The Convergent Validity and Nomological Net of Two Methods to Measure Retroactive Influences

Article *in* Psychology of Consciousness: Theory, Research, and Practice · May 2018
DOI: 10.1037/cns0000149

1 author:

Matt C. Howard
University of South Alabama
24 PUBLICATIONS **83** CITATIONS

SEE PROFILE

Psychology of Consciousness: Theory, Research, and Practice

The Convergent Validity and Nomological Net of Two Methods to Measure Retroactive Influences

Matt C. Howard

Online First Publication, May 3, 2018. http://dx.doi.org/10.1037/cns0000149

Psychology of Consciousness: Theory, Research, and Practice

The Convergent Validity and Nomological Net of Two Methods to Measure Retroactive Influences

Matt C. Howard

Online First Publication, May 3, 2018. http://dx.doi.org/10.1037/cns0000149

Psychology of Consciousness: Theory, Research, and Practice

2326-5523/18/$12.00 http://dx.doi.org/10.1037/cns0000149

The Convergent Validity and Nomological Net of Two Methods to Measure Retroactive Influences

Matt C. Howard
University of South Alabama

Two notable criticisms have been recently raised against research on retroactive influences (RIs). First, researchers have rarely tested the measurement properties of their methods to gauge RIs, and it is still relatively unknown whether any commonly used method is adequate for gauging RIs. Second, RI research needs more close replications in addition to conceptual replications, in which all aspects of the original study are reproduced. The current article investigates both of these criticisms. The current article explores the convergent validity and nomological network of two methods to measure RIs, and it also provides three studies to test these effects: an initial investigation, a close replication, and a conceptual replication. The relationship of the two methods did not meet even minimal requirements for convergent validity in all three studies, failing to provide support for their ability to measure RIs. The overall sample-size weighted correlation was .03 ($\bar{r}$, 95% CI [−.07, .12]). Also, the two measures did not share any significant relationships with any other variables, suggesting that they do not share a common nomological net. Together, the current article emphasizes the need for future research to further test the measurement of RIs, as researchers of RI cannot convincingly claim that their results are valid without first supporting their measurement methods. The current article also highlights the importance of valid measurement across all domains of psychology and beyond. Additional implications and directions for future research are discussed.

Keywords: retroactive influences, anomalous experiences, measurement, convergent validity, nomological net

Psychologists have long been in interested in the concept of psi (James, 1890, 1902; Jung, 1917, 1958; Meehl & Scriven, 1956), which is "the anomalous process of information or energy transfer that are currently unexplained" (Bem, 2011, p. 407). While several phenomena are included within the label of psi, such as clairvoyance and psychokinesis, two phenomena that have garnered increased recent attention are precognition and premonition (Caswell, Hunter, & Tessaro, 2014; Galak, Leboeuf, Nelson, & Simmons, 2012; Gaona, Colinas, Rouleau, Tessaro, & Caswell, 2014; Parra, 2013; Ritchie, Wiseman, & French, 2012; Traxler, Foss, Podali, & Zirnstein, 2012). Precognition is a conscious cognitive awareness of an event that has yet to occur, whereas premonition is the affective apprehensive of an event that has yet to occur (Rattet & Bursik, 2001). Both of these phenomena involve the "anomalous retroactive influence of some future event on an individual's current responses" (Bem, 2011, p. 407), and are often thought of as effects that are "time-reversed" (Bem, 2011, p. 407). That is, the observed causal effect flows backward in time, such that thoughts or emotions are influenced by an event prior to it occurring rather than after.

Skeptics have always been apprehensive toward the study of precognition and premonition (Diaconis, 1978; Pratt & Price, 1938; Price, 1955; Ritchie et al., 2012), henceforth called RIs, but the topic has endured a recent flood of attention after the publication of Bem (2011). Bem found support in eight of nine experiments to test for the RI of several well-established psychological effects. Further yet, Bem also found repeated significant relationships between RIs

Correspondence concerning this article should be addressed to Matt C. Howard, Marketing and Quantitative Methods, The University of South Alabama, 337 Mitchell College of Business, 5811 USA S Dr, Mobile, AL 36608. E-mail: mhoward@southalabama.edu

and sensation seeking, suggesting that RIs not only occur—but they relate to measurable antecedents. Responses to Bem were immediate, appearing on the very next page of the journal issue (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011).

Most of these responses focused on statistical and methodological issues of the social sciences (Brandt et al., 2014; Rouder, Morey, Speckman, & Province, 2012; Schimmack, 2012), such as the reliance on null hypothesis significance testing (Wagenmakers et al., 2011; Wetzels et al., 2011), but two notable criticisms were solely focused on psi research. First, LeBel and Peters (2011) noted that Bem (2011) used nine different methods to measure RIs, but it would have been more impressive if Bem supported that a single method measured RIs exceptionally well. To our knowledge, no one has yet to provide this support for any method to measure RIs. Second, several subsequent authors argued that RI research needs more close replications in addition to conceptual replications (Brandt et al., 2014; LeBel & Peters, 2011; Schimmack, 2012; Wetzels et al., 2011). Close replications recreate most all aspects of the original study, but perhaps use a different sample and/or expand on the original by applying more measures. Conceptual replications are meant to test the same theoretical effect, but may use a different measure or manipulation to do so. The current article investigates both of these criticisms.

The current article explores the convergent validity and nomological network of two methods to measure RIs, both of which are administered online and ask participants to predict words or pictures that will randomly appear on the following page. Convergent validity refers to the interrelationship(s) of methods to measure the same or similar constructs, whereas a nomological network is the representation of theoretical constructs (latent variables), their manifestations (observed variables), and the interrelationships among and between these (Cronbach & Meehl, 1955; Trochim, 2006). These pieces of information can support (or not support) whether the two studied measurement methods are appropriate for gauging RIs. Also, the current article satisfies prior calls by presenting three studies to test these effects: an initial investigation, a close replication, and a conceptual replication. Thus, the current article allows for a better understanding of the measurement of RIs while adhering to the methodological recommendations of recent authors.

