

## Michael Slepian – Job Market Paper

My job talk will span a number of papers, and thus multiple papers are attached below as my job market paper, a paper on judging trustworthiness, a paper on the consequences of secrecy, and a paper on detecting deception.

Slepian, M. L., & Ames, D. R. (in press). Internalized impressions: The link between apparent facial trustworthiness and deceptive behavior is mediated by targets' expectations of how they will be judged. *Psychological Science*.

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*, 31-42.

Slepian, M. L., Young, S. G., Rutchick, A. M., & Ambady, N. (2013). Quality of professional players' poker hands is perceived accurately from arm motions. *Psychological Science*, *24*, 2335-2338.

# Internalized Impressions: The Link Between Apparent Facial Trustworthiness and Deceptive Behavior Is Mediated by Targets' Expectations of How They Will Be Judged



Michael L. Slepian and Daniel R. Ames

Department of Management, Columbia University

Psychological Science

1–7

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797615594897

pss.sagepub.com



## Abstract

Researchers have debated whether a person's behavior can be predicted from his or her face. In particular, it is unclear whether people's trustworthiness can be predicted from their facial appearance. In the present study, we implemented conceptual and methodological advances in this area of inquiry, taking a new approach to capturing trustworthy behavior and measuring targets' own self-expectations as a mediator between consensual appearance-based judgments and the trustworthiness of targets' behavior. Using this novel paradigm to capture 900 observations of targets' behavior (as trustworthy or untrustworthy), we found that face-based judgments predicted trustworthiness. We also found that this effect was mediated by targets' expectations of how other people would perceive them and by their intentions to act in accordance with those expectations. These results are consistent with an internalized-impressions account: Targets internalize other people's appearance-based expectations and act in accordance with them, which leads facial-appearance-based judgments to be accurate.

## Keywords

social cognition, social perception, face perception, open data, open materials

Received 2/21/15; Revision accepted 6/16/15

For decades, scholars have debated whether a person's behavior can be predicted from his or her face. In particular, can judgments of individuals' faces predict their trustworthiness? One possibility (the *essentialist-impressions* account) is that genetic expression leads to both untrustworthy-looking faces and untrustworthy behavior. Such a correspondence would resemble the once-popular but now discredited claims of physiognomy. An alternative possibility (the *misleading-impressions* account) is that although people reliably agree on which faces look untrustworthy or trustworthy (Rule, Krendl, Ivcevic, & Ambady, 2013; Todorov, 2008), those judgments show no predictive validity (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015) because there is no reliable correspondence with actual trustworthiness (see Todorov & Porter, 2014). We believe that a third possibility exists: A lifetime of being treated as trustworthy or untrustworthy as a

result of one's appearance may lead one to internalize these expectations and act in accordance with them, which eventually results in appearance-based accuracy (our *internalized-impressions* account).

Recent research presents a mixed picture: Some work has provided evidence suggestive of accuracy in face-based judgments of trustworthiness (Stirrat & Perrett, 2010), yet other research has found no such accuracy (Rule et al., 2013). We see two reasons why prior work might have obtained mixed results. First, some past scholarship has revolved around single, relatively extreme,

## Corresponding Author:

Michael L. Slepian, Department of Management, Columbia Business School, Columbia University, Uris Hall, 3022 Broadway, New York, NY 10027

E-mail: michael.slepian@columbia.edu

and heterogeneous behaviors (e.g., targets' criminal history). Instead, we think a relationship between faces and behavior bears testing in a paradigm with three features: (a) an interactive face-to-face context, (b) a constrained set of trust-related behaviors, and (c) multiple observations of potentially trustworthy or untrustworthy behavior. Second, prior studies have not measured psychological mediators between face-based judgments and behavior. We expect that any link between facial appearance and behavior would be mediated by psychological variables, such as targets' expectations.

