Journal of Experimental Psychology: General
2015, Vol. 144, No. 4, e65–e72

0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000090

The Burden of Secrecy? No Effect on Hill Slant Estimation and Beanbag Throwing

Diane Pecher and Heleen van Mierlo
Erasmus University Rotterdam

Rouwen Cañal-Bruland
VU University Amsterdam

René Zeelenberg
Erasmus University Rotterdam

Slepian, Masicampo, Toosi, and Ambady (2012, Experiment 1) reported that participants who recalled a big secret estimated a hill as steeper than participants who recalled a small secret. This finding was interpreted as evidence that secrets are experienced as physical burdens. In 2 experiments, we tried to replicate this finding, but, despite larger power, did not find a difference in slant estimates between participants who recalled a big secret and those who recalled a small secret. This finding was further corroborated by a meta-analysis that included 8 published data sets of exact replications, which indicates that thinking of a big secret does not affect hill slant estimation. In a third experiment, we also failed to replicate the effect of recalling a secret on throwing a beanbag at a target (Slepian et al., 2012, Experiment 2). Together, our findings question the robustness of the original empirical findings.

Keywords: conceptual metaphor, embodied cognition, secrets, perceptual judgment, replication

Several recent theories of cognition are based on the idea that cognition shares processing mechanisms with perception and action (Barsalou, 1999, 2008; Glenberg, 1997). When people mentally represent concepts such as *coffee* or *car*, they mentally simulate interacting with these entities, such as smelling fresh coffee or seeing a car drive by, using the same processing mechanisms as for perception and action. For example, van Dantzig, Pecher, Zeelenberg, and Barsalou (2008) showed that mental representations are influenced by modality-specific stimuli such as noise bursts and light flashes. Participants verified object-property statements, such as "a banana is yellow," and performed better when these statements were preceded by stimuli from the same modality than by stimuli from a different modality. Such evidence supports the idea that cognition shares processing mechanisms with perception and action.

A major question for the grounded cognition view is how sensory-motor systems are involved in mental representations of abstract concepts such as *secret* or *power* (Dove, 2009; Machery, 2007; Pecher, Boot, & Van Dantzig, 2011). Abstract concepts do not have sensory-motor features that can be used in a mental simulation. Cognitive linguists (Gibbs, 1994; Lakoff & Johnson, 1980, 1999) have proposed that people map concrete image schemas on abstract concepts. Their argument was initially based on the observation that language about abstract concepts often contains metaphors to concrete domains. For example, people talk about *life* as a *journey* in which people choose a path and have a destination, and meet other travelers on the way. Empirical studies have shown that concrete image schemas are activated when people mentally represent abstract concepts (Boot & Pecher, 2010, 2011; Casasanto & Boroditsky, 2008; Cienki, 2005; Gentner, Imai, & Boroditsky, 2002; Meier & Robinson, 2004; Teuscher, McQuire, Collins, & Coulson, 2008). For example, people think of a powerful person as being higher than a powerless person, causing visual attention to move upward or downward (Zanolie et al., 2012), and of similarity as being closer in space, resulting in faster similarity ratings when items are presented closer together and faster dissimilarity ratings when items are presented further apart (Boot & Pecher, 2010). Thus, abstract concepts might be grounded in sensory-motor processing via metaphorical mappings between the abstract concept and a concrete domain.

In keeping with this general framework, Slepian, Masicampo, Toosi, and Ambady (2012) hypothesized that the metaphor "secrets are burdens" might play a role in people's experiences of keeping a secret. They proposed that secrets have the effect of weighing people down, similar to carrying a physical weight. Following this line of reasoning, they hypothesized that secrets would have similar effects on perception as physical burdens. To test this assumption, Slepian et al. (2012, Experiment 1) asked

This article was published Online First June 22, 2015.

Diane Pecher and Heleen van Mierlo, Department of Psychology, Erasmus University Rotterdam; Rouwen Cañal-Bruland, Department of Human Movement Sciences, VU University Amsterdam; René Zeelenberg, Department of Psychology, Erasmus University Rotterdam.

We thank Michael Slepian for sharing his stimuli and instructions, and for many helpful suggestions on the procedure of Experiment 2. We thank Christiaan Tieman, Elske van den Broek, Ivonne Canits, Tjeerd Schoonderwoerd, Patricia Arends, Nic Stötefalk, Alma Fiere, Femke Stolte, and Jurriaan Manintveld for their help with Experiment 3. Finally, we thank Samantha Bouwmeester and Steven Raaijmakers for their helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Diane Pecher, Psychology Department, T12-33, Erasmus University Rotterdam, Postbus 1738, 3000 DR Rotterdam, The Netherlands. E-mail: pecher@fsw.eur.nl

participants to recall either a big secret or a small secret. They speculated that thinking of a big secret would lead participants to experience a heavier burden than thinking of a small secret. Based on a study by Bhalla and Proffitt (1999), who found that participants who carried a heavy backpack estimated a hill as steeper than participants who did not carry a backpack, Slepian et al. tested the idea that keeping a big secret (heavy burden) would lead participants to estimate a hill as steeper than keeping a small secret (light burden). After participants recalled their big or small secret, they rated four photographs of objects and places, the fourth of which was a hill for which they estimated the slant. The results showed that participants who recalled a big secret estimated the hill about 13 degrees steeper than participants who recalled a small secret, thereby confirming Slepian et al.'s hypothesis and leading the authors to conclude that participants experienced their secret as an embodied physical burden.

