

University of Groningen

## An affect misattribution pathway to perceptions of intrinsic reward

Leander, N. Pontus; Kay, Aaron C.; Chartrand, Tanya L.; Payne, B. Keith

*Published in:*  
 Social Cognition

*DOI:*  
[10.1521/soco.2017.35.2.163](https://doi.org/10.1521/soco.2017.35.2.163)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
 Publisher's PDF, also known as Version of record

*Publication date:*  
 2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Leander, N. P., Kay, A. C., Chartrand, T. L., & Payne, B. K. (2017). An affect misattribution pathway to perceptions of intrinsic reward. *Social Cognition*, 35(2), 163-180.  
<https://doi.org/10.1521/soco.2017.35.2.163>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

## AN AFFECT MISATTRIBUTION PATHWAY TO PERCEPTIONS OF INTRINSIC REWARD

N. Pontus Leander  
*University of Groningen*

Aaron C. Kay and Tanya L. Chartrand  
*Duke University*

B. Keith Payne  
*University of North Carolina–Chapel Hill*

Intrinsic rewards are typically thought to stem from an activity's inherent properties and not from separable rewards one receives from it. Yet, people may not consciously notice or remember all the subtle external rewards that correspond with an activity and may misattribute some directly to the activity itself. We propose that perceptions of intrinsic reward can often be byproducts of misattributed causal inference, and present some initial evidence that perceptions of intrinsic reward can in fact increase when words pertaining to an activity are subtly paired with pleasant context cues. Importantly, these effects follow classic boundary conditions of both misattribution and intrinsic motivation, insofar as they were extinguished when participants could make a proper source attribution and/or when the activity became associated with a blatant external reward. We further propose a distinction can be made between authentically "intrinsic" rewards and the illusion of intrinsic rewards caused by misattributed positive affect.

*Keywords:* intrinsic reward, misattribution, positive affect

Illusion is the first of all pleasures.

Voltaire, *"The Maid of Orleans"* (*"La Pucelle d'Orléans"*)

---

Address correspondence to N. Pontus Leander, Department of Psychology, University of Groningen, 2/1 Grote Kruisstraat, 9712TS Groningen, The Netherlands; E-mail: n.p.leander@rug.nl.

Do all perceptions and experiences of intrinsic reward truly signal the satisfaction of psychological needs? Imagine someone who claims to like the piano and who seemingly revels in its simple pleasures. One might assume this person is *intrinsically* rewarded by the activity and, therefore, does not pursue it as a means to some other outcome but for the rewards intrinsic to the activity. Yet, perhaps this person did not always enjoy playing the piano; maybe it is just that, whenever he used to sit down to play as a child, his parents would lavish praise and be more attentive to him. Could these long-forgotten reward experiences have led him to associate playing the piano with feeling good? His subjective experience suggests intrinsic rewards, but it traces back to external rewards. We propose that perceptions of intrinsic reward can indeed be shaped by external rewards—particularly rewards that are forgotten or not noticed to begin with.

Intrinsic motivation is "...the doing of an activity for its inherent satisfactions rather than for some separable consequence" (Ryan & Deci, 2000, p. 56). The gist is that people can perceive activities to be ends in themselves rather than as means to some other outcome (Kruglanski, 1975; Lepper, Greene, & Nisbett, 1973). The rewards are internally generated rather than externally derived. If someone were to reflect on his or her lay beliefs about intrinsic rewards, the person might assume that it means to like an activity for its own sake, or he or she might say there is just something about the activity that fits his or her idiosyncratic preferences. However, just because people perceive an activity to be intrinsically rewarding does not mean they are right—many simple pleasures could stem from subtle external rewards that people do not consciously realize are separable from the activity itself.

We specifically consider the possibility that some external rewards are too subtle to consciously notice or remember, engendering a lack of source awareness about the positive feelings that correspond with one's thoughts about a given activity. When this happens, perceptions of intrinsic reward might arise through a process of *affect misattribution*—wherein one misattributes the positive affect that stems from an external reward to the activity itself. Theories of affect misattribution assume that people are not always consciously aware of the sources of their moods and other affective states, so they attribute them to something else (Leander, Moore, & Chartrand, 2009; Payne, Cheng, Govorun, & Stewart, 2005). Affect misattribution has been used to explain why people who are interviewed on a sunny day report greater life satisfaction (Schwarz & Clore, 1983), and why people who have recently exercised find erotic films more arousing (Cantor, Zillmann, & Bryant, 1975). From a misattribution perspective, an activity could seem intrinsically rewarding if it corresponds with positive affect of unknown origin.

Perhaps a distinction should thus be made between authentically intrinsic rewards and the simple pleasures produced by context cues. The longstanding view is that activities are (authentically) intrinsically rewarding when they satisfy innate psychological needs—such as needs for competence and self-determination (Deci & Ryan, 2000). However, Shah and Kruglanski (2000) noted there are some activities, such as lying on the beach, that could be experienced as intrinsic because they seem like ends in themselves, but not for reasons of competence or self-determination. We take a step further to note that lying on the beach confers an

array of external rewards—beaches are warm and sunlight has biological effects that can improve one's mood. Perhaps warmth and sunlight are so integral to our mental representation of *lying on the beach* that they seem inseparable from the behavior. Though remove the sun, and the behavior will probably lose its seemingly intrinsic appeal.

The distinction becomes ever more important when considering how to unpack intrinsic motivation as a construct, especially if one is concerned with the authenticity of intrinsic reward experiences and their implications for one's phenomenological sense of self. On the one hand, to maintain a coherent sense of self, misattributions of "intrinsic reward" may be a simple way to organize and provide meaning to incidental associations between activities and rewards—especially rewards that otherwise seem random or do not have clear causal relationship with the activity. On the other hand, if people misattribute subtle external rewards as intrinsic to a given activity, it would suggest that some perceptions of "intrinsic" reward are illusory—stemming not from the satiation of inherent psychological needs *per se*, but from conflating activities with the pleasant contextual stimuli they correspond with. This could lead to inaccurate beliefs about the self and misrepresentation of one's likes and dislikes. For example, educational toys and classrooms can be made cheery and colorful so they please the senses and thus attract interest and engagement, yet such features are separable from the function of the toy or the content of the class itself. These contextual rewards are experienced *while* one engages in the activity, not necessarily *from* the activity. Can one still come to authentically enjoy learning if one's initial attraction is based on misattribution?