Lastly, one additional question should be addressed before continuing: *why* should RIs be studied? Researchers on the topic have been ridiculed, directly insulted, and even pitied for wasting their time (Galak et al., 2012; Rouder et al., 2012; Wetzels et al., 2011). Findings on the topic can be considered mixed at best, and it is impossible to point to a series of studies that claim decisive support for the phenomenon. Conversely, a series of studies has yet to clearly contradict the phenomenon. Further research on the existence of RIs can bring more closure to this issue. If the existence of RI is supported, the results could provide extreme benefits for research and several applications could benefit from even the slightest manipulation of RIs. If support is *not* found for RIs, then parapsychology researchers could devote their time to more fruitful topics. Thus, the current article does not hold any assumptions about the existence of RIs, but it can provide support for either side of the RI argument—both for and against.

Background

Measurement of RIs

Bem (2011) presented nine experiments using different methodologies to test the effects of RI, each with particular advantages and disadvantages. While Bem was unique in many regards, the methodologies of the studies were almost entirely adapted from prior research. In fact, Bem's nine experiments could be considered representative of the majority of prior methodologies to study RIs (Honorton & Ferrari, 1989; Luke, Roe, & Davison, 2008; Morris, 1982; Parker & Sjödén, 2010; Steinkamp, 2000, 2001; Storm, Tressoldi, & Di Risio, 2012), and they are certainly representative of the methodologies afterward (Bem, Tressoldi, Rabeyron, & Duggan, 2015; Galak et al., 2012; Luke & Zychowicz, 2014; Ritchie et al., 2012; Traxler et al., 2012). For this reason, we review the methodologies separated by the four dominant types as differentiated by Bem, followed by the description of a fifth type that was not applied by Bem but nevertheless popular in prior research (Bem et al., 2015; Honorton & Ferrari, 1989; Storm et al., 2012). It should be

noted, however, that the boundaries between these methods are not entirely firm. That is, some methods to gauge RI are representative of multiple types as differentiated by Bem, but this does not mean that this framework cannot appropriately categorize RI methods.

The first method is to present participants with concealed options, and the participants are asked to select the correct choice (Bem, 2011, Studies 1 and 2; Luke & Zychowicz, 2014; Maier et al., 2014; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). For example, the participant may be shown curtains or pictures with similar arousal levels and valences. In the case of curtains, the participant is then asked to select the curtain with the picture behind it, and they are shown the result of their choice. Most often, the picture is erotic or positively valenced, suggesting that correct participants could foresee the positive results from their choice. In the case of similar pictures, the participant is asked to click on the correct or preferred picture. An erotic or positively valenced picture is then presented if the correct option is chosen, whereas a negatively valenced picture is presented if the incorrect option is chosen. Again, correct participants are believed to have foreseen the positive results from their choice. If a RI effect exists, then participants should select the correct curtain or picture more often than random chance alone.

A second method is to retroactively prime participants (Bem, 2011, Studies 3 and 4; Bierman & Bijl, 2014; Broderick & Goertzel, 2014). Participants are presented a series of pictures, and asked to press a certain key corresponding to whether the picture is positive or negative. Afterward, participants are briefly shown a positively or negatively valenced word or picture, most often subliminally, with the assumption that the presented word could affect the prior categorization. If an RI effect exists, then participants should take longer to categorize pictures when the valence of the picture differs from the word presented afterward.

A third method is to retroactively habituate participants (Bem, 2011, Studies 5, 6, and 7). Participants are presented similarly valenced pictures, and they are asked to select the picture that they like the most. Afterward, they are presented one of the pictures, most often subliminally, with the assumption that the picture could influence the participants' prior preference. A "hit" is when the participant selected the picture shown immediately afterward, and participants should select the hit more than random chance alone if an RI effect exists.[1]

A fourth method is to retroactively facilitate recall (Bem, 2011, Studies 8 and 9; Barušs & Rabier, 2014; Galak et al., 2012; Ritchie et al., 2012; Robinson, 2011; Subbotsky, 2013; Traxler et al., 2012; Vernon, 2015). Participants are presented a list of words and then asked to recall all words that they can remember immediately after. After the free recall test, participants are asked to practice selected words from the worst list, most often through retyping the words, with the assumption that practice afterward can influence prior performance. If an RI effect exists, participants should recall the practice words more often than the others.

A fifth and final method is to directly ask participants to predict the immediate future (Ertel, 2005a, 2005b, 2010; Honorton & Ferrari, 1989; Storm, Ertel, & Rock, 2013a, 2013b; Storm et al., 2012). Participants may be asked to predict the next card drawn from a deck, the next roll of a dice, or the next number shown on a randomly selected ping pong ball. If an RI effect exists, participants should predict the result more so than random chance alone.

While each of these methods test RI in a similar manner, prior authors have expressed different concerns for each (Brandt et al., 2014; Galak et al., 2012; LeBel & Peters, 2011; Rouder et al., 2012; Schimmack, 2012; Wagenmakers et al., 2011). Most notably, the first and second methods influence participant decisions by showing pictures after a decision is made, and many researchers choose to present erotic and/or highly arousing negative stimuli to obtain an effect. Not all university institutional review boards (IRBs) would allow researchers to present erotic and/or highly arousing negative stimuli to participants, especially to study RI. Also, the second and third methods are often used to habituate or prime participants in a subliminal manner. To do so, these methods must be applied in lab settings to use the required computer programs. Not all researchers

[1] A "hit" may also be when the participant does not select the picture shown immediately afterwards in certain cases, such as when erotic stimuli are used and/or when the habituation is meant to induce boredom.

have access to a research lab to perform these RI studies, thereby limiting the number of researchers that can study the effect. Because RI is a highly debated topic, some authors have suggested that the limited number of researchers that study RI is a severe limiting factor for the field. Those that study RI may be overly zealous to find effects, and any results may be biased by the researchers. For this reason, these authors have suggested that RI studies should be open to anyone, which is not possible with the second and third methods. Lastly, due to the reliance on lab studies, sample sizes in RI studies have been criticized for being too small, often ranging from 20 to 50 participants (Bem, 2011; Bierman & Bijl, 2014; Ertel, 2005a, 2005b, 2010; Luke & Morin, 2009; Luke et al., 2008; Luke & Zychowicz, 2014; Maier et al., 2014; Watt, 2014).