Given that this effect is of both theoretical importance, for example, for theories on embodiment of cognition, and practical relevance in terms of its implications for the relation between secrets, physical burdens, and perception of space, it is of utmost importance to establish the robustness of the effect. Interestingly, in a recent study, Slepian, Masicampo, and Ambady (2014) replicated the effect of secrets on the estimation of hill slant (both in direction and in effect size) and also obtained an effect of secrets on explicit distance estimation. By contrast, LeBel and Wilbur (2014) recently failed to replicate the effect of secrets on hill slant estimation, even though they used exactly the same procedure as in the original experiment by Slepian et al. (2012), and tested larger samples than those in the original studies. More recently, Slepian, Camp, and Masicampo (2015) suggested that the effect of the secret size manipulation might be small. They replicated the original Slepian et al. (2012) effect of secret size on hill slant estimation in one experiment, but failed to replicate the effect in another experiment. They further investigated the role of preoccupation in two additional experiments by asking participants to recall a preoccupying or a nonpreoccupying secret, and obtained effects on hill slant estimation in both experiments. Given these mixed findings, another exact replication of Slepian et al. (2012) seems mandatory to establish the robustness of the effect. In addition, assuming that we successfully replicate the results, we also made an attempt to investigate whether demand characteristics might have played a role, as this was suggested as a potential mechanism for the relation between actual physical burdens and perceptual estimations (Durgin et al., 2009; Durgin, Klein, Spiegel, Strawser, & Williams, 2012; Shaffer, McManama, Swank, & Durgin, 2013; Woods, Philbeck, & Danoff, 2009). We also performed a meta-analysis of the results reported in the literature thus far and our own findings in order to get an estimate of the effect.

Next to the effect on hill slant estimation, Slepian et al. (2012) presented three other results that showed a relation between secrets and measures of fatigue. In our view, of the four studies reported by Slepian et al., the effect on hill slant estimation obtained in their Experiment 1 provided the strongest test of the burden of secrecy theory. Secret "weight" was manipulated experimentally and an attempt was made to mask the purpose of the experiment by presenting the secret question and the rating task as different studies. In their Experiment 2, they reasoned that the metaphorical burden of holding a secret would cause people to overestimate distances. They asked participants to throw a beanbag at a target while thinking of a secret. Slepian et al. speculated that the experienced physical burden would lead participants to overestimate distances, another indirect measure of physical fatigue. Consequently, participants in the big-secret condition were predicted to throw the beanbag further than participants in the small-secret condition. Their results confirmed their hypothesis: The big-secret group threw the beanbag about 16 cm further than the small-secret group. Interpretation of the effect on beanbag throwing is more complicated because it requires the additional assumption that differences in perceived distance lead to differences in throwing distance, a relation that has been questioned by research on the effect of visual illusions on throwing behavior (Cañal-Bruland, Voorwald, Wielaard, & van der Kamp, 2013, Experiment 1). On the other hand, measuring throwing behavior might be less susceptible to top down biases than asking for explicit judgments (Firestone, 2013; Firestone & Scholl, 2014).

Two other results that Slepian et al. (2012) presented are also consistent with the idea that secrecy leads to experiencing physical burdens, but seem less convincing because alternative explanations are plausible. In Experiment 3, the "weight" of the secret was not manipulated; rather, participants who were known to have committed an infidelity were asked how much they thought of the secret and how much it bothered them (as a measure of the burden). Subsequently, they rated the energy and effort involved in three physical and three nonphysical tasks. The results showed a correlation between how much participants were burdened by their secret and their estimates of the physical tasks, but not their estimates of the nonphysical tasks. Although these results are consistent with the theory of secrets as physical burdens, the results are correlational in nature and, consequently, open to alternative explanations, such as a general tendency of participants to give higher or lower responses, or some other third factor that might affect both the extent to which someone is bothered by infidelity and the estimated effort of physical tasks, such as the degree to which a person feels bothered by things in general. In Experiment 4, Slepian et al. (2012) asked gay men to conceal their sexual orientation ("big secret") or their extraversion ("small secret"), and after some other tasks asked them to help move book stacks. The participants who concealed their sexual orientation moved fewer stacks than those concealing their extraversion. However, the number of book stacks they moved might indicate several other things than just physical fatigue, such as how well they liked the experimenter. Perhaps gay men do not like being asked to conceal their sexual orientation, and this changed their willingness to help the experimenter and move stacks of books. Thus, although the results of all four experiments are consistent with Slepian et al.'s claim that secrets are experienced as physical burdens, in our opinion, Experiments 1 and 2 provided the most convincing support for this claim.