In this article, we introduce a new paradigm that addresses both of these points. We created a novel research design to capture multiple instances of trust-related behaviors in a face-to-face context. We focused on a single class of behaviors: People (targets) repeatedly chose to make and defend a true or false claim to different counterparts, a false claim (i.e., deceptive, untrustworthy behavior) entailing the chance for private material gains but also imposing costs on the counterpart. Past paradigms have typically placed participants either in computer-mediated interactions or in asocial contexts. We expected that giving targets the option to lie (for potential gain) to a live face-to-face counterpart would elicit meaningful variance in trustworthiness of behavior not captured in prior research. Our paradigm also constrained behavior into a dichotomous choice (to lie or tell the truth) made repeatedly (in 10 independent interactions with different counterparts), yielding a clean and reliable measure of trustworthiness.

In addition, before interactions, we assessed targets' metaperceptions of their own trustworthiness (targets predicted how frequently they would be trusted by their counterparts) and their predictions of how frequently they would act in a trustworthy manner. Measuring these variables allowed us to test a potential route through which judgments of the face might predict behavior. That is, given that people reach high consensus on which faces look trustworthy or untrustworthy, individuals with trustworthy- or untrustworthy-looking faces should have a lifetime of experience of being treated like trustworthy or untrustworthy people. Such experiences would range from the banal (e.g., whether strangers smile at them) to the life-changing (e.g., whether they get particular jobs). We believe that the cumulative effect of such treatment is likely to be powerful, as implied by work on self-fulfilling prophecies (Rosenthal, 1994) and the looking-glass self (Cooley, 1902).

## Method

We first measured participants' apparent facial trustworthiness by having independent judges rate photographs of them. Two days later, participants were told that they would be interacting with other participants and were

asked to report how they expected to be judged by their counterparts and how they expected themselves to act. They subsequently interacted as both targets and counterparts in a novel mixed-motive game. As targets, they repeatedly chose whether to behave in a trustworthy manner (i.e., to tell the truth) or in an untrustworthy manner (i.e., to lie) to a series of 10 different counterparts. We predicted that if ratings of facial trustworthiness showed an ability to predict trustworthy behavior, then this link would be mediated by targets' expectations, which would be consistent with an internalized-impressions account.

## Participants

Our participant pool consisted of all the M.B.A. students in a particular course; sample size was determined by the number of students who were enrolled in the course and present on the day the study was conducted ( $N = 118$ ). Ninety-five participants' faces were photographed, but 5 of these participants did not provide self-expectation judgments. Thus, the final sample consisted of 90 students (65.60% male; mean age = 28.10 years,  $SD = 1.76$ ).

## Mixed-motive game

Participants played a two-person game in which each person privately drew a random card (labeled "high" or "low"). In a face-to-face interaction, they then freely chose to claim that the card was "high" or "low," independently of the card drawn, thereby choosing to tell the truth or to lie.

The mixed-motive paradigm was implemented in two testing sessions (accounting for testing session in our analyses did not alter the results). Targets and counterparts were randomly paired within sessions, with no repeat pairings. Each participant was randomly paired with 10 other participants in succession. In each interaction, a given participant served as both a target and a counterpart. As the target, the participant decided whether to tell the truth or to lie to the counterpart; as the counterpart, the participant decided whether to trust the target. After both members of a pair had drawn a card, and independently and privately decided whether to tell the truth or to lie, they then claimed that their cards were either high or low (thereby telling the truth or not). They next spent 2 to 3 min attempting to persuade one another of their trustworthiness. After this persuasion phase, each participant independently and privately judged whether he or she trusted that the target was telling the truth. Once both parties had made their private judgments about one another, both revealed whether they had lied or told the truth, and whether they had trusted their fellow participant.