## ON THE MISATTRIBUTION OF INTRINSIC REWARD

We essentially propose there are at least two pathways through which subjective experiences and perceptions of intrinsic reward can occur: first is when there is some *internal* source of positive feelings that stems from engagement in an activity (e.g., the activity satiates a psychological need); second is what we refer to as a misattribution pathway: when there is some *external* source of positive feelings that does not necessarily stem from the activity *per se*, but from a cognitive association between the activity and positive affect. A misattribution pathway would suggest people could believe a reward experience is intrinsic to an activity when they cannot attribute their positive feelings to the original source. The reward may be implicit, subtle, or so entangled with one's mental representation of the activity it is virtually impossible to consciously separate it. However, the reward is still a "separable consequence" in the sense it can be removed without changing the behavioral affordances of the activity itself.

Our logic builds upon a history of research suggesting that when people are not consciously aware of the source of a primed thought or feeling, they may misattribute it as internally generated (Loersch & Payne, 2011). For example, Bar-Anan, Wilson, and Hassin (2010) demonstrated a post-priming misattribution process, wherein an externally primed goal increased people's preference for certain be-

havioral choices but, because they were unaware of the influence, they misattributed the reason for their choice to some other internal state—such as interest in the activity itself. Accordingly, Parks-Stamm, Oettingen, and Gollwitzer (2010) found that people readily interpret nonconscious goal-directed behavior in terms of another goal that is more salient. In other words, people easily invent motivations to explain “why” a particular behavior occurred. Perhaps people also invent notions of intrinsic reward to interpret unexplained phenomenological states.

Our logic is also consistent with the classic covariation principle of causal attribution, wherein a cause-effect relationship is inferred when an effect corresponds with a possible cause over time (Kelley, 1973). If an activity corresponds with unexplained positive affect, it may invite an attribution of intrinsic reward. Research has shown that positive affect triggered by one object (e.g., a pleasant image) can increase evaluations of another object that follows (e.g., “liking” of neutral symbols; Murphy & Zajonc, 1993; Payne et al., 2005). Beliefs about intrinsic rewards may accordingly be increased by the correspondence between a behavior and unexplained affective rewards. For example, research on evaluative conditioning suggests that when an object word is repeatedly and implicitly paired with positively valenced stimuli, evaluations of the object word become more positive (De Houwer, Thomas, & Baeyens, 2001). Custers and Aarts (2005) similarly showed that when a word pertaining to an otherwise-neutral behavior is implicitly paired with positively valenced words (e.g., *sun*, *beach*, *friend*, *smile*, *home*), evaluations of the behavior became more positive (on a Likert-type scale that ranged from *very negative* to *very positive*). This suggests increased liking for the behavior. What remains unclear, however, is whether participants make an external or internal attribution for such liking. One could argue the conditioning creates a means-end relationship in the person’s mind (between the behavior and reaching desirable rewards), in which case the liking may be associated with the pursuit of external rewards. Yet Custers and Aarts (2005) offered an associative account for the relationship between the behavior and positive affect—which implies the liking would be perceived as intrinsic to the activity.

The subtlety of the external rewards may be key to any misattribution of intrinsic reward. Indeed, an external reward can be present and not undermine intrinsic motivation so long as one does not attribute one’s behavior to it (Kruglanski, Alon, & Lewis, 1972; Lepper et al., 1973); this is what we assume happens when the reward is subtle. For example, Eisenberger and Cameron (1996) suggested that, whereas blatant external rewards weaken intrinsic motivation, subtle rewards often do not (at least not to the same extent; e.g., Ross, 1975). Therefore, when we refer to “external rewards,” we do not necessarily mean material objects or prizes (e.g., gold stickers, money); rather, we refer to context cues that increase positive affect implicitly. Subtle context cues may be especially effective at fostering perceptions of intrinsic reward because they could trigger activation of positive subjective experiences without necessarily invoking an external attribution. Even if one is consciously aware of the context cues that provide the experiential rewards, one may still be unaware of their influence (Nisbett & Wilson, 1977; Wilson & Brekke, 1994); in social cognition research, the notion of supraliminal priming is

based on this idea. If the external rewards are sufficiently subtle, one may get to enjoy their positive affective outcomes yet still attribute the experience to the corresponding activity.

## THE PRESENT STUDIES

We present two pilot studies to show when people are likely to misattribute subtle external rewards as intrinsic to an activity. We consider these to be pilot studies because we used Internet-based samples of the general population, and thus sought to pair mental representations of a general activity category (learning) with positive pictures, to create an association between learning and positive affect. Human beings are presumably innately motivated to learn (Deci & Ryan, 2000), but we aim to show that external cues to positive affect can inflate the perceived intrinsic rewards of learning. Note that in these studies, we made no assumption the person had to be engaged in the activity when the reward is received, but rather the mental representation of the activity should be coactivated with the affective reward. Our logic is similar to that of evaluative conditioning, wherein increased evaluations of liking of a behavior can occur from pairing the behavior word (e.g., “studying”) with positive affect words (Custers & Aarts, 2005).

We did, however, assume that for affect misattribution to adequately explain some perceptions of intrinsic reward, we must fit our model to the theoretical frameworks of not just misattribution but also intrinsic motivation. Both frameworks provide insights into the attributions that could extinguish misattributed perceptions of intrinsic reward. From a misattribution perspective, any potential effects should, theoretically, be extinguished if one can properly attribute the reward experience to its true cause (i.e., precluding misattribution, Study 1). From an intrinsic motivation perspective, the effects should, theoretically, be extinguished if one can attribute one’s motivation to a blatant extrinsic reward (i.e., facilitating an external attribution, Study 2). Both boundary conditions must apply.