In the present study, we thus aimed to replicate the results of Experiments 1 (hill slant estimation) and 2 (beanbag throwing) reported by Slepian et al. (2012) to investigate whether these effects were robust. We followed the procedure of the original experiments as closely as possible and used a sample size that was sufficiently large to have substantial statistical power to reject the null hypothesis of no effect. That is, their study had a sample size of 40. In their Experiment 1, they found an effect size (Cohen's d) of 0.76. To obtain a statistically significant effect of similar effect size with a power of .95 and an alpha of .05 (using a two-tailed independent samples t test), we would need a sample size of 92

(G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007). To be on the safe side, we decided to test a total of 100 participants. In Experiments 1 and 2, like Slepian et al., we tested participants from Amazon's Mechanical Turk participant pool, we used the same stimuli, and used a procedure very similar to that reported by Slepian et al. After participants finished the experiment, we administered a questionnaire to investigate to what extent they had been aware of a relation between recalling their secret and estimating hill slant, and if this induced them to use that relation strategically in their estimate of hill slant. Because they were tested online (just as in Slepian et al., 2012, Experiment 1), there was no opportunity for experimenter effects. We report all experiments on the effects of secrecy on hill slant estimation (and beanbag throwing) that we have done.

Experiment 1

Method

Participants. One hundred participants (50 female) were recruited through Amazon's Mechanical Turk and received \$0.50 for their participation. The participants were randomly assigned to the big-secret and small-secret groups with the restriction that there were 50 participants in each group. On Amazon's Mechanical Turk, we set the worker requirements such that only people who were in the United States could participate. The mean age of the participants was 34.7 years (range 18 to 71). Fifty-eight participants reported having an associate's degree or higher. Two participants reported many distractions during the study, and four reported a lot of noise.

Stimuli. To manipulate importance of the secret, participants were asked to answer the following question about either a big or a small secret (from Slepian et al., 2012):

> Before we ask you to rate objects and places, we are also interested in the psychology of secrets. We ask you to think about a big [small] secret that you have, one that you are purposefully keeping as a secret.
>
> Without revealing specific details about your secret, we are curious what it pertains to. Please write about your big [small] secret in the provided box, revealing as much or as little details as you'd like. As a reminder we do not have a way to link your responses to Mechanical Turk IDs so this is completely anonymous.

Michael Slepian was very generous to send us his materials. Therefore, the four color pictures that we used are identical to those used by Slepian et al. (2012). A photograph of a hill was used to elicit hill slant estimation from the participant. Three other pictures were used as fillers: a table, a water bottle, and a park.

A funneled questionnaire (not used in the original Slepian et al., 2012, study) was subsequently administered to measure awareness of a relation between recalling a secret and estimating hill slant, and to assess potential strategic use of such relation. The questionnaire consisted of these questions:

1. While working on these 2 studies (the study on secrets and the picture rating study), was there anything special that you noticed?

2. While working on these 2 studies, what did you think their purpose was?

3. Did any of the pictures in the second study stand out? If yes, which picture, and why did it stand out?

4. Did you notice any relation between the first study and the second study? If yes, which relation?

5. Did anything you did in the first study affect your response to the second study? If yes, how?

6. How did you go about rating the pictures? did you have a particular strategy or goal?

7. Did thinking about your secret affect how you estimated the hill slant? If yes, how?

Procedure. The procedure of the experiment itself was similar to that used by Slepian et al. (2012, Experiment 1). Participants were recruited for two short studies. On the start page of the experiment, they were told the second study would be on rating objects and places. Before they started, they were asked to answer the question about their secret. After they had provided their answer, they were informed that the first study was finished and the second study would start. Subsequently, they were asked to rate the four pictures which were each presented in fixed order on a separate page. For each rating, a text box was provided in which the participant typed the rating. They rated the sturdiness of the table on a 7-point scale from 1 (*not sturdy*) to 7 (*very sturdy*). They rated the durability of the water bottle on a 7-point scale from 1 (*not durable*) to 7 (*very durable*). They gave their estimate of the outside temperature of the park by typing degrees in Fahrenheit. Finally, they estimated the hill slant in degrees. Only numbers (i.e., not letters) were allowed as answers. After each answer, participants had to click an arrow button to continue to the next item. If participants provided a nonallowed answer, a red box appeared on the screen to remind the participant to provide a valid answer.

After rating the four pictures, participants answered the questions of the funneled questionnaire. Below each question, a text box was provided in which participants could type their answer. Finally, they provided their gender, age, native language, education level, and how quiet or noisy their surroundings were.

Results

For each participant, we recorded the four ratings to the four pictures. No participants were excluded from the data analysis (following Slepian et al., 2012).[1] To facilitate comparison with Slepian et al.'s data, the average z scores for the four picture rating tasks for both secret condition groups are shown in Figure 1. The actual averages are displayed in Table 1. A two-tailed independent samples t test showed no significant difference between the groups on hill slant ratings, $t(98) = 0.41$, $p = .68$, Cohen's $d = 0.08$, 95% confidence interval (CI) $[-0.31, 0.47]$. We also calculated the ratio of Bayesian probabilities for the null and alternative hypotheses, given the data. This provided us with a more informative measure of the probability that either hypothesis is true, rather than

[1] No exclusion criteria are mentioned in Slepian et al. (2012). In later publications (Slepian et al., 2014, 2015), participants were excluded based on different sets of criteria. Because our aim was to replicate Slepian et al. (2012), we followed their method.

the *p* value, which only gives the probability of the observed or more extreme data given that the null hypothesis is true (Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). The corresponding JZS Bayes Factor is equal to 6.00, indicating that the data provide 6 times more evidence for the null hypothesis than for the alternative hypothesis. We should note that the JZS Bayes Factor for Slepian et al.'s original finding is estimated to be 2.64 times in favor of the alternative hypothesis, providing only weak evidence for the effect (see Rouder et al., 2009, and Wagenmakers, 2007, for a detailed explanation of the JZS Bayes Factor).

