. As previously found for judgments of other's ideas (Licuanan, 2007), our study revealed a resilient bias across various conditions in the direction of underestimation. The robustness of underestimation in originality judgments is extraordinary in light of the ubiquitous overconfidence in the metacognitive literature (see Dunlosky & Metcalfe, 2009; Koriat, 2016, for reviews). Moreover, the fact that participants correctly identified the fluctuations in their originality level based on generation order and their robust

resolution strengthen the notion that judgments were not given at random or a preset level. Rather, it reflects a variation in judgments based on reliable item-related cues.

Why then are people so under-confident regarding their originality? A possible explanation stems from the uniqueness of this judgment: while other metacognitive judgments are made in reference to objective performance (e.g., “How confident are you that your solution is correct?”; Ackerman & Thompson, 2017; Bjork et al., 2013), the correspondence of originality judgments to performance relies on a comparison to peers, which involves greater uncertainty. In fact, judgments about objective performance and judgments based on a comparative component sometimes demonstrate an opposite calibration bias, even within the same task (Moore & Healy, 2008). Underestimation, in this case, reflects that participants estimate their responses as relatively common in the sample. This bias is in line with the False Consensus Bias (Ross, Green, & House, 1977), which describes the belief that others in the group think in a similar manner to oneself (Nickerson, Baddeley & Freeman, 1987; see Dunning, 2015, for a review). This notion is supported by our finding that originality judgments were affected by information about others’ performance, as underestimation was reduced, yet not eliminated when participants received high-originality false feedback regarding their originality compared to their peers.

Together, these findings strengthen the task-dependency of calibration (see Thiede & de Bruin, 2011, regarding reading comprehension and Ackerman & Zalmanov, 2012, regarding problem-solving), emphasizing the importance of investigating judgments in various relevant contexts. Moreover, they suggest that originality judgments rely, at least in part, on social cues, as by definition original ideas should be uncommon in a given context. By drawing from research on the influence of social context and roles on creativity evaluation (Fuchs et al., 2019; Mueller et al., 2018;

Runco & Smith, 1992), our study contributes to the scarce metacognitive research which takes social considerations into account (Ackerman & Goldsmith, 2008; Eskenazi et al., 2016; Sidi, Ackerman, & Erez, 2018). Further research is required for delving into social aspects of our findings and for considering other methods for minimizing the bias of originality judgments, as achieved with other metacognitive biases (e.g., Szpunar, Jing & Schacter, 2014).

Interestingly, while participants assessed their uses to be similar to others', they were able to discriminate between more and less original ideas, as reflected in a consistent good resolution. Beyond the discrimination itself, this finding also demonstrates that calibration and resolution are distinct aspects of originality judgments reliability, as previously found for other metacognitive judgments (Fleming & Lau, 2014; Koriat & Goldsmith, 1996; Sidi, Ophir, & Ackerman, 2016).

Finally, the study demonstrates a dissociation between originality judgments and actual originality. In Experiment 2, we found that originality judgments were insensitive to performance fluctuations, while Experiment 3 revealed that originality judgments corresponded to the false-feedback participants received during training, while performance was unaffected. The study establishes a double dissociation between originality and its assessment, as previously found for judgments regarding memorization (e.g., Besken & Mulligan, 2013; Metcalfe & Finn, 2008). This dissociation should be taken into consideration when external factors are used to influence originality level (e.g., thought probes that encourage mind wandering, Hao, Wu, Runco & Pina, 2015) under the assumption that originality judgments would correspond, and vice versa.

Notably, in Experiment 1 we used a sample of engineering students and in Experiment 2 and Experiment 3, we used an online sample from the general population. These two diverse populations clearly differ in both education level and profession. Thus, our results are robust across populations.