## STUDY 1

The primary aim of this study is to show that the subtle application of pleasant context cues can increase subjective perceptions of intrinsic reward. The second aim is to demonstrate that this type of influence follows the pattern for other misattribution effects, in that it only persists so long as the person cannot attribute the affective reward to its true cause (Murphy & Zajonc, 1993). For instance, Schwarz and Clore (1983) found the effects of interviewing people on a sunny day, on perceived life satisfaction, persisted only so long as the interviewer did not mention the weather. In a study on nonconscious goal failure, participants who were informed the difficult task may have made them feel bad were less likely to engage in self-enhancement afterwards (Chartrand, Cheng, Dalton, & Tesser, 2010). Accordingly, any influence of subtle external rewards should be extinguished when

participants are explicitly reminded of how the pleasant context cues made them feel.

## METHOD

*Participants.* Two hundred eighty American adults (141 female, Age  $M = 30.41$ ) completed the study via Amazon's Mechanical Turk. Demographic information from 23 of these participants was not obtained.

*Procedure.* This study used a 2 (context condition: positive vs. neutral)  $\times$  3 (attribution: none vs. reminder of images vs. specific source attribution) factorial design. Participants first completed a novel reward paradigm in which learning-related concepts were repeatedly paired with either positive or neutral context cues. Participants were given 20 matrices of letters ( $10 \times 4$ ), in a fixed sequence, along with instructions to search for a word between 4–7 characters long and then type it into the space provided. The words were always printed left to right, sometimes spilling over to the next line. Sixteen of the words pertained to academic learning (*intern, book, memory, study, train, inform, school, guide, tutor, pupil, pencil, teach, note, adept, class, mind*), and four were fillers (*bail, board, bracket, chair*). Importantly, the letter matrices were each superimposed over background images that were either positive or neutral in valence, according to normed ratings from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1995; see Appendix 1). Note that the pleasantness of images often corresponds with physiological indicators of experienced affect (Lang, Greenwald, Bradley, & Hamm, 1993), which means the images are not just positively evaluated but also rewarding. The instructions for the specific word length to search for on each trial were also slightly more pleasant in the positive condition.<sup>1</sup> There was no mention of the images. Thus, the task involved searching for learning-related words in either a pleasant or neutral context. No performance contingency was suggested and nearly all participants performed flawlessly.

Participants then received the attribution manipulation. In the “no reminder” condition, participants simply moved on to rate their intrinsic and extrinsic motivation. In the “reminder of images” condition, participants were instructed to think back to the pictures that appeared in the background of the word search task and rate their pleasantness and the extent to which they thought the images might have influenced them (on 7-point unnumbered scales with endpoints: “not at all” and “extremely pleasant” and “no influence at all” and “very strong influence,” respectively). Unsurprisingly, participants in the positive condition rated their images more pleasant ( $M = 5.88$ ,  $SD = 1.79$ ) than those in the neutral condition ( $M = 4.35$ ,  $SD = 2.24$ ),  $F(1, 96) = 13.92$ ,  $p < .001$ . However, perceived influence did not differ by condition and was generally quite low ( $M = 2.08$ ,  $SD = 1.80$  vs.  $M = 2.47$ ,  $SD = 1.94$ ,  $F < 1.1$ ). Participants who received the reminder were aware of the

---

1. In the positive condition, participants were instructed to *look for* as opposed to *find* the words (e.g., “please look for [find] the five-letter word in this letter string.”). A pilot test indicated that being asked to “look for” an object is slightly more pleasant than being asked to “find” an object ( $M = 2.49$  vs.  $2.34$  on a 4-point scale [where 1 = neutral, 2 = slightly pleasant, 3 = pleasant, and 4 = very pleasant]),  $F(1, 397) = 6.76$ ,  $p = .01$ ,  $\eta^2_p = .017$  (within-subjects). Both conditions used *look for* on the filler trials.

pleasantness of the images but did not perceive any influence in particular. This is consistent with the idea that people may consciously notice a stimulus, but not notice any influence occurring and thus be unlikely to correct for it (Payne et al., 2005; Winkielman, Zajonc, & Schwarz, 1997).

Participants in the “specific source attribution” condition were made explicitly aware of the influence. Rather than rating the images, they instead read, “Please be aware that the previous task might have affected your mood and your mood can affect the way you respond to the next part of the survey (which asks about your motivation for doing things).” Note that we referred to “mood” rather than “affect” only to ensure that participants understood the gist of the influence. The main objective was to inform participants of the affective influence of the images.

Participants then rated their perceptions of the intrinsic rewards of learning. Five items assessed intrinsic rewards, “I like to learn,” “I think it is interesting to learn,” “Learning is a pleasant activity,” “When I know I have learned something new, I feel good inside,” and “I like to learn things for reasons beyond my understanding” (rated on 7-point unnumbered scales with endpoints “Not at all” and “Extremely,”  $\alpha = .87$ ). On the next screen, five additional items assessed various extrinsic rewards—separable outcomes that could alternatively explain why someone would report a liking for learning: “I like learning because it helps me get along with others better,” “I try to learn different things in order to better obtain tangible rewards (e.g., nourishment, success, belonging),” “Learning helps me achieve outcomes I would otherwise not be able to achieve,” “I learn things in order to look smarter or to avoid being seen as unintelligent,” “I learn for my own good,” ( $\alpha = .64$ ).

We also assessed competence-based attributions to distinguish the misattribution effect from beliefs about satisfying a psychological need. One item assessed competence experiences that stem from overcoming effortful challenges (“I enjoy class assignments that are challenging”) and the other was meant to assess self-perceptions of intellectual ability (“I am good at trivia games because I have a natural tendency to pick up on lots of ‘useless’ information”); these two items were uncorrelated with each other ( $r = .08, p = .21$ ). All the measures were positively correlated with intrinsic motivation: extrinsic attributions ( $r = .31, p < .001$ ), beliefs about talent ( $r = .20, p = .001$ ), and in particular, enjoyment of challenges ( $r = .62, p < .001$ ). The high correlation between enjoyment of challenges and perceptions of intrinsic reward is consistent with the idea that the psychological need for competence explains a large proportion of variance in people’s beliefs about the intrinsic rewards of learning.