Perhaps the methods that have garnered the most attention, however, are the fourth and fifth (Barušs & Rabier, 2014; Ertel, 2005a, 2005b, 2010; Galak et al., 2012; Ritchie et al., 2012; Robinson, 2011; Storm et al., 2013a, 2013b; Storm et al., 2012; Subbotsky, 2013; Traxler et al., 2012; Vernon, 2015). These methods provide several advantages to research methodology. They do not require aversive pictures to be shown to participants, and most any university IRB would allow the study. While participants need to be engaged, these methods also do not require participants to participate in a lab scenario using certain computer programs. Participants can instead participate via an online survey. Because of their methodological benefits, the fourth and fifth methods are often considered the ideal methods to test for RIs (Barušs & Rabier, 2014; Ertel, 2005a, 2005b, 2010; Galak et al., 2012; Ritchie et al., 2012; Robinson, 2011; Storm et al., 2013a, 2013b; Storm et al., 2012; Subbotsky, 2013; Traxler et al., 2012; Vernon, 2015), and they are the chosen methods applied in the current series of studies.

The Current Studies

The current article tests two similar versions of the fourth and fifth methods. Because the convergent validity of RI measures has yet to be shown, it was believed that testing two similar methods would be more appropriate than testing two very different methods, such as two and four above. If the two similar versions do not demonstrate sufficient convergent validity, then it is unlikely that two very different methods would instead show sufficient convergent validity.

The first method is almost identical to the description above for Method 4, but the task is made as direct as possible. As further detailed below, participants are presented a list of 20 words with a similar valence, and they are asked to select the five that appear on the following page. Then, on the following page, they are randomly shown five of these words and asked to type them in the space provided. If RIs exists, then the participants should select the practice words more so than random chance alone (6.25% chance of selecting correct word, 1.25 correct words selected on average). Thus, this direct version of Method 4 could also be considered a version of method five, because participants are directly asked to predict an immediate future event.

The second method is similar, but uses pictures instead of words. In a series of 12 trials, participants are shown four pictures at a time, and they are asked to identify the picture that appears on the following page. Then, they are randomly shown one of the pictures on the following page. If RIs exists, then the participants should select the correct picture more so than random chance alone (25% chance of selecting correct picture, 3 correct pictures selected on average). This direct method is most similar to method five above, but it also contains aspects of method four.

Testing these two methods can provide further insight into the nature of RI, but four further notes should be made before continuing. First, prior authors have suggested that measures must demonstrate a correlation of .70 or above to fully support their convergent validity; however, authors have also suggested that much smaller correlations are acceptable when the measures use different methods, and correlations as small as .30 may be sufficient (Campbell & Fiske, 1959; Kenny & Kashy, 1992). In the current article, a correlation of .30 is considered sufficient to support the convergent validity of the two methods.

Second, the current article does not only test the methods' relationship with each other, but also compares their relationships with other variables. If both methods gauge RI, they would be expected to share similar relationships with

other variables. Perhaps the most commonly studied relationships of RI are with perceived RI abilities and sensation seeking (Bem, 2011; LeBel & Peters, 2011), which are investigated in the current article. Some authors have also shown interest in the relationship of RI with age, gender, and the Big Five (Honorton & Ferrari, 1989; Storm et al., 2012). Each of these relationships is likewise studied. If the two methods have similar relationships, then nomological network for measuring RI can also be supported.

Third, most prior authors have tested the existence of RIs through comparing study results to random chance alone (i.e., one-tailed *t* test). This is also done in the current article, but testing the convergent validity of the two methods may provide a more robust test of RI effects. Through comparing with random chance alone, researchers assume that everyone approaches or avoids RI stimuli as expected—toward positive stimuli and away from negative stimuli. It is possible that certain people avoid (approach) RI stimuli, even if they are positive (negative), due to their unexplained feelings toward the stimuli. If this is the case, then comparing results with random chance alone may not uncover a RI effect, because an equal number of people may have approached and avoided the RI stimuli. When testing the relation of two RI methods, however, these people should approach (avoid) stimuli with both methods, and a strong correlation could be found in the absence of any difference from random chance alone. Thus, the current study may provide benefits for the methodology of RI research in addition to measurement benefits.

Fourth, studying the convergent validity and nomological network of RI measures implicitly assumes that certain individual differences may exist in regards to RI abilities. That is, some people may have a natural tendency to be more or less susceptible to RI effects, which can be supported by validity and nomological network information. An RI effect may still exist without the presence of any individual differences, however. All people may have relatively equal abilities to identify RI effects, and any deviations from the norm may be randomly distributed throughout trials of RI tasks. If this is indeed the case, then RI measures would not produce notable relationships with other variables, but tests of RI effects would provide findings in which successful results (i.e., identifying the correct picture or word) are achieved at a greater rate than random chance alone. In the current article, this possibility is also tested. Thus, while the primary focus of the current article is the measurement of RI abilities and related individual differences, the current studies can also test for other RI effects.

While considering these aspects, three studies are performed in the following. In Study 1, two RI methods are tested together. In Study 2, a close replication is performed. In Study 3, a conceptual replication is performed. Recent authors have argued that different dynamics occur when RI is tested consciously and subconsciously. For example, RI may have an effect when participants are asked to predict the word that will appear, but it may have an even stronger effect when participants are asked about their favorability of certain words. This difference is tested in Study 3, in which participants are asked about their favorability of words and pictures, rather than predicting the words and pictures that will appear.