The groups also did not differ in their ratings of the control pictures, $t(98) = 0.60$, $p = .55$, for the bottle durability ratings, $t(98) = 1.44$, $p = .15$, and for the park temperature ratings, although the difference for table sturdiness ratings approached significance, $t(98) = 1.89$, $p = .06$. Participants who recalled a small secret rated the table as sturdier than participants who recalled a big secret. This effect was not reported by Slepian et al. (2012) and unexpected by us.

The answers to the questionnaire indicated that participants were not aware at all of the relation between secrets as burdens and slant estimations.[2] One participant from the big-secret group reported to have thought about the hill and how difficulties can be an uphill battle (the secret of this participant pertained to a personal difficulty). The other 99 participants reported seeing no relation whatsoever between recalling secrets and judging hill slant, and many seemed to find the question odd.

We reported our results to Michael Slepian. He was very helpful and went through our experiment and then suggested a couple of changes to our experimental procedure. These changes concerned aspects of the procedure that differed from the original experiment but were not specified in the Method section of Slepian et al. (2012) (see the following Method section for details). We followed his suggestions in Experiment 2 and registered the Methods and Analyses plan on the Open Science Framework website (https://osf.io/hmfxt/). The general advantage of preregistering experiments is that it limits researcher degrees of freedom and selective reporting (cf. Simmons, Nelson, & Simonsohn, 2011; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

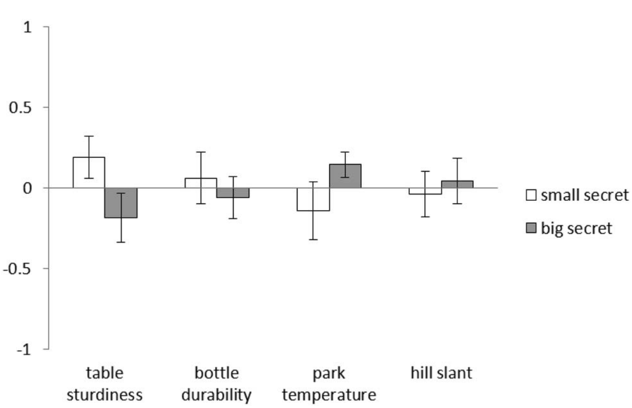

Figure 1. Ratings for the four pictures (*z* scores) in Experiment 1. Error bars represent standard error of the mean.

Table 1
Average Ratings for the Four Pictures in Experiments 1 and 2

		Table sturdiness		Bottle durability		Park temperature		Hill slant	
	N	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Experiment 1									
Small secret	50	5.48	0.15	4.94	0.21	67.00	2.44	40.98	2.94
Big secret	50	5.04	0.18	4.78	0.17	70.86	1.13	42.70	2.96
Experiment 2									
Small secret	50	5.14	0.18	5.10	0.15	68.08	1.23	38.44	2.44
Big secret	50	5.60	0.16	4.92	0.14	67.84	1.85	42.08	2.39

Experiment 2

Method

Participants. One hundred participants (44 female) were recruited through Amazon's Mechanical Turk with the same assignment to conditions as in Experiment 1. The mean age was 32.6 years (range 19 to 72). Fifty-four participants reported having an associate's degree or higher. Two participants reported many distractions during the study, and two reported a lot of noise.

Stimuli and procedure. The stimuli and procedure were identical to those of Experiment 1, with the following exceptions. To exclude participants who had already participated in a similar study, we asked the question, "Have you ever done another study like this one, in which you had to recall a secret and estimate the hill slant?" after the funnel questionnaire. Two participants answered positively to this question and were replaced. No other participants were excluded from the data analysis. The experiment started with a welcome screen informing them that they would be asked to rate objects and places. Then a new page was displayed asking about the secret. After the secret question, we did not mention that the first study was finished or that the second study would start. A new page simply asked the participants to rate the pictures. Note that this procedure might increase the likelihood that participants noticed a link between the two parts of the experiment. Rather than on separate pages, all four pictures were presented on the same page and participants could freely scroll up and down. For the ratings of the table and the bottle, we used a multiple choice format so that participants could click on their choice. For the hill slant estimate, we added the instruction (following Michael Slepian's suggestion), "Note: 0 degrees is a flat surface, while 90 degrees is a vertical surface. Your estimation should be in between those two numbers." If participants did not enter a number between 0 and 90, an error message was displayed and participants could not continue until they corrected their answer.

Results

For each participant, we recorded the four ratings of the four pictures. The average *z* scores for the four picture-rating tasks for

[2] Participants did seem to give thoughtful answers. For example, to the first question, many participants answered with a full sentence, such as "I didn't notice anything special," even though a simple "no" would have answered the question. This suggests that participants did take the task seriously.

both secret condition groups are shown in Figure 2. The actual averages are displayed in Table 1. A two-tailed independent samples *t* test showed no significant difference between the groups on hill slant ratings, $t(98) = 1.06$, $p = .29$, Cohen's $d = 0.21$, 95% CI [−0.18,0.60]. The corresponding JZS Bayes Factor is equal to 3.84, indicating that the data provide almost 4 times more evidence for the null hypothesis than for the alternative hypothesis.