**Table 1.** Payoff Table for a Single Round of Mixed-Motive Game

	Partner's trust in recipient			
	Partner (counterpart) trusts participant (target)		Partner (counterpart) distrusts participant (target)	
	Participant is trustworthy (+10 for participant)	Participant is untrustworthy (+20 for participant)	Participant is trustworthy (0 for participant)	Participant is untrustworthy (0 for participant)
Participant's trust in partner				
Participant (counterpart) trusts partner (target)				
Partner is trustworthy (+10 for participant)	+20 [+20]	+30 [-10]	+10 [+10]	+10 [+10]
Partner is untrustworthy (-20 for participant)	-10 [+30]	0 [0]	-20 [+20]	-20 [+20]
Participant (counterpart) distrusts partner (target)				
Partner is trustworthy (0 for participant)	+10 [+10]	+20 [-20]	0 [0]	0 [0]
Partner is untrustworthy (0 for participant)	+10 [+10]	+20 [-20]	0 [0]	0 [0]

Note: A given participant's payoff for a given round was determined by two components: The participant's score as the counterpart (i.e., whether the participant trusted his or her partner) and the participant's score as a target (i.e., whether the participant's partner trusted the participant). As counterparts, participants earned points for trusting their partners when their partners were trustworthy and lost points for trusting their partners when their partners were untrustworthy. As targets, participants earned points for getting their partners to trust them and gained no points if their partners did not trust them. Each cell contains two values; the number outside the brackets shows the total payoff for the person identified as the participant, and the number inside the brackets shows the total payoff for that participant's partner.

We describe the outcome of the decision to tell the truth or to lie as trustworthy or untrustworthy *behavior*. The only way for targets to earn points on the basis of their behavior was to earn trust. If a target chose to tell the truth and was trusted by his or her counterpart, the target earned a modest payoff in the game (10 points). If a target chose to lie and was trusted by his or her counterpart, the target earned double that payoff (20 points). If a target's counterpart did not trust the target, the target earned nothing (0 points). In game-theory terms, lying was a weakly dominant strategy (see Kohlberg & Mertens, 1986).

We describe the outcome of the decision to trust or distrust a target as a *judgment*. Counterparts' judgments about whether to trust targets also had payoffs. If a target's counterpart correctly trusted a target who told the truth, the counterpart received a modest reward (10 points). Incorrectly trusting a target who lied entailed a significant loss (-20 points). If a counterpart decided to not trust the target, he or she neither earned nor lost points (0 points). Thus, counterparts' payoffs for their trust judgments were contingent on whether a target was telling the truth or lying. In game-theory terms, there was no dominant strategy for judgments (if participants assumed that lies and truths were equally likely but undiagnosable). The three top performers in each session

received prizes (a \$50 Amazon gift card for the top performer and \$25 Amazon gift cards to the second- and third-place performers).

The payoff table shown in Table 1 summarizes participants' payoffs for each possible outcome. In game-theory terms, lying is a weakly dominant strategy; if we assume that each player recognizes that lying is weakly dominant for the other, lying combined with distrusting is the Nash equilibrium (see Kohlberg & Mertens, 1986). However, we did not expect that most interactions would result in mutual lying and distrust. Note that in each of the 10 rounds, each participant was both a target (choosing how to behave) and a counterpart (judging a fellow participant). These choices were made separately, with behavioral decisions made privately before mutual discussion (i.e., behavioral choices were made before the interaction) and judgments made privately after discussion. Thus, the calculation of the payoff matrix should not be taken to suggest that one decision was contingent on the other; they were independent. We designed this paradigm with the expectation that it would produce variance in the frequency with which people would behave in a trustworthy manner (i.e., some people would choose to lie frequently, and others would choose to tell the truth frequently); this variance was critical for testing our predictions.

In this study, we examined whether individuals' behavior (not their judgments of other people) could be predicted from their faces, and thus the focus of our analysis was predicting targets' behavior toward others, not counterparts' judgments of others. Our two central questions were (a) whether the apparent facial trustworthiness of targets (based on ratings from an independent set of judges) predicted how they behaved (i.e., their frequency of telling the truth) in the 10-round game and (b) whether that link was mediated by the targets' expectations reported before the game.