Study 1

Method

Participants. Participants ($N = 157$, $M_{age} = 33.66$, $SD_{age} = 9.43$, 42% female, 77% American) were recruited from Mechanical Turk (MTurk) and provided US$0.55 as compensation. MTurk is a website that connects individuals willing to perform tasks on a computer, such as taking a survey, with those needing the tasks completed. Several prior studies have shown that results obtained from MTurk samples are reliable and valid, even when studying special populations (Buhrmester, Kwang, & Gosling, 2011; Paolacci & Chandler, 2014; Shapiro, Chandler, & Mueller, 2013). All participants that did not follow all aspects of the study procedures were removed, as described below. All statistics, including descriptive information, reflect the sample after removing these participants.

Procedure. All procedures for Studies 1, 2, and 3 were reviewed and approved by the IRB of the primary author's university.

Study 1 had four primary sections that were completed in a single sitting. First, participants gained access to the survey via a link provided

on MTurk, allowing them to complete all study procedures in the location of their choosing. The link was advertised as, "Test Your Ability to Predict the Future." Then, they provided their informed consent and completed a prequestionnaire.

Second, participants completed the first RI task, which was closely adapted from Bem (2011) and many prior studies (Barušs & Rabier, 2014; Galak et al., 2012; Ritchie et al., 2012; Subbotsky, 2013; Traxler et al., 2012; Vernon, 2015). Participants were first told,

> This section tests your ability to correctly predict the future. Twenty words appear below, and five of these words will be presented on the next page. Your task is to select the five words that will appear on the next page. Also, it is absolutely critical that you focus only on this task and do not perform any other tasks (e.g. check email).

Then, 20 words were presented to the participant in a randomized order, and they could not continue until they selected five of these words. On the following page, five randomly selected words were shown, and participants were asked to type the five words that were presented. Participants that did not type all five provided words were removed from the final data analyses.

Third, participants completed the second RI task, which was also closely adapted from many prior studies (Ertel, 2005a, 2005b, 2010; Honorton & Ferrari, 1989; Storm et al., 2013a, 2013b; Storm et al., 2012). Participants were first told,

> This section also tests your ability to correctly predict the future. In this section of the survey, you will now be asked to answer 12 similar questions. The questions will ask you to select one of four photos that will appear on the page following the question. Your task is to correctly select the photo that will appear on the page following the question. Again, it is absolutely critical that you focus only on this task and do not perform any other tasks (e.g., check e-mail).

The task then included 12 trials, in which four pictures were presented at a time in a randomized order. Participants were asked to select which of the four pictures would appear on the following page, then a single picture would be randomly presented on the following page. Participants that did not complete all 12 trials were removed from the final data analyses.

Fourth, participants completed a postquestionnaire, were thanked for their time, and were disclosed the purpose of the study.

Stimuli. The same set of words and pictures were used for all three studies. The 20 words were taken from the database of Warriner, Kuperman, and Brysbaert (2013), in which participants were asked to rate the valence, arousal, and dominance of 13,915 words. The 20 selected words each had the same valence rating, which was the median of the entire database, and similar arousal and dominance ratings. That was intended to prevent participants from selecting words (or not selecting words) that had any natural appeal (or repulsion) beyond other words. The 48 pictures were taken from the database of Dan-Glauser and Scherer (2011), in which participants were asked to rate the valence, arousal, and the congruence of the represented scene with internal (moral) and external (legal) norms of 730 pictures. The 48 selected pictures were those surrounding the median valence of the database, and they had similar arousal and congruence ratings. Again, this was intended to prevent participants from selecting pictures (or not selecting pictures) that had any natural appeal (or repulsion) beyond other pictures.

It should also be noted that the "correct" words and pictures for each participant were randomly selected using the survey software's randomizer, resulting in a different set of correct words and pictures for each participant. While it is believed that no participant was able to view the task before participating, randomizing the correct choices helps ensure that prior knowledge (as opposed to future knowledge) would not aid performance on the task. Also, the survey software randomly selected the correct choices after the items or pictures were shown. This helps ensure that any deviations from random chance are due to RI, rather than participants being omniscient and able to envision the selected correct choices before making their own choice.

Measures.

Retroactive influencing Task 1 (RI Task 1). The results of RI Task 1 were recorded as the number of words that the participant correctly predicted appearing on the next page.

Retroactive influencing Task 2 (RI Task 2). The results of RI Task 2 were recorded as the number of pictures that the participant correctly predicted appearing on the next page.

Perceived RI abilities. A self-created three-item perceived RI abilities scale was applied in the current study. The three items are, "I believe that I am able to predict the future to a certain extent," "I can sense some things before they happen," and "I can sometimes feel when an event is going to occur."

Sensation seeking. The two-item sensation seeking scale created by Bem (2011) was applied in the current study. The two items are, "I am easily bored" and "I often enjoy seeing movies I've seen before" (reverse coded).

Age. Participants were asked to provide their age in years.

Gender. Participants were asked to indicate their gender, which was coded as 0 (male) and 1 (female). An "other" option was provided, but no participant responded as other.

Results and Discussion

Correlations and Cronbach's alphas of all study variables are presented in Table 1. The two RI tasks had a small, positive, and nonsignificant relationship ($r = .09$, two-tailed $p > .10$). Each of the RI tasks also had nonsignificant relationships with all other measured variables (all two-tailed $p > .10$), and the Cronbach's alphas of the RI tasks were essentially zero.

Also, one-sample t tests were performed to determine whether a general RI effect occurred for both tasks. The result of RI Task 1 was compared against an a priori value of 1.25, as this would be the average number of predicted words due to random chance alone. The average number of words correctly predicted was 1.27 for Study 1. The one-sample t test was not statistically significant ($t = .354$, two-tailed $p > .10$, 95% CI [−.10, .15]. The result of RI Task 2 was compared against an a priori value of 3, as this would be the average number of predicted pictures due to random chance alone. The average number of pictures correctly predicted was 2.95 for Study 2. The one-sample t test was not statistically significant ($t = -.427$, two-tailed $p > .10$, 95% CI [−.29, .19].