The groups also did not differ in their ratings of the control pictures, $t(98) = 0.88$, $p = .38$ for the bottle durability ratings, $t(98) = 0.11$, $p = .91$, for the park temperature ratings, although again the difference for table sturdiness ratings approached significance, $t(98) = 1.94$, $p = .06$. The pattern was opposite to that in Experiment 1; participants who recalled a small secret rated the table as less sturdy than participants who recalled a big secret.

The answers to the questionnaire indicated that the large majority of participants were not aware at all of the relation between secrets as burdens and slant estimations. When asked whether any of the pictures stood out, one participant from the small-secret group reported that the hill reflected the difficulties related to the secret. When asked explicitly whether recalling the secret affected their hill slant estimation, this participant and five others (three form the big-secret group and two from the small-secret group) stated that their estimation might have been influenced. Two participants (one from the big-secret group) speculated that they might have underestimated the slant, and the other four speculated that they might have overestimated the slant. Thus, as in Experiment 1, only a very small portion of the participants noticed a possible relation, and most of these only mentioned this when we explicitly asked them about this relation.

Meta-Analysis: The Effect of Secrecy on Hill Slant Estimates

Next to our failed replications in Experiments 1 and 2, we are aware of six more (almost exact replication) experiments that have been published in the meantime (i.e., Slepian et al., 2012, 2014, 2015; LeBel & Wilbur, 2014). To obtain a better estimate of the effect of the burden of secrecy on hill slant estimates, we conducted a meta-analysis on these studies (calculated with ESCI software; Cumming, 2012). The two experiments reported here are the only data that we gathered using this paradigm. We searched

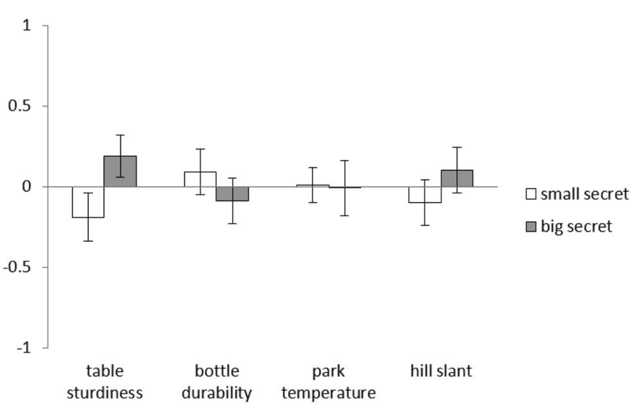

Figure 2. Ratings for the four pictures (*z* scores) in Experiment 2. Error bars represent standard error of the mean.

for additional replications on Scopus, Web of Science, Open Science Framework, and PsychFileDrawer.org. One additional study was found on PsychFileDrawer.org by Perfect, Moon, and Nelson (2014), which we included in the meta-analysis. They did not replicate the original finding in a sample of 312 participants. We included only the "recall" condition from Slepian et al. (2014), because the "reveal" condition included different instructions. In cases in which sample sizes were not reported separately for each condition, we assumed that participants were distributed evenly over conditions. Because there was significant heterogeneity between studies (see statistical results), we performed a random-effects analysis. Figure 3 shows the difference between the big-secret and small-secret conditions (or big- and no-secret conditions in Slepian et al., 2014) for each experiment with 95% confidence intervals. The meta effect is a nonsignificant difference of 3.28 degrees in hill slant estimation between the big-secret and small-secret group, which includes zero in its 95% confidence interval. The analysis also showed significant and substantial heterogeneity across studies, $Q_8 = 22.15$, $p = .0046$, $I^2 = 63.9\%$. Inspection of Figure 3 suggests that Experiment 1 from Slepian et al. (2012), Experiment 1 from Slepian et al. (2014), and Experiment 2 from Slepian et al. (2015) differ in some crucial way from the other experiments. Usually when a meta-analysis reveals heterogeneity across studies, one should look for possible moderators. In this case, however, it is difficult to identify such moderators. After all, the studies used the exact same materials, almost identical procedures, and tested samples from the same participant pool. Only the study by Slepian et al. (2014) differed from the others because they used a no-secret condition instead of a small-secret condition. However, as their results are very similar to those in Slepian et al. (2012), it does not seem that this difference can explain the heterogeneity. The only substantive difference between the studies is their sample size. The studies by Slepian et al. (2012) and Slepian et al. (2014) used smaller samples (about 20 participants per condition) than the other studies (between 45 and 120 participants per condition). Hence, it is conceivable that the replication studies resulted in a better estimate of the effect than the Slepian et al. (2012, 2014) studies (Cohen, 1994; Cumming, 2012; Halsey, Curran-Everett, Vowler, & Drummond, 2015). Differences in sample size alone, however, cannot explain heterogeneity. Apart from the sample sizes, the replication attempts were exact replications, and therefore it seems unlikely that the differences in results can be explained by a moderator variable. In any case, the meta-analysis shows a small overall effect that is not significantly different from zero, providing no support for the idea that thinking of a secret affects hill slant estimation. However, because the effect on hill slant estimation was not the only evidence in favor of the interpretation offered by Slepian et al. (2012), we next tried to replicate another finding that was reported as support for the secrecy as physical burden theory.