## RESULTS AND DISCUSSION

A 2 (context condition: positive vs. neutral)  $\times$  3 (attribution: none vs. reminder of images vs. specific source attribution) analysis of covariance (ANCOVA) was conducted on participants’ attributions of intrinsic reward. We controlled for the extrinsic attributions to account for any baseline liking that could be attributed to separable outcomes for engaging in learning, with the aim to isolate the effect to



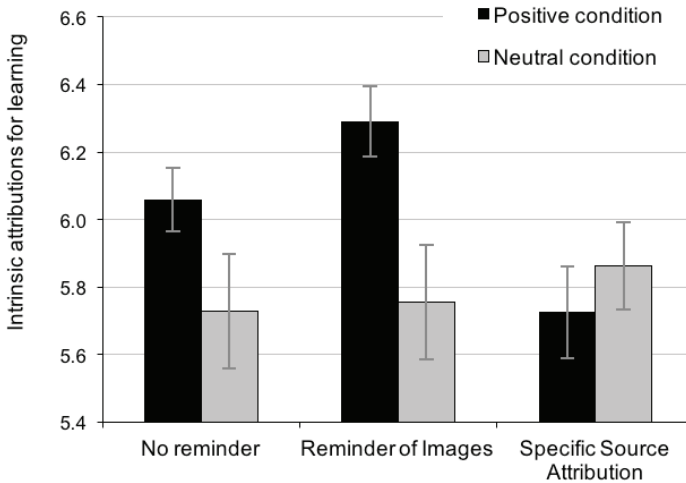


FIGURE 1. Perceived intrinsic rewards of learning, as a function of context condition and attribution condition. Error bars represent standard errors.

intrinsic rewards. Results indicated a main effect of context condition,  $F(1, 273) = 3.88, p = .050, \eta_p^2 = .014$ , qualified by a two-way interaction of context condition and attribution condition,  $F(2, 273) = 3.46, p = .033, \eta_p^2 = .025$ .<sup>2</sup> As illustrated in Figure 1, the positive condition increased attributions of intrinsic reward in the “no reminder condition” ( $M = 6.06$  vs.  $5.76, p = .092$ ), and in the “reminder of images” condition ( $M = 6.29$  vs.  $5.76, p < .001$ ); these two conditions did not statistically differ from one another ( $F < 1$ ). However, the effect was extinguished in the “specific source attribution” condition ( $M = 5.72$  vs.  $5.86, p > .25$ ); participants showed no misattribution of intrinsic reward when they could make a proper source attribution.

Subsequent discriminant validity analyses indicated no effects of the manipulations on the extrinsic or competence-based attributions ( $F_s < 2$ ). The only notable effect was that, when additionally controlling for the competence-based attributions, the marginal simple effect of the positive condition, when there was no reminder, reached conventional levels of significance (from  $p = .092$  to  $p = .037$ ). The subtle external rewards only affected beliefs about intrinsic rewards, independently of the extrinsic and competence-based attributions.

In sum, the subtle external rewards increased participants’ beliefs about the intrinsic pleasures of learning. In addition, the attribution manipulation helps to fit our model to a misattribution framework, insofar that the effect of the subtle external rewards was extinguished when participants could make a proper source attribution. Interestingly, a mere reminder of the images was not enough to extinguish the misattribution effect, which is consistent with the idea that people can consciously notice a stimulus, but not notice any influence occurring and thus be unable to correct for it (Payne et al., 2005; Winkielman et al., 1997). The misattribu-

2. Excluding the covariate only slightly altered the main effect for the context cue condition,  $F(2, 274) = 4.23, p = .041, \eta_p^2 = .015$ , and two-way interaction of context condition and attribution condition,  $F(2, 274) = 2.97, p = .055, \eta_p^2 = .021$ .

tion of intrinsic rewards was contingent on a lack of awareness that the context cues were having an effect.

## STUDY 2

For affect misattribution to adequately explain some perceptions of intrinsic reward, we must also fit our model to an intrinsic motivation framework. One way to do this is to demonstrate that the effect can be extinguished by introducing a blatant external reward contingency (Deci, Koestner, & Ryan, 1999; Eisenberger & Cameron, 1996). In Study 2, we added a learning task and manipulated whether or not participants expected payment for their performance on it. Befitting an intrinsic motivation framework, any effect of the subtle external rewards on perceived intrinsic rewards should be extinguished by the blatant, external, reward contingency.

Through adding the learning task, we can also explore how the misattribution of intrinsic reward connects to one's behavior on the learning task. The learning task can be another contingency that could alter one's subjective experience with learning, in that a positive task experience may be essential to justify any attribution of intrinsic reward.

## METHOD

*Participants.* One hundred sixty-seven American adults (64 female, Age  $M = 30.39$ ) completed the study via Amazon's Mechanical Turk.

*Procedure.* This study used a 2 (context condition: positive vs. neutral)  $\times$  2 (reward contingency: no reward vs. pay-for-performance) factorial design. Participants first completed the word-search paradigm from Study 1 wherein they were exposed to either positive or neutral context cues. Next, participants were presented with a learning task that either included a blatant reward contingency or not. In the "no reward" condition, participants read: "The next part of this study involves reading short lessons on U.S. history and then answering questions about those lessons. This is an opportunity to pick up new knowledge: For each history lesson you read, you will subsequently be given a set of questions to answer about it; the more lessons you complete, the greater your knowledge acquisition. There is no penalty for wrong answers; more correct answers simply means that you are acquiring more knowledge." In the "pay-for-performance" condition, the paragraph was changed to read "...This is an opportunity to pick up extra Amazon credit... your total number of correct answers will determine your bonus payment (this will be in addition to what you are already being paid to participate). There is no penalty for wrong answers; more correct answers simply means you are picking up more Amazon credit (up to an additional 25 cents)."

Participants then completed the learning task. They were first given a dense passage of a few hundred words detailing some pivotal moment in U.S. history. On the next screen were multiple-choice questions about the passage, after which participants could choose to either go to the next lesson or skip the remaining lessons.