Study 2: Close Replication

Method

Participants. Participants ($N = 181$, $M_{age} = 32.52$, $SD_{age} = 9.18$, 38% female, 94% American) were recruited from MTurk and provided US$0.55 as compensation. All participants that (a) did not follow all aspects of the study procedures, and/or (b) reported participating in Studies 1 or 3 were removed.[2] All statistics reflect the sample after removing these participants.

Procedure. All procedures for Study 2 were identical to Study 1.

Measures. All measures applied in Study 1 were applied in Study 2 along with the measure below.

Big Five. The 10-item measure of the Big Five created by Gosling, Rentfrow, and Swann (2003) was applied in the current study. Each item starts with the stem, "I see myself as: . . .," followed by the item content. An example item is, "Extraverted, enthusiastic" (Extraversion).

Results and Discussion

Correlations and Cronbach's alphas of all study variables are presented in Table 2. The two RI tasks had a small, negative, and nonsignificant relationship ($r = -.09$, two-tailed $p > .10$). Each of the RI tasks also had nonsignificant relationships with most all other measured variables (all two-tailed $p > .05$), but RI Task 1 had a significant relationship with gender ($r = -.22$, two-tailed $p < .01$). Given the large number of analyses, however, this single significant result cannot be considered nonspurious. The Cronbach's alphas of the RI tasks were again essentially zero.

Also, one-sample t tests were performed to determine whether a general RI effect occurred for both tasks. The result of RI Task 1 was compared against an a priori value of 1.25. The average number of words correctly predicted was exactly 1.25 for Study 2. The one-sample t test was not statistically significant ($t = -.042$, two-tailed $p > .10$, 95% CI [−.13, .12]. The result of RI Task 2 was compared against an a priori value of 3. The average number of pic-

[2] Chronologically, Study 2 was performed after Studies 1 and 3, but it is reported second for clarity purposes.

Table 1
Correlations and Cronbach's Alphas of Study 1 Variables

Measure	1	2	3	4	5	6	7
1. RI Task 1	<.00[a]						
2. RI Task 2	.09	.01					
3. RI ability prequestionnaire	−.02	−.10	.93				
4. RI ability postquestionnaire	−.02	.12	.76**	.94			
5. Sensation seeking	−.02	.01	−.11	−.12	.28		
6. Age	.01	−.12	−.18*	−.23**	−.19*	N/A	
7. Gender	−.01	−.02	−.12	−.10	−.10	.18*	N/A

Note. RI = retroactive influence.
[a] Cronbach's alpha was a negative number because the covariance among the items was negative.
* $p < .05$. ** $p < .01$.

tures correctly predicted was exactly 3.00 for Study 2. The one-sample *t* test was not statistically significant ($t = .000$, two-tailed $p > .10$, 95% CI [−.21, .21].

Study 3: Conceptual Replication

Method

Participants. Participants ($N = 178$, $M_{age} = 37.71$, $SD_{age} = 11.72$, 47% female, 90% American) were recruited from MTurk and provided US$0.55 as compensation. All participants that (a) did not follow all aspects of the study procedures, and/or (b) reported participating in Study 1 were removed. All statistics reflect the sample after removing these participants.

Procedure. The procedures for Study 3 were mostly identical to Study 1, except participants were asked to indicate their favorite words/pictures instead of predicting those that would appear. The study on MTurk was advertised as, "Select Your Preference for Certain Words and Pictures." The instructions for the first RI task read, "Of the following 20 words, please select the 5 that you like the best. Also, it is absolutely critical that you focus only on this task and do not perform any other tasks (e.g. check email)."

The instructions for the second RI task read,

> In this section, you will be shown a series of four pictures at a time. Of the four pictures, select the one that you like the best. Again, it is absolutely critical that you focus only on this task and do not perform any other tasks (e.g., check e-mail).

Table 2
Correlations and Cronbach's Alphas of Study 2 Variables

Measure	1	2	3	4	5	6	7	8	9	10	11	12
1. RI Task 1	<.00[a]											
2. RI Task 2	−.10	<.00[a]										
3. RI ability prequestionnaire	−.14	−.14	.92									
4. RI ability postquestionnaire	−.14	.01	.87**	.94								
5. Sensation seeking	.21	−.04	.07	.07	.24							
6. Openness	.04	−.03	.17*	.19*	−.03	.56						
7. Conscientiousness	.00	.07	−.12	−.05	−.23**	.00	.74					
8. Extraversion	−.10	.11	.12	.14	−.20**	.33**	.35**	.74				
9. Agreeableness	−.05	.04	−.13	−.12	−.28**	.07	.25**	.32**	.46			
10. Emotional Stability	−.02	.04	−.13	−.11	−.33**	.10	.48**	.71**	.44**	.83		
11. Age	.07	−.07	−.21**	−.18*	−.23**	−.16*	.29**	.10	.31**	.25**	N/A	
12. Gender	.22**	−.02	−.06	−.09	−.01	−.14	.08	−.10	.15*	−.05	.16*	N/A

Note. RI = retroactive influence.
[a] Cronbach's alpha was a negative number because the covariance among the items was negative.
* $p < .05$. ** $p < .01$.

Thus, participants were asked to indicate their favorite picture instead of forecasting events.

Measures. All measures applied in Study 2 were applied in Study 3; however, a prequestionnaire measure of perceived RI abilities was not provided, which may have otherwise revealed the study purpose to participants. Also, an item was added to the postquestionnaire, which allowed participants to freely respond regarding the purpose of the study. All participants either reported that the purpose was to identify their favorite pictures or that they were unsure about the purpose.

Results and Discussion

Correlations and Cronbach's alphas of all study variables are presented in Table 3. The two RI tasks had a small, positive, and nonsignificant relationship ($r = .09$, two-tailed $p > .10$). Each of the RI tasks also had nonsignificant relationships with all other measured variables (all two-tailed $p > .10$). The Cronbach's alphas of the RI tasks were again essentially zero.