### Measures

Two days before the game, during a video-based exercise occurring in the students' class, photographs were taken of the players. In the photographs, the players assumed a neutral expression. No specific rationale was given for taking photographs other than that it was part of the video-based class exercise. That is, these photographs were taken outside the context of the mixed-motive game; participants were not aware of the game or its rules when the photographs were taken. We recruited independent judges ( $n = 30$  per rating) via Amazon.com's Mechanical Turk. These judges used a 7-point scale to rate each face for trustworthiness (1 = *not at all trustworthy*, 7 = *very trustworthy*;  $M = 4.127$ ,  $SD = 0.616$ , 95% confidence interval, or CI = [3.998, 4.256],  $\alpha = .887$ ), attractiveness (1 = *not at all attractive*, 7 = *very attractive*;  $M = 3.264$ ,  $SD = 0.805$ , 95% CI = [3.096, 3.433],  $\alpha = .948$ ), *babyfacedness* (1 = *not at all babyfaced*, 7 = *very babyfaced*;  $M = 3.332$ ,  $SD = 0.811$ , 95% CI = [3.162, 3.502],  $\alpha = .918$ ), and apparent affect (1 = *appears angry*, 7 = *appears happy*;  $M = 3.962$ ,  $SD = 0.830$ , 95% CI = [3.788, 4.135],  $\alpha = .952$ ).

During the game session, after the process and payoffs had been described to the players, they predicted how frequently (0%–100%) they would (a) act in a trustworthy manner ( $M = 52.6\%$ ,  $SD = 31.8$ , 95% CI = [45.9%, 59.2%]) and (b) be trusted ( $M = 55.2\%$ ,  $SD = 17.5$ , 95% CI = [51.6%, 58.9%]).<sup>1</sup> The game yielded 10 observations of each participant's behavior as a target, the focal measure for the current study (coded as 1 = trustworthy, 0 = untrustworthy;  $M = 62.4\%$ ,  $SD = 35.5$ , 95% CI = [55.0%, 69.9%]), and 10 observations of each participant's judgment (of other targets) as a counterpart (coded as 1 = trust, 0 = distrust;  $M = 61.7\%$ ,  $SD = 18.4$ , 95% CI = [57.8%, 65.5%]).

Given the nature of the student sample, players had varying levels of familiarity with each other before the game, and familiarity could influence behavior. To assess and control for this possibility, we provided the players with a list of their 10 counterparts' names 4 days after the game and asked them to rate how familiar they had been with each counterpart before the game (1 = *not at all*, 2

= *slightly*, 3 = *somewhat*, 4 = *mostly*, 5 = *highly*;  $M = 2.264$ ,  $SD = 1.515$ , 95% CI = 2.164, 2.363]).

### Results

To maximize statistical power, we fitted outcomes to a linear mixed-effects model examining all 900 trust judgments and 900 trust behaviors, controlling for random variance (from targets, counterparts, and round of the mixed-motive game). All analyses were conducted in the R software environment (Version 3.1.1; R Development Core Team, 2014). We used the R package lme4 to implement mixed-effects models (Bates, Maechler, Bolker, & Walker, 2015). In calculating  $p$  values, we used the R package lmerTest to run lme4 models through Kenward-Rogers approximation tests to estimate the degrees of freedom (these estimated degrees of freedom scale the model estimates to best approximate the  $F$  distribution, and thus can be fractional and differ slightly across tests; Kuznetsova, Brockhoff, & Christensen, 2013). R package confint was used to implement Wald-tests to calculate 95% CIs.