Experiment 3

With this experiment, we aimed to replicate Experiment 2 of Slepian et al. (2012). In their Experiment 2, they reasoned that the metaphorical burden of holding a secret causes people to overestimate distances. Slepian et al. measured perceived distance indirectly by asking participants to throw a beanbag at a target. Their

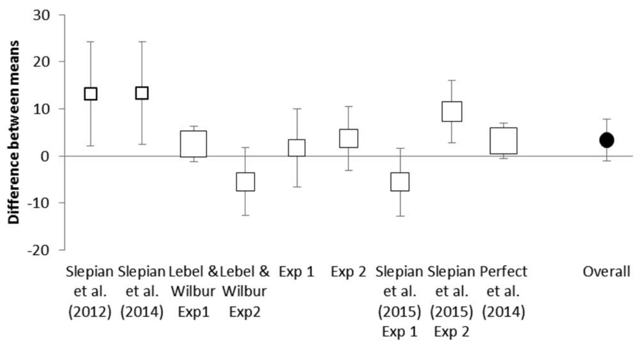

Figure 3. Results of the meta-analysis. The points show the difference between mean hill slant estimates in the big- and small-secret conditions for each experiment separately (big- and no-secret condition in Slepian et al., 2014). The error bars show the 95% confidence interval for each difference. The size of the marker indicates the weight of each experiment into the meta-analysis. The Overall point is the estimated effect size based on Cumming (2012).

results confirmed their hypothesis: They found that people who recalled a big secret overthrew the beanbag more than participants who recalled a small secret. In our replication attempt, we tried to follow Slepian et al.'s procedure as much as possible and registered the Methods and Analyses plan on the Open Science Framework website (https://osf.io/qc3pz/) before we started collecting data.

Method

Participants. A group of 118 participants was recruited at Erasmus University Rotterdam to participate for course credit or a snack. Slepian et al. (2012) had a sample size of 36 in their Experiment 2 and found an effect size (Cohen's *d*) of 0.67. To obtain a statistically significant effect of similar effect size with a power of .95 and an alpha of .05 (using a two-tailed independent samples *t* test), we would need a sample size of 118 (G*Power 3; Faul et al., 2007).

Stimuli. The importance of the secret was manipulated in the same way as in the previous experiments. For the throwing task, we used a beanbag that was made from soft cloth with roughly the shape of a tetragon (four-sided polygon). It was filled with beads with a total weight of 71 g. Two sides were approximately 7.5 cm, and the two other sides were approximately 10.0 cm. Four strips of Velcro were glued on the outside so that it would stick to the carpeted floor, decreasing the likelihood that the beanbag would tumble or bounce after landing (for similar precautions, see Cañal-Bruland et al., 2013). A small white dot at the midpoint of each side served as a reference point for the distance measurement. Three strips of white tape on the dark carpet floor of the lab were used to indicate the position of the foremost foot of the participant. The target was indicated by a cross of two strips of tape (10 cm each) that were stuck on the floor such that the center of the cross was at a distance of 265 cm from the participant position. In Slepian et al.'s (2012) study, the target was a steel colander (M. L. Slepian, personal communication, February 20, 2014). We chose to use a cross as a target, because it allowed a more precise measurement of distance between the target and the landing position of the beanbag. Note that the colander results in a discontinuous measure of distances because all bean bags that land in the colander will be assigned the exact same distance score (i.e., zero), and small distances away from the target cannot be obtained. This is likely a major factor contributing to the observation that distance data obtained by Slepian et al. were non-normal. Moreover, beanbags that land on the rim of the colander can fall inside or outside the colander, thereby increasing variability of the measurement. For this reason, we decided to use a marking on the floor rather than a colander as the target.

Procedure. At the start of the experiment, participants were informed that their responses would be saved anonymously on the computer. Participants were then asked to answer the question about their big or small secret on the computer. The computer program randomly assigned the participant to either the big- or the small-secret condition. The experimenter was blind to the hypotheses of the study and to the condition (big vs. small secret) of each participant. After starting, the experiment program the experimenter sat at a position from where the computer screen could not be seen. Participants were instructed to not discuss the question or their answer with the experimenter. They typed their answer, revealing as much or as little information as they wanted. When they were finished typing the answer they, pressed "Enter" to remove their answer from the screen so that the experimenter could not see it. Subsequently, the participant was asked to throw the beanbag at the cross. They were instructed to aim such that the beanbag would land at the middle of the cross. After their throw, participants were dismissed.

Slepian et al. (2012) did not specify how distance from the target was measured, for example, in *x* (i.e., horizontal error) and/or *y* (longitudinal error) coordinates, or in one straight line from beanbag to target (i.e., radial errors). We measured both *x* and *y* coordinates and considered the *y*-distance as the dependent measure to test whether distance thrown differed between conditions, because the *y*-axis (longitudinal error) was aligned with the distance between the participant and the target (see also Cañal-Bruland et al., 2013). The experimenter measured the *y* and *x* coordinates of the position of the midpoint of the beanbag in relation to the midpoint of the target. Positions to the left of the target and positions closer to where the participant had been standing were recorded as negative distances; positions to the right and beyond the target were recorded as positive distances. In cases in which the beanbag moved after landing, the final position was measured, and a note was made that the landing position was different from the final position.