Learning task performance was assessed via the number of correct responses they accrued. On average, participants read 3.69 passages and accrued 11.01 correct answers. No participants completed all the history lessons, which means everyone eventually exhausted their motivation for the task and skipped through it.

We assessed attributions of intrinsic reward in two different ways. After completing the history task, participants reported their attributions of intrinsic reward ( $\alpha = .86$ ), extrinsic reward ( $\alpha = .71$ ), and the two competence-based attributions for learning as per Study 1. However, given that participants in this study had just engaged in a learning task, we were concerned their subjective experience of reward would be contingent on their task performance. We, therefore, also assessed participants' retrospective attributions for participating in online studies generally. This measure might be cleaner because it can assess the idea that contextual positive cues paired with an activity (i.e., participating in an online study) were indeed misattributed to that activity. A broader, retrospective measure may also not be as contingent on performance: in a classic study of retrospective misattribution (Kruglanski et al., 1972), elementary school students who received (or simply saw their peers receive) an unexpected reward for winning games later misattributed their own initial participation to that reward (instead of interest in the games or liking for competition). We accordingly assessed participants' retrospective attributions for participating in online studies, which past work suggests can include intrinsic reasons (entertainment), extrinsic reasons (earning additional money), and unspecified reasons ("killing time"; see Paolacci, Chandler, & Ipeirotis, 2010). Participants read, "*I complete HITs [jobs on Amazon's Mechanical Turk] because...*" Two items assessed intrinsic attributions ("*...it is fun,*" "*...I just like doing it,*"  $r = .73, p < .001$ ), two assessed extrinsic attributions in terms of personal outcomes ("*...it is a way to make money,*" "*...it is a way to earn extra money/credit for personal purchases,*"  $r = .62, p < .001$ ), and two assessed extrinsic attributions in terms of prosocial outcomes ("*...it is a way to help people who need it,*" "*...it is a way to do something useful and, perhaps, important,*"  $r = .75, p < .001$ ). We included "*it is a way to pass the time*" as a lone filler item. Participants gave their responses on 7-point, unnumbered scales with endpoints labeled "*Not at all*" and "*Extremely*." Participants were debriefed at the end of the study.

## RESULTS AND DISCUSSION

We present the results of the retrospective attributions first because they were relatively straightforward: We conducted a 2 (context condition: positive vs. neutral)  $\times$  2 (reward contingency: no reward vs. pay-for-performance) analysis of covariance (ANCOVA) on participants' retrospective attributions of intrinsic reward, controlling for the extrinsic attributions. Results indicated a two-way interaction of context condition and reward contingency condition,  $F(1, 161) = 4.18, p = .042, \eta^2_p = .025$ , and no main effects ( $F_s < 1$ ).<sup>3</sup> As illustrated in Figure 2, participants

3. Excluding the covariates only slightly altered the two-way interaction,  $F(1, 163) = 3.86, p = .051, \eta^2_p = .023$ .

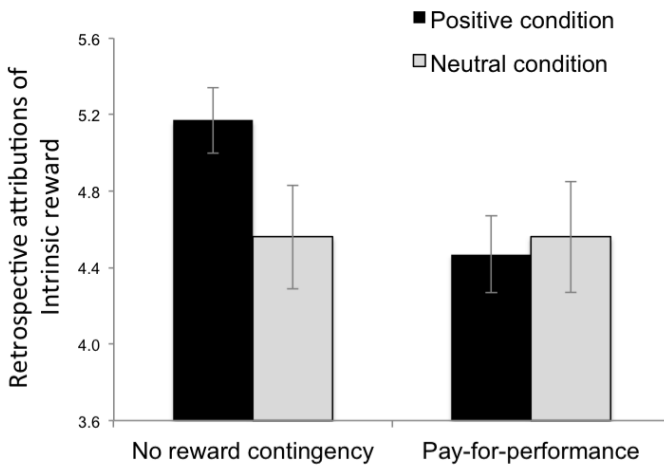


FIGURE 2. Retrospective attribution of intrinsic motivation to participate in Mechanical Turk studies, as a function of context condition and learning task reward contingency. Error bars represent standard errors.

in the positive condition only reported higher perceptions of intrinsic reward in the *absence* of the pay-for-performance reward contingency: they reported higher intrinsic rewards than their counterparts in the neutral condition ( $M = 5.17$  vs.  $M = 4.57$ ,  $p = .014$ ), and than others in the positive condition who received the pay-for-performance reward contingency (vs.  $M = 4.44$ ,  $p = .002$ ). In other words, the increased perceptions of intrinsic reward were extinguished by the blatant external reward. There were no single or interactive effects on either of the two extrinsic attributions ( $F_s < 1$ ). This result helps to fit the effect to an intrinsic motivation framework.

The attributions of intrinsic rewards for learning followed a similar overall pattern, but were additionally contingent on idiosyncratic differences in participants' learning task behavior. First note there were no single or interactive effects of the context condition on task performance ( $F_s < 2$ ), only a main effect of the reward contingency (pay-for-performance:  $M = 13.44$ ,  $SD = 9.09$ , vs. no reward contingency:  $M = 8.10$ ,  $SD = 7.44$ ),  $F(1, 163) = 17.37$ ,  $p < .001$  ( $B = -.29$ , 95% CI:  $-0.43, -0.15$ ). The positive context cues did not noticeably affect behavior on the learning task. However, participants' behavior on the learning task did affect their misattributions of intrinsic reward: A regression analysis predicted participants' attributions of intrinsic rewards for learning from their context condition (positive = 1, neutral = -1), task reward (no reward = 1, pay-for-performance = -1), performance on the history task (standardized), and all possible interactions, controlling for attributions of extrinsic reward ( $r = .40$ ,  $p < .001$ ). Results indicated a positive direct effect of task performance,  $B = .20$  (95% CI:  $0.06, 0.36$ ),  $F(1, 158) = 7.45$ ,  $p = .007$ , a positive interaction of (no-) reward and task performance,  $B = .16$  (95% CI:  $0.01, .31$ ),  $F(1,$