Clearly, this is an initial study that introduces a new metacognitive judgment type that takes place at the beginning of the creative process, during the ideation stage. Metacognitive research typically aims to expose biases in judgments for understanding their unique bases that are not directly associated with the performance they should reflect (see Ackerman, 2019). It also focuses on proving associations between judgments and the regulatory processes people are involved in based on these judgments. Metacognitive research already established that judgments guide behavior (e.g., Metcalfe & Finn, 2008). In a recent paper, focused on the initial solving process, Lauterman and Ackerman (in press) found that the initial judgment of solvability of a problem, provided after a brief impression of the task, predicted the time people invested in solving attempts later on, above and beyond problems' difficulty and actual solvability. As for originality judgments, it is an important future direction to examine their predictive value for the following creative processes, such as screening ideas, delving to some of them, and deciding on the best fit for the task at hand. In particular, it is interesting to examine their role in contexts that have relevant research into other aspects of creativity, like organizational contexts (e.g., Berg, 2016; Mueller et al., 2018).

In conclusion, our study introduces a new metacognitive judgment that accompanies idea generation, demonstrating both similarities to other metacognitive judgments, alongside some unique characteristics. Most exceptional is the underestimation that seems to characterize originality judgments distinctively from other metacognitive judgments, which tend to show over-confidence. From a practical

standpoint, underestimation bias might have negative implications for those relying on originality assessments of their own ideas, such as engineers, entrepreneurs, and scientists. Namely, underestimation might lead to abandoning promising ideas that are worth further examination and investment of unnecessary effort in the ideation process. This research presents the first step in understanding the heuristic cues underlying originality judgments and their relationship to original thinking outcomes. Future research is called to further investigate other dimensions of creative evaluation (Christensen & Ball, 2016), and the numerous open questions this research highlights.

### **Disclosers**

Declarations of interest: None.

Funding: none.

Ethics: The study was approved by a Social & Behavioral Sciences Institutional Review Board at the Technion—Israel Institute of Technology (Approval Number 2017 – 27).

### **References**

- Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 153-158. <http://dx.doi.org/10.1037/aca0000231>
- Ackerman, R. (2019). Heuristic Cues for Meta-Reasoning Judgments: Review and Methodology. *Psihologijske teme, 28*(1), 1-20. <https://doi.org/10.31820/pt.28.1.1>
- Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting – with and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1224–1245. <https://doi.org/10.1037/a0012938>

Ackerman, R., & Thompson, V. A. (2017). Meta-Reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607-617.

<https://doi.org/10.1016/j.tics.2017.05.004>

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychonomic Bulletin & Review*, 19(6), 1187-1192.

<https://doi.org/10.3758/s13423-012-0305-z>.

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology*, 45(2), 357-376.

<http://dx.doi.org/10.1037/0022-3514.45.2.357>

Arkes, H. R. (1979). Competence and the overjustification effect. *Motivation and Emotion*, 3(2), 143-150. <https://doi.org/10.1007/BF01650599>

Baas, M., De Dreu, C. K., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus?.

*Psychological bulletin*, 134(6), 779-806. <https://doi.org/10.1037/a0012815>.

Bates, D. M., Maechler, M., & Bolker, B. (2015). lme4: Linear mixedeffects models using S4 classes (R package version 1.1-6) [Software].

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks.

*Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 309-319.

<http://dx.doi.org/10.1037/a0029171>

Benedek, M., Mühlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of*

*aesthetics, creativity, and the arts*, 7(4), 341-349.

<http://dx.doi.org/10.1037/a0033644>

Berg, J. M. (2016). Balancing on the creative high-wire: Forecasting the success of novel ideas in organizations. *Administrative Science Quarterly*, 61, 433–468.

<https://doi.org/10.1177/0001839216642211>

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, 41, 897–903.

<https://doi.org/10.3758/s13421-013-0307-8>

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444.

<https://doi.org/10.1146/annurev-psych-113011-143823>

Boot, N., Baas, M., Van Gaal, S., Cools, R., & De Dreu, C. K. W. (2017). Creative cognition and dopaminergic modulation of fronto-striatal networks: Integrative review and research agenda. *Neuroscience & Biobehavioral Reviews*, 78, 13–23.

<https://doi.org/10.1016/j.neubiorev.2017.04.007>

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, 36(2), 429-437.