Results

No participants were excluded from this data analysis. A two-tailed independent samples *t* test showed no significant difference between the groups on distance thrown ($M = 0.15$ cm, $SE = 4.25$ and $M = 7.74$ cm, $SE = 5.18$ for the small- and big-secret group, respectively), $t(116) = 1.13$, $p = .26$, Cohen's $d = .21$, 95% CI [−0.15, 0.57]. The corresponding JZS Bayes Factor is equal to 3.84 in favor of the null hypothesis. Thus, we did not replicate Slepian et al.'s finding that participants who recalled a big secret threw a beanbag further than participants who recalled a small secret. Our data provided almost four times more evidence for the null than for the alternative hypothesis.

An exploratory analysis showed that the groups also did not differ in the distance from the target on the *x*-axis ($M = -0.15$ cm, $SE = 1.69$ and $M = -0.38$ cm, $SE = 2.13$ for the small- and big-secret group, respectively), $t(116) = 0.08$, $p = .93$, Cohen's $d = .01$, 95% CI [−0.38, 0.35]. For 24 participants, the beanbag moved after landing. Analyses that excluded the data from these participants also yielded no significant effects on *x*-distance, $t(91.9) = 0.64$, $p = .52$, Cohen's $d = .13$, 95% CI [−0.28, 0.53], and *y*-distance, $t(92) = 0.62$, $p = .54$, Cohen's $d = .13$, 95% CI [−0.27, 0.54].

General Discussion

In three experiments, we failed to replicate findings by Slepian et al. (2012) that recalling a secret leads to steeper hill slant estimates and overthrowing a beanbag aimed at a target. Because, next to our exact replications of the hill slant experiments (Experiments 1 and 2), other exact replications exist (LeBel & Wilbur, 2014; Perfect et al., 2014; Slepian et al., 2014, 2015), we were able to perform a meta-analysis based on nine data sets, representing nine attempts of the exact same experiment, and hence were able to get a better estimate of the effect of secrets on hill slant estimation. As illustrated in Figure 3, the results of the meta-analysis also fail to support the idea that thinking of a big secret affects hill slant estimation.

Although in Experiments 1 and 2, thanks to Michael Slepian providing us with all the necessary information, we used almost identical procedures as Slepian et al. (2012), in Experiment 3, we had to slightly deviate from the original, as we did not have sufficiently specific information to allow for exact replication.[3] Notably, in all our experiments, the sample sizes were much larger than those of the original study, and this resulted in a power of at least .95 for each experiment to find an effect of the same size as that reported by Slepian et al. Still, taken together—that is, the findings of our three experiments and the meta-analysis—there was no indication that recalling a secret leads to steeper hill slant estimation or farther beanbag throws. Our study, like those of many others, shows the value of independent exact replications of original findings (Doyen, Klein, Pichon, & Cleeremans, 2012; Harris, Coburn, Rohrer, & Pashler, 2013; Klein et al., 2014; Pashler, Coburn, & Harris, 2012; Pashler & Harris, 2012; Shanks et al., 2013).

The failure to replicate these effects and the nonsignificant overall effect of our meta-analysis shows that the burden of secrecy on perceptual judgments such as hill slant estimates and beanbag throwing distances is not robust. We will refrain from drawing more general theoretical conclusions. As Slepian et al. (2015) suggest, the particular manipulation to ask participants about a big or small secret might not have been strong enough. They suggest that level of preoccupation with a secret is a better predictor of how much a secret is experienced as a physical burden. It is possible that their new manipulation will show more robust effects. At present, we have no explanation why the first two experiments on the effect of secret size showed large effects on hill slant estimation and the five follow-up experiments from independent labs did not. It is possible that the lower sample sizes in earlier studies were more susceptible to random variability, whereas the later studies with larger sample sizes provided more accurate estimates of the population effect size (Cohen, 1994; Cumming, 2012; Halsey et al., 2015). Given these findings, researchers should be cautious when interpreting Slepian et al.'s (2012) findings, and are advised to use different methods if they are interested in investigating the relation between secrecy and perceptual judgments.

[3] Our decision to use a more precise target (a cross on the floor) than Slepian et al. (who used a large steel colander) arguably resulted in a more accurate measure of distance from the target. We do not know to what extent our beanbag differed from the one used by Slepian et al. The beanbag we used was a practice beanbag for jugglers. It was quite light and thus unlikely to be perceived as an additional physical burden. Its shape and the added Velcro reduced the likelihood of moving after landing. Because there is no obvious reason why the shape of the neutral target or the properties of the beanbag would interact with the secret manipulation, it seems unlikely that these features caused the differences between our result and that of Slepian et al.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645. http://dx.doi.org/10.1146/annurev.psych.59.103006.093639

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1076–1096. http://dx.doi.org/10.1037/0096-1523.25.4.1076

Boot, I., & Pecher, D. (2010). Similarity is closeness: Metaphorical mapping in a conceptual task. *Quarterly Journal of Experimental Psychology (2006), 63,* 942–954. http://dx.doi.org/10.1080/17470210903134351

Boot, I., & Pecher, D. (2011). Representation of categories: Metaphorical use of the container schema. *Experimental Psychology, 58,* 162–170. http://dx.doi.org/10.1027/1618-3169/a000082

Cañal-Bruland, R., Voorwald, F., Wielaard, K., & van der Kamp, J. (2013). Dissociations between vision for perception and vision for action depend on the relative availability of egocentric and allocentric information. *Attention, Perception, & Psychophysics, 75,* 1206–1214. http://dx.doi.org/10.3758/s13414-013-0476-3

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition, 106,* 579–593. http://dx.doi.org/10.1016/j.cognition.2007.03.004