### Perception

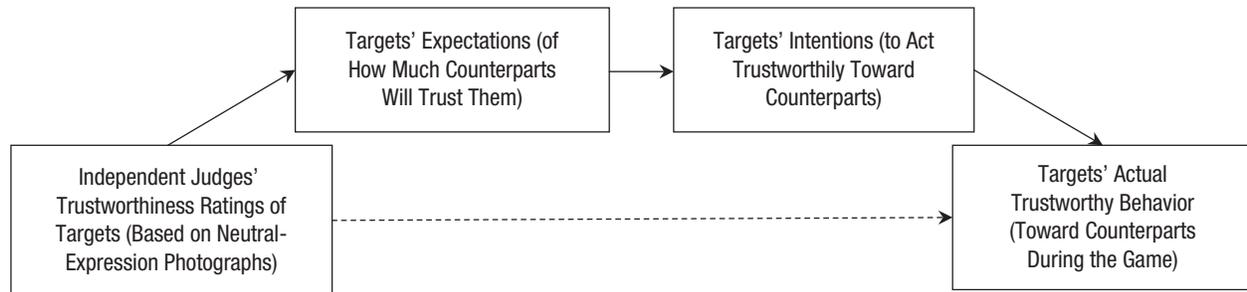
Photograph-based trustworthiness judgments predicted how often counterparts chose to trust targets after the live interactions,  $b = 0.074$ , 95% CI = [0.022, 0.126],  $SE = 0.027$ ,  $t(79.76) = 2.77$ ,  $p = .007$ . Counterparts had access to a multitude of cues in the face-to-face interactions, yet their trustworthiness judgments corresponded to independent ratings of the targets' faces. If this effect emerged in the absence of accuracy, our results would fit with the misleading-impressions account noted earlier. However, our internalized-impressions account suggests that trustworthiness judgments could be accurate, and we next tested for such accuracy.

### Accuracy

Photograph-based trustworthiness judgments predicted how often targets actually behaved in a trustworthy manner toward counterparts,  $b = 0.124$ , 95% CI = [0.007, 0.242],  $SE = 0.060$ ,  $t(87.99) = 2.077$ ,  $p = .041$ . This finding is consistent with the notion that facial trustworthiness predicts trustworthy behavior.

### Other predictors

Before turning to our central prediction (concerning how targets' expectations might mediate the link between their facial trustworthiness and trustworthy behavior), we considered a number of other possible predictors and alternative explanations. Accuracy of trustworthiness



**Fig. 1.** Mediation model of the predicted mediation between targets' facial trustworthiness (as judged by the independent raters) and the targets' behavior, as mediated by the targets' expectations of how they would be judged and how they would act.

judgments might derive from other features of targets' neutral-expression faces (e.g., emotional resemblances; Hehman, Flake, & Freeman, 2015; Sacco & Hugenberg, 2009; Zebrowitz, 2011). Even though the photographs were taken outside the context of and before the game, perhaps targets still somehow conveyed their trustworthy intentions (e.g., by smiling slightly). Our findings were inconsistent with this suggestion: Ratings of attractiveness corresponded with counterparts' trustworthiness judgments,  $b = 0.042$ , 95% CI = [0.001, 0.083],  $SE = 0.021$ ,  $t(81.62) = 1.996$ ,  $p = .049$ , but other variables did not—apparent affect:  $b = 0.007$ , 95% CI = [−0.034, 0.047],  $SE = 0.021$ ,  $t(80.89) = 0.331$ ,  $p = .741$ ; babyfaceness:  $b = 0.025$ , 95% CI = [−0.018, 0.067],  $SE = 0.022$ ,  $t(85.17) = 1.147$ ,  $p = .255$ ; target's gender (0 = male, 1 = female),  $b = 0.060$ , 95% CI = [−0.009, 0.129],  $SE = 0.035$ ,  $t(80.95) = 1.712$ ,  $p = .091$ .

Ratings of babyfaceness predicted trustworthy behavior,  $b = 0.103$ , 95% CI = [0.014, 0.192],  $SE = 0.045$ ,  $t(87.99) = 2.272$ ,  $p = .026$ , but other variables did not—attractiveness:  $b = 0.022$ , 95% CI = [−0.070, 0.114],  $SE = 0.047$ ,  $t(87.99) = 0.477$ ,  $p = .635$ ; apparent affect:  $b = 0.065$ , 95% CI = [−0.023, 0.153],  $SE = 0.045$ ,  $t(87.99) = 1.44$ ,  $p = .153$ ; target's gender:  $b = 0.105$ , 95% CI = [−0.048, 0.259],  $SE = 0.078$ ,  $t(87.99) = 1.347$ ,  $p = .182$ .