<https://doi.org/10.3758/MC.36.2.429>

Christensen, B. T., & Ball, L. J. (2016). Dimensions of creative evaluation: Distinct design and reasoning strategies for aesthetic, functional and originality judgments.

*Design Studies*, 45, 116-136. <https://doi.org/10.1016/j.destud.2015.12.005>

- Christensen, P. R., Guilford, J. P., & Wilson, R. C. (1957). Relations of creative responses to working time and instructions. *Journal of Experimental Psychology*, *53*, 82–88. <https://doi.org/10.1037/h0045461>.
- De Dreu, C. K. W., Baas, M., & Nijstad, B. A. (2008). Hedonic tone and activation in the mood–creativity link: Towards a dual pathway to creativity model. *Journal of Personality and Social Psychology*, *94*, 739–756. <https://doi.org/10.1037/0022-3514.94.5.739>.
- Dunning, D. (2015). On Identifying Human Capital: Flawed Knowledge Leads to Faulty Judgments of Expertise by Individuals and Groups. *Advances in Group Processes*, *149–176*. <https://doi.org/10.1108/s0882-614520150000032006>
- Dunlosky, J., & Metcalfe, J. (2008). *Metacognition: A textbook for cognitive, educational, lifespan and applied psychology*. Thousand Oaks, CA: SAGE.
- Eliav, E., & Miron-Spektor, E. (2015). The recovery effect: The creative potential of frequent interruptions. *Academy of Management Best Paper proceedings*. *10.5465/AMBPP.2015.229*.
- Elsbach, K., & Kramer, R. (2003). Assessing creativity in Hollywood pitch meetings: Evidence for a dual-process model of creativity judgments. *Academy of Management Journal*, *46*: 283–301. <https://doi.org/10.5465/30040623>
- Eskenazi, T., Montalan, B., Jacquot, A., Proust, J., Grèzes, J., & Conty, L. (2016). Social influence on metacognitive evaluations: The power of nonverbal cues. *The Quarterly Journal of Experimental Psychology*, *69*(11), 2233–2247. <https://doi.org/10.1080/17470218.2015.1115111>

- Forthmann, B., Bürkner, P. C., Szardenings, C., Benedek, M., & Holling, H. (2019). A New Perspective on the Multidimensionality of Divergent Thinking Tasks. *Frontiers in psychology, 10*, 985. <https://doi.org/10.3389/fpsyg.2019.00985>
- Fuchs, C., Sting, F., Schlickel, M., & Alexy, O. (2019). The Ideator's Bias: How Identity-induced Self-efficacy Drives Overestimation in Employee-driven Process Innovation. *Academy of Management Journal*. <https://doi.org/10.5465/amj.2017.0438>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 1–9. <https://doi.org/10.3389/fnhum.2014.00443>
- Gabora, L. (2018). The neural basis and evolution of divergent and convergent thought. In O. Vartanian & R. Jung (Eds.) *The Cambridge Handbook of the Neuroscience of Creativity* (pp. 58-70). Cambridge MA: Cambridge University Press.
- Gabora, L. (2019). Reframing convergent and divergent thought for the 21st Century. In A. Goel, C. Seifert, & C. Freska (Eds.) *Proceedings of the 2019 Annual Meeting of the Cognitive Science Society*, (pp. 1794-1800). Austin TX: Cognitive Science Society.
- Goldsmith, M., Pansky, A., & Koriat, A. (2014). Metacognitive control of memory reporting. In T. J. Perfect, & D. S. Lindsay (Eds.), *Sage handbook of applied memory* (pp. 481–500). London, UK: Sage.
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.
- Guilford, J.P. (1959). Three faces of intellect. *The American Psychologist, 14*, 469–479. <http://dx.doi.org/10.1037/h0046827>