Cienki, A. (2005). Image schema and gesture. In B. Hampe (Ed., *From perception to meaning* (pp. 421–442). Berlin, Germany: Mouton de Gruyter. http://dx.doi.org/10.1515/9783110197532.5.421

Cohen, J. (1994). The earth is round (p <. 05). *American Psychologist, 49,* 997–1003. http://dx.doi.org/10.1037/0003-066X.49.12.997

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition, 110,* 412–431. http://dx.doi.org/10.1016/j.cognition.2008.11.016

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE, 7,* e29081. http://dx.doi.org/10.1371/journal.pone.0029081

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review, 16,* 964–969. http://dx.doi.org/10.3758/PBR.16.5.964

Durgin, F. H., Klein, B., Spiegel, A., Strawser, C. J., & Williams, M. (2012). The social psychology of perception experiments: Hills, backpacks, glucose, and the problem of generalizability. *Journal of Experi-*

mental Psychology: Human Perception and Performance, 38, 1582–1595. http://dx.doi.org/10.1037/a0027805

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Firestone, C. (2013). How "paternalistic" is spatial perception? Why wearing a heavy backpack doesn't-and couldn't-make hills look steeper. *Perspectives on Psychological Science, 8,* 455–473. http://dx.doi.org/10.1177/1745691613489835

Firestone, C., & Scholl, B. J. (2014). "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science, 25,* 38–46. http://dx.doi.org/10.1177/0956797613485092

Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space time metaphors. *Language and Cognitive Processes, 17,* 537–565. http://dx.doi.org/10.1080/01690960143000317

Gibbs, R. W. J. (1994). *The poetics of mind: Figurative thought, language, and understanding*. New York, NY: Cambridge University Press.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1–19.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle p value generates irreproducible results. *Nature Methods, 12,* 179–185. http://dx.doi.org/10.1038/nmeth.3288

Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLoS ONE, 8,* e72467. http://dx.doi.org/10.1371/journal.pone.0072467

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "Many Labs" replication project. *Social Psychology, 45,* 142–152. http://dx.doi.org/10.1027/1864-9335/a000178

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: Chicago University Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York, NY: Basic Books.

LeBel, E. P., & Wilbur, C. J. (2014). Big secrets do not necessarily cause hills to appear steeper. *Psychonomic Bulletin & Review, 21,* 696–700. http://dx.doi.org/10.3758/s13423-013-0549-2

Machery, E. (2007). Concept empiricism: A methodological critique. *Cognition, 104,* 19–46. http://dx.doi.org/10.1016/j.cognition.2006.05.002

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15,* 243–247.

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS ONE, 7,* e42510. http://dx.doi.org/10.1371/journal.pone.0042510

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7,* 531–536. http://dx.doi.org/10.1177/1745691612463401

Pecher, D., Boot, I., & Van Dantzig, S. (2011). Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. H. Ross (Ed., *The psychology of learning and motivation* (pp. 217–248). Burlington, MA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-385527-5.00007-3

Perfect, H., Moon, A., & Nelson, L. (2014, February 5). *Big secrets make hills look steeper*. Retrieved from http://www.PsychFileDrawer.org/replication.php?attempt=MTc2

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Shaffer, D. M., McManama, E., Swank, C., & Durgin, F. H. (2013). Sugar and space? Not the case: Effects of low blood glucose on slant estimation are mediated by beliefs. *i-Perception, 4,* 147–155. http://dx.doi.org/10.1068/i0592

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., . . . Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PLoS ONE, 8,* e56515.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

Slepian, M. L., Camp, N. P., & Masicampo, E. J. (2015). Exploring the secrecy burden: Secrets, preoccupation, and perceptual judgments. *Journal of Experimental Psychology: General, 144,* e31–e42. http://dx.doi.org/10.1037/xge0000052

Slepian, M. L., Masicampo, E. J., & Ambady, N. (2014). Relieving the burdens of secrecy: Revealing secrets influences judgments of hill slant and distance. *Social Psychological and Personality Science, 5,* 293–300. http://dx.doi.org/10.1177/1948550613498516

Slepian, M. L., Masicampo, E. J., Toosi, N. R., & Ambady, N. (2012). The physical burdens of secrecy. *Journal of Experimental Psychology: General, 141,* 619–624. http://dx.doi.org/10.1037/a0027598

Teuscher, U., McQuire, M., Collins, J., & Coulson, S. (2008). Congruity effects in time and space: Behavioral and ERP measures. *Cognitive Science, 32,* 563–578. http://dx.doi.org/10.1080/03640210802035084

van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science, 32,* 579–590. http://dx.doi.org/10.1080/03640210802035365

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14,* 779–804. http://dx.doi.org/10.3758/BF03194105

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

Woods, A. J., Philbeck, J. W., & Danoff, J. V. (2009). The various perceptions of distance: An alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1104–1117. http://dx.doi.org/10.1037/a0013622

Zanolie, K., van Dantzig, S., Boot, I., Wijnen, J., Schubert, T. W., Giessner, S. R., & Pecher, D. (2012). Mighty metaphors: Behavioral and ERP evidence that power shifts attention on a vertical dimension. *Brain and Cognition, 78,* 50–58. http://dx.doi.org/10.1016/j.bandc.2011.10.006

Received February 12, 2015
Revision received May 19, 2015
Accepted May 21, 2015 ■