Also, one-sample t tests were performed to determine whether a general RI effect occurred for both tasks. The result of RI Task 1 was compared against an a priori value of 1.25. The average number of words correctly predicted was 1.18 for Study 2. The one-sample t test was not statistically significant ($t = -1.157$, two-tailed $p > .10$, 95% CI [−.19, .05]). The result of RI Task 2 was compared against an a priori value of 3. The average number of pictures correctly predicted was 3.12 for Study 2. The one-sample t test was not statistically significant ($t = 1.071$, two-tailed $p > .10$, 95% CI [−.10, .35].

General Discussion

The primary purpose of the current article was to test the convergent validity and nomological network of two methods to measure RIs. The relationship of the two methods did not meet even minimal requirements for convergent validity in all three studies, failing to provide support for their ability to measure RI. The overall sample size weighted correlation from all three studies together was .03 ($\bar{r}$, 95% CI [−.07, .12], indicating that the relationship between the two variables cannot be considered anything more than random chance alone. Also, the two measures did not share any significant relationships with any other variables, suggesting that they do not share a common nomological network (unless their nomological network is to be unrelated to all other variables—including each other). The Cronbach's alphas of the measures were essentially zero across all studies. This indicates that the trials within the two measures did not significantly relate to each other, and virtually identical RI tasks do not produce a meaningful relationship—or much any relationship at all. Lastly, observed RI effects did not occur more than random chance alone, again suggesting that an RI effect did not occur in any study. Thus, the three studies did

Table 3
Correlations and Cronbach's Alphas of Study 3 Variables

Measure	1	2	3	4	5	6	7	8	9	10	11
1. RI Task 1	<.00[a]										
2. RI Task 2	.09	<.00[a]									
3. RI ability postquestionnaire	−.01	.01	.93								
4. Sensation seeking	−.07	.08	.04	.14							
5. Openness	.02	−.09	.13	−.33**	.59						
6. Conscientiousness	−.01	−.02	−.10	−.37**	.29**	.65					
7. Extraversion	−.02	.06	.19*	−.38**	.41**	.37**	.83				
8. Agreeableness	.06	−.08	.00	−.32**	.31**	.38**	.40**	.62			
9. Emotional stability	−.03	−.03	−.07	.44**	.26**	.51**	.68**	.52**	.77		
10. Age	.01	.06	−.09	−.08	.04	.26**	.07	.17*	.19*	N/A	
11. Gender	.11	−.03	−.01	−.01	.02	.11	−.07	.03	.01	.24**	N/A

Note. RI = retroactive influence.
[a]Cronbach's alpha was a negative number because the covariance among the items was negative.
* $p < .05$. ** $p < .01$.

not support that the two tested methods are able to accurately measure RIs.

Implications and Future Directions

Several implications and future directions should be considered. Most notably, several prior publications have relied on methods to measure RI that are almost identical to those applied in the current article (Ertel, 2005a, 2005b, 2010; Honorton & Ferrari, 1989; Storm et al., 2013a, 2013b; Storm et al., 2012). The results of the current studies draw these prior investigations into question. While these results do not suggest foul play by prior authors, they do suggest that prior significant findings were no more than spurious relationships. Further, the current article calls the broader scheme of RI studies into question, even those that did not apply almost identical measures to gauge RI. Because these studies did not confirm their RI measurement methods, they too cannot claim that their effects are anything more than spurious relationships.

The current studies provide a guide for future measurement research on RIs. Several authors have noted that RI research suffers from a paucity of measurement investigations (LeBel & Peters, 2011), but few authors have heeded this call for future research. In heeding this call, the current article not only showed that it is indeed important to test the measurement of RI, but it also provided a clear method to do so. Performing an initial study, followed by a close replication, followed by a conceptual replication can clearly support whether certain RI measurement methods are appropriate, but it also satisfies the criticisms of prior authors who have argued that not enough close or conceptual replication studies are performed in RI research (Brandt et al., 2014; LeBel & Peters, 2011; Schimmack, 2012; Wetzels et al., 2011). A series of three studies could also provide the "extraordinary evidence" that is required for "extraordinary claims" (Sagan, 1979; Wagenmakers et al., 2011; Wetzels et al., 2011).

Future research should continue to study the measurement of RIs, particularly with the other types as differentiated by Bem (2011). While ambitious, a study which tests the convergent validity of all five types together would provide impactful results and clear implications for subsequent research. However, a more reasonable expectation may be a series of studies that each test the convergent validity of two methods of a common type, which may result in the identification of a method or type that shows appropriate convergent validity.

It should be noted that the current article also investigated the general RI effect—the ability of participants to be influenced by events more so than random chance alone. Across three studies and two different tasks, for a total of six trials, none of the trials showed an RI effect that was significantly different than random chance alone. That is, participants only identified the correct words or pictures at a rate that was statistically similar to random chance. These results agree with the majority of replications after Bem (2011) that also failed to replicate any RI effects (Brandt et al., 2014; LeBel & Peters, 2011; Schimmack, 2012).

In addition to this traditional method, the current article emphasized the potential for convergent validity tests to identify RI effects. As mentioned, the traditional method assumes that everyone approaches or avoids RI stimuli as expected—toward positive and away from negative stimuli. It is possible that certain people avoid (approach) RI stimuli, even if they are positive (negative), due to their unexplained feelings toward the stimuli. These people would offset RI effects as gauged by the traditional method, but a convergent validity test could still provide significant results despite these people (although it did not in the current article). Thus, testing the convergent validity of RI measures can provide benefits beyond measurement alone.

In testing other possible methods to measure RI, an appropriate starting point may be to identify method(s) that produce acceptable internal consistency reliabilities. The observed relationships of a measure are limited by its unreliability (Campbell & Fiske, 1959; Howard & Jayne, 2015; Kenny & Kashy, 1992; Waechter, Nelson, Wright, Hyatt, & Oakman, 2014). A measure with little to no internal consistency, such as the measures applied in the current studies, cannot provide consistent and meaningful relationships. Further, some studies identify individuals with apparent RI abilities to conduct focused follow-up testing (Kenny & Kashy, 1992; Waechter et al., 2014), and these individuals are identified by their notable performance on RI tasks similar to those applied in the cur-

rent article. Measures with poor internal consistency cannot accurately identify a person's standing on a latent trait, and identifying those with apparent RI abilities using these methods only results in spurious findings. For these reasons, if the study of RIs is to progress in an appropriate manner, it may be best to first address this measurement limitation.