158) = 4.36,  $p = .038$ , and a positive three-way interaction of condition, (no-) reward, and task performance,  $B = .16$  (95% CI: 0.01, 0.31),  $F(1, 158) = 4.62$ ,  $p = .033$ .<sup>4,5</sup>

Note that the three-way interaction is not very strong, but tests of the simple two-way interactions, conducted at  $+1/-1$  SD mean task performance, at least suggest a positive interaction of the positive context cues and no-reward contingency only among those who performed well,  $B = .21$  (95% CI: -0.003, .43),  $F(1, 158) = 3.78$ ,  $p = .054$ , and not among those who performed poorly,  $B = -.12$  (95% CI: -.31, .08),  $F(1, 158) = 1.44$ ,  $p = .233$ . The general pattern of the data, as illustrated in Figure 3, is consistent with the idea that attributions of intrinsic reward were only higher if those participants, who received the positive context cues and no-reward contingency, also performed well.

To summarize, the effect of the subtle external rewards on the retrospective attributions of intrinsic reward mimicked a classic pattern of intrinsic motivation, insofar as the effect was extinguished by the pay-for-performance reward contingency. The effect on perceived intrinsic rewards for learning followed a similar pattern, but was additionally contingent on participants' learning task behavior. We could speculate that poor performance on the learning task extinguished the incidental positive affect; however, given that the other dependent measure—the retrospective misattribution—was unaffected, poor performance on the learning task may have simply undermined the plausibility that learning was the source of positive affect. Critically, if we set aside the difference between the two dependent measures, both showed the same basic pattern: that the increased perceptions of intrinsic reward were contingent on the presence of subtle external rewards, suggesting they were products of misattributed causal inference.

## GENERAL DISCUSSION

The notion of the “true self” was meant to describe a sense of self that is based on spontaneous authentic experience (Winnicott, 1965). Yet, we question whether people are always able to distinguish authentic experiences from those that are merely misattributed responses to external stimuli. Accordingly, we propose there are at least two pathways through which perceptions of intrinsic reward can arise: first is through the satisfaction of psychological needs, which has been extensively covered by self-determination theorists as well as others. Second is a hypothetical “misattribution” pathway: wherein there is some *external* source of positive affect that one does not distinguish from the activity itself.

4. Excluding the covariate significantly altered the three-way interaction: There was only a direct effect of task performance,  $B = .22$  (95% CI: 0.06, 0.39),  $F(1, 159) = 7.10$ ,  $p = .009$ , and no other effects ( $F_s < 2.34$ ,  $p_s > .12$ ). It makes sense that the covariate contributed to the significance of the three-way interaction—it was meant to account for any baseline liking due to separable outcomes for engaging in the activity, such as the pay-for-performance reward manipulation.

5. We observed the same three-way interaction pattern for one of the competence items “I enjoy class assignments that are challenging,”  $B = .30$  (95% CI: 0.09, 0.51),  $F(1, 158) = 7.95$ ,  $p = .005$ . One could speculate that high performance on the learning task provided feedback to one's competence and the positive affect facilitated beliefs about enjoyment.

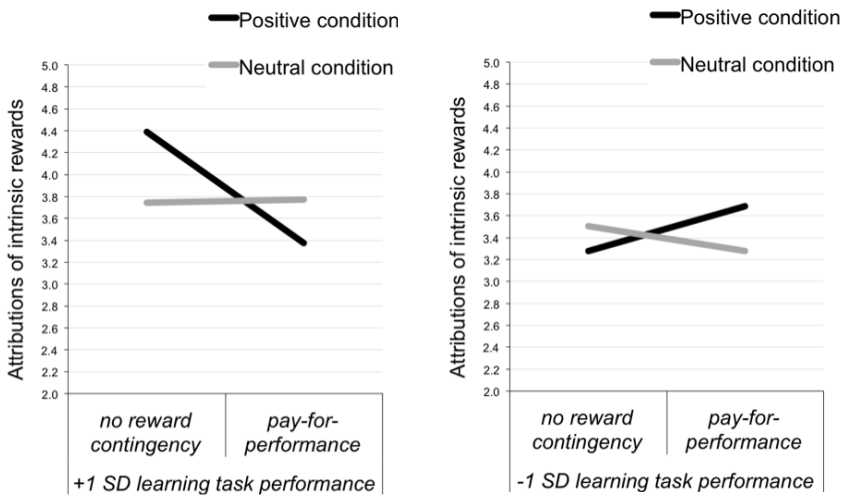


FIGURE 3. Perceived intrinsic rewards of learning, as a function of context condition, reward contingency, and performance on the learning task. Values are predicted from regression equation.

The two pilot studies presented here suggest some beliefs about intrinsic reward may indeed stem from the misattribution of external rewards. However, these misattributed beliefs appear to be contingent on several factors: they are extinguished (i) when the individual is led to make a proper source attribution for the influence, (ii) when the activity in question becomes associated with an external reward contingency, and (iii) when one's subsequent experience with an activity is poor enough to undermine its plausibility as a source of reward. Misattributed beliefs about intrinsic motivation seem to depend on a coherent subjective experience, which could be what differentiates them from authentic experiences of intrinsic reward that stem from the satisfaction of psychological needs.

## IMPLICATIONS FOR INTRINSIC MOTIVATION

An affect misattribution pathway may help to explain some of the more elusive properties of intrinsic motivation—particularly its occasional instability across situations. Indeed, intrinsic motivation sometimes only persists as long as the salience of external rewards is minimized (Ross, 1975; White & Kight, 1984). The seeming instability of intrinsic motivation has been the subject of much debate, but perhaps it signals the presence of a misattributed belief about intrinsic reward as opposed to the presence of authentic intrinsic motivation. If misattributed beliefs about intrinsic reward depend on a coherent subjective experience, it is easy to imagine how an incidental association between an activity and a specific subjective experience would be unstable: if the subjective experience changes (or one's attributions for the experience change), the perception of intrinsic rewards could also change or disappear entirely. Authentic forms of intrinsic motivation are pre-

sumably more robust—a true labor of love should be intrinsically rewarding even when fraught with external rewards, punishments, and situational changes.