- Hao, N., Wu, M., Runco, M. A., & Pina, J. (2015). More mind wandering, fewer original ideas: Be not distracted during creative idea generation. *Acta Psychologica, 161*, 110–116. <http://dx.doi.org/10.1016/j.actpsy.2015.09.001>
- Hocevar, D. (1979b). Ideational Fluency as a Confounding Factor in the Measurement of Originality. *Journal of Educational Psychology, 71*, 191-196. <https://doi.org/10.1037/0022-0663.71.2.191>
- Jaarsveld, S., & van Leeuwen, C. (2005). Sketches from a design process: Creative cognition inferred from intermediate products. *Cognitive Science, 29*(1), 79-101. [https://doi.org/10.1207/s15516709cog2901\\_4](https://doi.org/10.1207/s15516709cog2901_4)
- Kapatsinski, V., Olejarczuk, P., & Redford, M. A. (2017). Perceptual learning of intonation contour categories in adults and 9 to 11 year old children: adults are more narrow minded. *Cognitive science, 41*(2), 383-415. <https://doi.org/10.1111/cogs.12345>
- Kaufman, J. C. (2019). Self-assessments of creativity: Not ideal, but better than you think. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 187-192. <https://doi.org/10.1037/aca0000217>
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*(1), 1-24. <https://doi.org/10.1006/jmla.1993.1001>
- Kominsky, J. F., & Keil, F. C. (2014). Overestimation of knowledge about word meanings: The “misplaced meaning” effect. *Cognitive science, 38*(8), 1604-1633. <https://doi.org/10.1111/cogs.12122>

- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349-370. <http://dx.doi.org/10.1037/0096-3445.126.4.349>
- Koriat, A. (2016). Metacognition: Decision-making processes in self-monitoring and self regulation. In G. Keren, & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making*, Volume 1 (pp. 356-379). Malden, MA: Wiley–Blackwell.
- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643-656.  
<http://dx.doi.org/10.1037/0096-3445.133.4.643>
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490-517.
- Koriat, A. & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments. In S. Chaiken and Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483-502). New York: Guilford.
- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36-69, <http://dx.doi.org/10.1037/0096-3445.135.1.36>
- Koriat, A., & Sorka, H. (2015). The construction of categorization judgments: Using subjective confidence and response latency to test a distributed model. *Cognition*, *134*, 21-38. <https://doi.org/10.1016/j.cognition.2014.09.009>

- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package) (R package version 2.0-6) [Software].
- Lauterman, T., & Ackerman, R. (in press). Initial Judgment of Solvability in non-verbal problems – A predictor of solving processes. *Metacognition and Learning*.
- Licuanan, B. F., Dailey, L. R., & Mumford, M. D. (2007). Idea evaluation: Error in evaluating highly original ideas. *Journal of Creative Behavior*, 41(1), 1–27.  
<https://doi.org/10.1002/j.2162-6057.2007.tb01279.x>.
- Metcalf, J. & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15, 174–179.  
<https://doi.org/10.3758/PBR.15.1.174>
- Metcalf, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics?. *Personality and Social Psychology Review*, 2(2), 100-110. DOI: 10.1207/s15327957pspr0202\_3
- Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning*, 6, 303–314. <https://doi.org/10.1007/s11409-011-9083-7>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, 115, 502-517. <https://doi.org/10.1037/0033-295X.115.2.502>
- Moore, D. A., & Klein, W. M. P. (2008). Use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organizational Behavior and Human Decision Processes*, 107, 60–74.  
<https://doi.org/10.1016/j.obhdp.2008.02.005>