Lastly, in addition to these implications for RI research, the current article also highlights the importance of valid measurement across all domains of psychology and beyond. Many authors have criticized prior research for applying measures with little investigation into their psychometric properties and validity (Howard & Jayne, 2015; LeBel & Peters, 2011; MacKenzie, Podsakoff, & Podsakoff, 2011). The current series of studies can be used as an example of prior results being questioned by new discoveries regarding measurement. It can also be used to emphasize the importance of close and conceptual replications across all domains of psychology—a call that has been increasingly made since Bem (2011; Brandt et al., 2014; Wagenmakers et al., 2011).

Limitations

Certain limitations of the current article should be noted. First, the current article only tested two methods to measure RIs, which could both be considered variations of the same type of method as differentiated by Bem (2011). It was assumed that two dissimilar methods would not demonstrate sufficient convergent validity if two similar methods did not demonstrate sufficient convergent validity. For this reason, it was believed to be more appropriate to test two similar methods first, and two dissimilar methods could then be tested if support was found.

Second, the current article tested intentional and direct RIs (Studies 1 and 2) as well as unintentional and indirect RIs (Study 3). The current article did not test a middle ground between these two, however. For example, Bem (2011; Studies 8 and 9) told participants that the retroactive priming task was meant to gauge psi abilities, but they were not directly told that it was meant to specifically gauge RI abilities. It is possible that different results may have been observed while using this other approach, although no current theory supports this notion.

Third, a benefit of the two methods applied was their ability to be administered via an online survey, allowing most any researcher to conduct RI studies. It is also possible that RI effects can only be observed in highly controlled lab settings. While a detriment to the overall field of RI research, it is nevertheless a possibility that should be considered.

Fourth, while the sensation seeking and Big Five measures were taken from prior research (Bem, 2011; Gosling et al., 2003), they demonstrated poor internal consistencies in the current article. Particularly, the sensation seeking measure had a Cronbach's alpha below .30 in all studies. While Bem (2011) reported several significant relationships of the measure, subsequent authors have expressed concern regarding its use—questioning whether it gauges sensation seeking at all (Galak et al., 2012). Given these concerns, it appears appropriate to no longer use Bem's sensation seeking measure.

Conclusion

The goal of the current article was to determine whether two popular methods to measure RIs demonstrate sufficient convergent validity and nomological network overlap. The results showed that the two measures were largely unrelated, and their relationships with other constructs largely differed. The relationship between the two measures, and possibly the two measures themselves, cannot be considered anything more than random chance alone. These results shed further doubt on the study of RIs, but may also serve as a guide for the study of RIs moving forward.

References

Barušs, I., & Rabier, V. (2014). Failure to replicate retrocausal recall. *Psychology of Consciousness, 1,* 82–91. http://dx.doi.org/10.1037/cns0000005

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100,* 407–425. http://dx.doi.org/10.1037/a0021524

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000 Research, 4,* 1188.

Bierman, D. J., & Bijl, A. (2014). Anomalous 'retrocausal' effects on performance in a go/nogo task. *Journal of Scientific Exploration, 28,* 437–452.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., . . . Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50,* 217–224. http://dx.doi.org/10.1016/j.jesp.2013.10.005

Broderick, D., & Goertzel, B. (Eds.). (2014). *Evidence for psi: Thirteen empirical research reports*. Jefferson, NC: McFarland.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105. http://dx.doi.org/10.1037/h0046016

Caswell, J. M., Hunter, J., & Tessaro, L. W. (2014). Phenomenological convergence between major paradigms of classic parapsychology and cross-cultural practices: An exploration of paranthropology. *Journal of Consciousness Exploration & Research, 5,* 467–482.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva Affective Picture Database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods, 43,* 468–477. http://dx.doi.org/10.3758/s13428-011-0064-1

Diaconis, P. (1978). Statistical problems in ESP research. *Science, 201,* 131–136. http://dx.doi.org/10.1126/science.663642

Ertel, S. (2005a). Psi test feats achieved alone at home: Do they disappear under lab control? *Australian Journal of Parapsychology, 5,* 149–164.

Ertel, S. (2005b). The ball drawing test: Psi from untrodden ground. In M. A. Thalbourne & L. Storm (Eds.), *Parapsychology in the 21st century: Essays on the future of psychical research* (pp. 90–123). Jefferson, NC: McFarland.

Ertel, S. (2010). Psi in a skeptic's lab: A successful replication of Ertel's Ball Selection Test. *Journal of Scientific Exploration, 14,* 461–478.

Galak, J., Leboeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate ψ. *Journal of Personality and Social Psychology, 103,* 933–948. http://dx.doi.org/10.1037/a0029709

Gaona, J. M., Colinas, F., Rouleau, N., Tessaro, L. W., & Caswell, J. M. (2014). Premonitions: A global online statistical tracking study of precognitive predictions. *Journal of Consciousness Exploration & Research, 5,* 341–350.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37,* 504–528. http://dx.doi.org/10.1016/S0092-6566(03)00046-1

Honorton, C., & Ferrari, D. C. (1989). "Future Telling": A meta-analysis of forced-choice precognition experiments, 1935–1987. *Journal of Parapsychology, 53,* 281–308.

Howard, M. C., & Jayne, B. S. (2015). An analysis of more than 1,400 articles, 900 scales, and 17 years of research: The state of scales in cyberpsychology, behavior, and social networking. *Cyberpsychology, Behavior and Social Networking, 18,* 181–187. http://dx.doi.org/10.1089/cyber.2014.0418

James, W. (1890). *The principles of psychology*. New York, NY: Holt. http://dx.doi.org/10.1037/11059-000

James, W. (1902). *The varieties of religious experience*. Cambridge, MA: Harvard University Press.