This work may, accordingly, offer a step toward unpacking the many effects that together get labeled as “intrinsic” reward. When is a reward authentically *intrinsic*? One could argue it is only intrinsic when it serves a basic (psychological) need or motivation—such as when behavioral engagement in the activity satisfies basic needs for competence or autonomy, or when the structure of a task requires a level of eagerness or vigilance that fits one’s chronic approach-avoidance tendencies (Higgins, 2005). Yet the present work suggests a reward can be perceived to be “intrinsic” when it is incidentally associated with an activity and not because it addresses a particular need or motive. Actual intrinsic motivation might only come later, such as when one tries to re-experience the positive affect and thus establish control over its occurrence. Indeed, some consider human beings to be inherently motivated to establish control over outcomes (White, 1959); as an example, Eitam, Kennedy, and Higgins (2013) demonstrated that establishing control over trivial outcomes can be motivating, even if it does not achieve any tangible outcome *per se*. If an initial experience of positive affect triggers control motivation, it might lead to (authentic) intrinsic rewards through one’s efforts to re-experience the outcome.

Affect misattribution could even be an authentic pathway through which extrinsically motivated behavior becomes intrinsically motivated over time. The valence and quality of affect associated with a goal or desired outcome can often get transferred to the means to its attainment (Fishbach, Shah, & Kruglanski, 2004). This suggests misattributed positive affect can eventually become chronically associated with the activity. A question is whether the association—and reward experience—would be self-sustaining if the true source of reward were later removed. Indeed, even if a person is conditioned to associate his or her mental representation of an activity with positive affect—such that subsequent thoughts about the activity co-activate the positive affect—it remains unclear whether the conditioned reward experience would ever be fully intrinsic.

## IMPLICATIONS FOR THE PHENOMENAL SELF AND NEED FULFILLMENT

If a distinction is to be made between authentic and misattributed experiences of intrinsic reward, it raises questions for how misattributed intrinsic rewards contribute to need fulfillment and the phenomenal self. The illusion of intrinsic rewards could be misleading—it could motivate the pursuit of unattainable end states, or inflate expectations about the quality of affect one will experience from an activity. Some pursuits may not really suit one’s regulatory orientations and internalizing them can undermine psychosocial functioning (Baumann, Kaschel, &

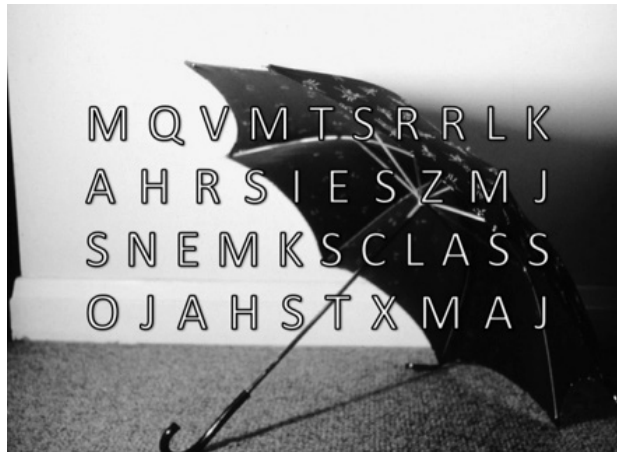
Kuhl, 2005). Put simply, sometimes we fall more in love with the idea of a pursuit than its actual pursuit because we have illusory beliefs about its reward potential. That said, illusory beliefs are not valued simply because they are perceived to be true, but also because they are functional (Preston & Epley, 2005). As with misattributions of free will and causality—or even the existence of God (Kay, Gaucher, Napier, Callan, & Laurin, 2008)—misattributions of intrinsic reward may fulfill certain needs. For example, being able to infer meaning, from an incidental activity-reward association, could help to foster a coherent and willful sense of self. The misattribution of intrinsic reward turns incidental associations into personal likes and dislikes; it also makes a person's intrinsic interests indicative of a purposive entity rather than a byproduct of random associations in memory.

A final possibility that is also worth briefly considering is that the process of misattribution may itself be a pathway to need satisfaction of sorts: that is, people may have certain epistemic needs and the activation of unexplained subjective states may necessitate an explanation due to the explanatory vacuum they create. The mere presence of an unexplained phenomenological state implies a lack of (conscious) control over one's internal experiences, which could conceivably trigger motivation to establish control over it. In other words, one could be motivated to control not just the occurrence of the reward experience, but also the explanation for it. Landau, Kay, and Whitson (2015; see also Kay, Whitson, Gaucher, & Galinsky, 2009) proposed one strategy people use to restore psychological control is to seek simple, clear, and consistent explanations in the world—in other words, to imbue the world with structure. Perhaps the same logic applies to how people explain unexplained subjective experiences: they form a simple, clear representation of the self that fits the experience and offers a causal mechanism. If the subjective state is pleasant, the intuitively appealing interpretation is that one likes and enjoys whatever one is doing at the time. Such an interpretation is not only logically coherent, but it also makes the "self" the controlling agent of the experience rather than assigning control to external stimuli or non-volitional sensory systems. In other words, the misattribution of intrinsic reward could provide a sense that the world is controllable, structured, and predictable rather than random and chaotic.

To conclude, perceptions of intrinsic reward may result, in part, from external rewards that are forgotten or not noticed as rewards to begin with. When rewards are applied subtly, the most obvious source of pleasant affect is the activity itself. Thus, the "intrinsic" in intrinsic reward may not always describe the source of a motivation, but the subjective experience of it.



## APPENDIX



Samples of positive (fireworks, top) and corresponding neutral (umbrella, below) word search matrices. Participants in this specific example were instructed to search for a 5-letter word (e.g., “CLASS,” line 3). Images were presented in full color.