- Mueller, J., Melwani, S., Loewenstein, J., & Deal, J. J. (2018). Reframing the decision-makers' dilemma: Towards a social context model of creative idea recognition. *Academy of Management Journal*, *61*(1), 94-110. <https://doi.org/10.5465/amj.2013.0887>
- Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know? *Acta Psychologica*, *64*(3), 245–259. [https://doi.org/10.1016/0001-6918\(87\)90010-2](https://doi.org/10.1016/0001-6918(87)90010-2)
- Osburn, H. K., & Mumford, M. D. (2006). Creativity and planning: Training interventions to develop creative problem-solving skills. *Creativity Research Journal*, *18*, 173–190. [https://doi.org/10.1207/s15326934crj1802\\_4](https://doi.org/10.1207/s15326934crj1802_4)
- Perry-Smith, J. E., & Mannucci, P. V. (2017). From creativity to innovation: The social network drivers of the four phases of the idea journey. *Academy of Management Review*, *42*(1), 53-79. <https://doi.org/10.5465/amr.2014.0462>
- Prowse Turner, J. A., & Thompson, V. A. (2009). The role of training, alternative models, and logical necessity in determining confidence in syllogistic reasoning. *Thinking & Reasoning*, *15*(1), 69-100. <https://doi.org/10.1080/13546780802619248>
- Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 144-152. <http://dx.doi.org/10.1037/aca0000227>
- Rietzschel, E. F., De Dreu, C. K. W., & Nijstad, B. A. (2010). What are we talking about when we talk about creativity: Group creativity as a multifaceted multistage phenomenon. In E. A. Mannix, M. Neale, & J. A. Goncalo (Eds.). *Research on managing groups and teams: Creativity in Groups* (vol. 12, pp. 1–23). Palo Alto, CA: Jossey-Bass.

- Roskes, M., De Dreu, C. K., & Nijstad, B. A. (2012). Necessity is the mother of invention: Avoidance motivation stimulates creativity through cognitive effort. *Journal of personality and social psychology, 103*(2), 242-256.  
<https://doi.org/10.1037/a0028442>
- Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279-301. [https://doi.org/10.1016/0022-1031\(77\)90049-X](https://doi.org/10.1016/0022-1031(77)90049-X)
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66-75.  
<https://doi.org/10.1080/10400419.2012.652929>
- Runco, M. A., & Chand, I. (1995). Cognition and creativity. *Educational psychology review, 7*(3), 243-267. <https://doi.org/10.1007/BF02213373>
- Runco, M. A., & Charles, R. (1993). Judgments of originality and appropriateness as predictors of creativity. *Personality and Individual Differences, 15*, 537-546.  
[https://doi.org/10.1016/0191-8869\(93\)90337-3](https://doi.org/10.1016/0191-8869(93)90337-3)
- Runco, M.A. & Smith, W.R., (1992). Interpersonal and intrapersonal evaluations of creative ideas. *Personality and Individual Differences, 14*(3), 295-302.  
[https://doi.org/10.1016/0191-8869\(92\)90105-X](https://doi.org/10.1016/0191-8869(92)90105-X)
- Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition, 1*, 145-151. <https://doi.org/10.1016/j.jarmac.2012.06.002>

- Sidi, Y., Ackerman, R., & Erez, A. (2018). Feeling happy and (over) confident: the role of positive affect in metacognitive processes. *Cognition and Emotion*, *32*(4), 876-884. <https://doi.org/10.1080/02699931.2017.1346501>
- Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority-does the medium, screen versus paper, affect performance even with brief tasks?. *Metacognition and Learning*, *11*(1), 15-33. <https://doi.org/10.1007/s11409-015-9150-6>
- Sidi, Y., Shpigelman, M., Zalmanov, H., & Ackerman, R. (2017). Understanding metacognitive inferiority on screen by exposing cues for depth of processing. *Learning and Instruction*, *51*, 61-73. <https://doi.org/10.1016/j.learninstruc.2017.01.002>
- Sowden, P. T., Pringle, A., & Gabora, L. (2015). The shifting sands of creative thinking: Connections to dual-process theory. *Thinking & Reasoning*, *21*, 40-60. <https://doi.org/10.1080/13546783.2014.885464>
- Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, *3*(3), 161-164. <https://doi.org/10.1016/j.jarmac.2014.02.001>
- Thiede, K. W., & de Bruin, A. B. (2011). Self-regulated learning in reading. In Schunk, D. (Ed.), Greene, J. (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 124-137). Routledge.
- Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive psychology*, *63*(3), 107-140. <https://doi.org/10.1016/j.cogpsych.2011.06.001>

Torrance, E. P. (1966). *Torrance tests of creative thinking: Norms-technical manual: Verbal tests, forms a and b: Figural tests, forms a and b*. Personal Press, Incorporated.