Jung, C. G. (1917). *Two essays on analytical psychology*. New York, NY: Meridian.

Jung, C. G. (1958). S. DeLaszlo Violet, *Psyche & symbol: A selection from the writings of G. Jung*. Garden City, NJ: Doubleday Anchor Books.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165–172. http://dx.doi.org/10.1037/0033-2909.112.1.165

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011). evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology, 15,* 371–379. http://dx.doi.org/10.1037/a0025172

Luke, D., & Morin, S. (2009). *Luck beliefs, PMIR, psi and the sheep-goat effect: A replication*. Paper presented at the 33rd International Conference of the Society for Psychical Research, University of Nottingham, U.K.

Luke, D. P., Roe, C. A., & Davison, J. (2008). Testing for forced-choice precognition using a hidden task: Two replications. *Journal of Parapsychology, 72,* 133–154.

Luke, D., & Zychowicz, K. (2014). Comparison of outcomes with nonintentional and intentional precognition tasks. *Journal of Parapsychology, 78,* 223–234.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *Management Information Systems Quarterly, 35,* 293–334. http://dx.doi.org/10.2307/23044045

Maier, M. A., Büchner, V. L., Kuhbandner, C., Pflitsch, M., Fernández-Capo, M., & Gámiz-

Sanfeliu, M. (2014). Feeling the future again: Retroactive avoidance of negative stimuli. *Journal of Consciousness Studies, 21,* 121–152.

Meehl, P. E., & Scriven, M. (1956). Compatibility of science and ESP. *Science, 123,* 14–15. http://dx.doi.org/10.1126/science.123.3184.14

Morris, R. L. (1982). Assessing experimental support for true precognition. *Journal of Parapsychology, 46,* 321–336.

Paolacci, G., & Chandler, J. (2014). Inside the Turk. *Current Directions in Psychological Science, 23,* 184–188. http://dx.doi.org/10.1177/0963721414531598

Parker, A., & Sjödén, B. (2010). Do some of us habituate to future emotional events? *Journal of Parapsychology, 74,* 99–115.

Parra, A. (2013). A phenomenological examination of premonition experiences in dreams and waking states: A survey study. *Australian Journal of Parapsychology, 13,* 187–212.

Pratt, J. G., & Price, M. M. (1938). The experimenter-subject relationship in tests for ESP. *Journal of Parapsychology, 2,* 84–94.

Price, G. R. (1955). Science and the supernatural. *Science, 122,* 359–367. http://dx.doi.org/10.1126/science.122.3165.359

Rattet, S. L., & Bursik, K. (2001). Investigating the personality correlates of paranormal belief and precognitive experience. *Personality and Individual Differences, 31,* 433–444. http://dx.doi.org/10.1016/S0191-8869(00)00148-3

Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's 'retroactive facilitation of recall' effect. *PLoS ONE, 7,* e33423. http://dx.doi.org/10.1371/journal.pone.0033423

Robinson, E. (2011). Not feeling the future: A failed replication of retroactive facilitation of memory recall. *Journal of the Society for Psychical Research, 75,* 142–147.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001

Sagan, C. (1979). *Broca's brain, reflections on the romance of science*. New York, NY: Random House.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17,* 551–566. http://dx.doi.org/10.1037/a0029487

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science, 1,* 213–220. http://dx.doi.org/10.1177/2167702612469015

Steinkamp, F. (2000). Does precognition foresee the future? A postal experiment to assess the possibility of true precognition. *Journal of Parapsychology, 64,* 3–18.

Steinkamp, F. (2001). Does precognition foresee the future? Series 2, a laboratory replication and Series 3, a world wide web replication. *Journal of Parapsychology, 65,* 17–40.

Storm, L., Ertel, S., & Rock, A. J. (2013a). The sheep-goat effect as a matter of compliance vs. noncompliance: The effect of reactance in a forced choice ball selection test. *Journal of Scientific Exploration, 27,* 393–411.

Storm, L., Ertel, S., & Rock, A. J. (2013b). Paranormal effects and behavioural characteristics of participants in a forced-choice Psi task: Ertel's ball selection test under scrutiny. *Australian Journal of Parapsychology, 13,* 111–131.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analysis of ESP studies, 1987–2010: Assessing the success of the forced-choice design in parapsychology. *Journal of Parapsychology, 76,* 243–273.

Subbotsky, E. (2013). Sensing the future: Reversed causality or a non-standard observer effect. *The Open Psychology Journal, 6,* 81–93.

Traxler, M. J., Foss, D. J., Podali, R., & Zirnstein, M. (2012). Feeling the past: The absence of experimental evidence for anomalous retroactive influences on text processing. *Memory & Cognition, 40,* 1366–1372. http://dx.doi.org/10.3758/s13421-012-0232-2

Trochim, W. (2006). The nomological network: What is a nomological network? Retrieved from https://www.socialresearchmethods.net/kb/nomonet.php

Vernon, D. J. (2015). Exploring precognition using a repetition priming paradigm. *Journal of the Society for Psychical Research, 79,* 65–79.

Waechter, S., Nelson, A. L., Wright, C., Hyatt, A., & Oakman, J. (2014). Measuring attentional bias to threat: Reliability of dot probe and eye movement indices. *Cognitive Therapy and Research, 38,* 313–333. http://dx.doi.org/10.1007/s10608-013-9588-2

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100,* 426–432. http://dx.doi.org/10.1037/a0022790

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7,* 632–638. http://dx.doi.org/10.1177/1745691612463078

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45,* 1191–1207. http://dx.doi.org/10.3758/s13428-012-0314-x

Watt, C. (2014). Precognitive dreaming: Investigating anomalous cognition and psychological factors. *Journal of Parapsychology, 78,* 115–125.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6,* 291–298. http://dx.doi.org/10.1177/1745691611406923

Received June 22, 2017
Revision received November 11, 2017
Accepted December 23, 2017 ■