## REFERENCES

- Bar-Anan, Y., Wilson, T. D., & Hassin, R. R. (2010). Inaccurate self-knowledge formation as a result of automatic behavior. *Journal of Experimental Social Psychology, 46*, 884-894. doi: 10.1016/j.jesp.2010.07.007
- Baumann, N., Kaschel, R., & Kuhl, J. (2005). Striving for unwanted goals: Stress-dependent discrepancies between explicit and implicit achievement motives reduce subjective well-being and increase psychosomatic symptoms. *Journal of Personality and Social Psychology, 89*(5), 781-799. doi: 10.1037/0022-3514.89.5.781
- Cantor, J., Zillmann, D., & Bryant, J. (1975). Enhancement of experienced sexual arousal in response to erotic stimuli through misattribution of unrelated residual excitation. *Journal of Personality and So-*

- cial Psychology*, 32, 69-75. doi: 10.1037/h0076784
- Chartrand, T. L., Cheng, C. M., Dalton, A. N., & Tesser, A. (2010). Nonconscious goal pursuit: Isolated incidents or adaptive self-regulatory tool? *Social Cognition*, 28(5), 569-588. doi: 10.1521/soco.2010.28.5.569
- Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology*, 89(2), 129-142. doi: 10.1037/0022-3514.89.2.129
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627-668. doi: 10.1037/0033-2909.125.6.627
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. doi: 10.1207/S15327965PLI1104\_01
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853-869. doi: 10.1037/0033-2909.127.6.853
- Eisenberger, R., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist*, 51, 1153-1166. doi: 10.1037/0003-066X.51.11.1153
- Eitam, B., Kennedy, P. M., & Higgins, E. T. (2013). Motivation from control. *Experimental Brain Research*, 229, 475-484. doi: 10.1007/s00221-012-3370-7
- Fishbach, A., Shah, J. Y., & Kruglanski, A. W. (2004). Emotional transfer in goal systems. *Journal of Experimental Social Psychology*, 40(6), 723-738. doi: 10.1016/j.jesp.2004.04.001
- Higgins, E. T. (2005). Value from regulatory fit. *Current Directions in Psychological Science*, 14(4), 209-213. doi: 10.1111/j.0963-7214.2005.00366.x
- Kay, A. C., Gaucher, D., Napier, J. L., Callan, M. J., & Laurin, K. (2008). God and the government: Testing a compensatory control mechanism for the support of external systems. *Journal of Personality and Social Psychology*, 95(1), 18-35. doi: 10.1037/0022-3514.95.1.18
- Kay, A. C., Whitson, J. A., Gaucher, D., & Galinsky, A. D. (2009). Compensatory control: Achieving order through the mind, our institutions, and the heavens. *Current Directions in Psychological Science*, 18(5), 264-268. doi: 10.1111/j.1467-8721.2009.01649.x
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 28(2), 107-128. doi: 10.1037/h0034225
- Kruglanski, A. W. (1975). The endogenous-exogenous partition in attribution theory. *Psychological Review*, 82(6), 387-406. doi: 10.1037/0033-295X.82.6.387
- Kruglanski, A. W., Alon, S., & Lewis, T. (1972). Retrospective misattribution and task enjoyment. *Journal of Experimental Social Psychology*, 8, 493-501. doi: 10.1016/0022-1031(72)90075-3
- Landau, M. J., Kay, A. C., & Whitson, J. A. (2015). Compensatory control and the appeal of a structured world. *Psychological Bulletin*, 141(3), 694-722. doi: 10.1037/a0038703
- Lang, P. J., Bradley, M. M., & Cuthbert, B. (1995). *International affective picture system*. Gainesville, FL: University of Florida, Center for Research in Psychophysiology.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3), 261-273. doi: 10.1111/j.1469-8986.1993.tb03352.x
- Leander, N. P., Moore, S. G., & Chartrand, T. L. (2009). Mystery moods: Their origins and consequences. In G. Moskowitz & H. Grant (Eds.), *The psychology of goals* (pp. 480-504). New York: Guilford.
- Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic rewards: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, 28, 129-137. doi: 10.1037/h0035519
- Loersch, C., & Payne, B. K. (2011). The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation. *Perspectives on Psychological Science*, 6, 234-252. doi: 10.1177/1745691611406921
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal

- stimulus exposures. *Journal of Personality and Social Psychology*, 64(5), 723-739. doi: 10.1037/0022-3514.64.5.723
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259. doi: 10.1037/0033-295X.84.3.231
- Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.
- Parks-Stamm, E. J., Oettingen, G., & Gollwitzer, P. M. (2010). Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving. *Journal of Experimental Social Psychology*, 46, 531-542. doi: 10.1016/j.jesp.2010.02.004
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89(3), 277-293. doi: 10.1037/0022-3514.89.3.277
- Preston, J., & Epley, N. (2005). Explanations versus applications: The explanatory power of valuable beliefs. *Psychological Science*, 16(10), 826-832.
- Ross, M. (1975). Salience of reward and intrinsic motivation. *Journal of Personality and Social Psychology*, 32, 245-254. doi: 10.1037/0022-3514.32.2.245
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54-67.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45(3), 513-523. doi: 10.1037/0022-3514.45.3.513
- Shah, J. Y., & Kruglanski, A. W. (2000). The structure and substance of intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance*. (pp. 105-127). San Diego, CA: Academic Press.
- White, G. L., & Kight, T. D. (1984). Misattribution of arousal and attraction: Effects of salience of explanations for arousal. *Journal of Experimental Social Psychology*, 20, 55-64. doi: 10.1016/0022-1031(84)90012-X
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66(5), 297-333. doi: 10.1037/h0040934
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and emotions. *Psychological Bulletin*, 116, 117-142. doi: 10.1037/0033-2909.116.1.117
- Winkielman, P., Zajonc, R. B., & Schwarz, N. (1997). Subliminal affective priming resists attributional interventions. *Cognition and Emotion*, 11(4), 433-465. doi: 10.1080/026999397379872
- Winnicott, D. W. (1965). Ego distortion in terms of true and false self. *The maturational process and the facilitating environment: Studies in the theory of emotional development* (pp. 140-157). New York: International Universities Press.