It is also possible that the level of familiarity between players influenced the trustworthiness of their behavior. The greater targets' familiarity with their counterparts, the more likely the targets were to behave in a trustworthy manner toward those counterparts,  $b = 0.041$ , 95% CI = [0.025, 0.057],  $SE = 0.008$ ,  $t(835.08) = 5.027$ ,  $p < .001$ . Critically, when we accounted for targets' familiarity with

their counterparts, greater perceived trustworthiness, as judged from the target's face by the independent raters, was still associated with greater trustworthy behavior,  $b = 0.123$ , 95% CI = [0.007, 0.240],  $SE = 0.060$ ,  $t(88.02) = 2.070$ ,  $p = .041$ .

### Mediation by self-expectations

We next turned to the mediation prediction implied by the internalized-impressions account. We expected that the link between face-based judgments and behavior would be mediated by the targets' expectations of how they would be judged and how they would act (Fig. 1). Results of regression analyses were consistent with this prediction, revealing that photograph-based trustworthiness judgments predicted targets' self-expectations of how often they would be trusted,  $b = 0.073$ ,  $SE = 0.029$ , 95% CI = [0.015, 0.131],  $t(88) = 2.500$ ,  $p = .014$  (i.e., the targets anticipated other people's naive expectations). These expectations about being trusted, in turn, predicted how often targets intended to act in a trustworthy manner,  $b = 0.685$ ,  $SE = 0.180$ , 95% CI = [0.328, 1.042],  $t(88) = 3.809$ ,  $p < .001$  (i.e., targets internalized these expectations and intended to act consistently with them). Targets' intentions of acting in a trustworthy manner, in turn, predicted the trustworthiness of their actual behavior,  $b = 0.805$ ,  $SE = 0.082$ , 95% CI = [0.641, 0.968],  $t(88) = 9.785$ ,  $p < .0001$  (see Table 2 for zero-order correlations of these variables).<sup>2</sup> A formal bootstrapped mediation analysis (5,000 iterations), in which attractiveness, apparent affect, babyfaceness, and target's gender were entered as covariates, confirmed this mediational path, mean indirect effect = .0512,  $SE = .0349$ , 95%

**Table 2.** Zero-Order Correlations Between the Main Variables in the Mediation Model

Variable	Targets' facial trustworthiness	Targets' expectations	Targets' intentions
Targets' expectations	.26	—	—
Targets' intentions	.27	.38	—
Targets' behavior	.22	.21	.72

Note: All correlations are significant,  $p \leq .05$ .

CI = [.0062, .1548]; excluding the covariates did not alter statistical significance, mean indirect effect = .0363, *SE* = 0.0195, 95% CI = [.0095, .0915].

## Discussion

In a novel paradigm featuring trusting behavior in face-to-face interactions, trustworthiness ratings of targets (based on neutral-expression photographs) corresponded with targets' behavioral trustworthiness. We found that targets seemed to have an awareness of how people would judge them, and they internalized these expectations and behaved in accordance with them. Such internalized impressions are similar to self-fulfilling prophecies (Rosenthal, 1994), although we suggest that the effects of the internalized impressions are somewhat broader and more cumulative. Much work on self-fulfilling prophecies has focused on a single context, such as whether a teacher's expectations (e.g., that a student is particularly intelligent) can bring about outcomes consistent with those expectations (e.g., improvement in the student's performance; Rosenthal, 1994). We believe that these effects can play out over longer periods as well, as implied by work on the looking-glass self (Cooley, 1902). Our participants' (somewhat accurate) expectations of how other people would judge them in a particular game corresponded to strangers' ratings of photographs of participants' faces, which suggests that these accurate metaperceptions may be derived from a range of contexts across a lifetime of treatment.

Participants' behavior seemed to live up, or down, to how they expected to be judged. Those participants who thought they would be trusted were more likely to be trustworthy, and those who thought they would be distrusted were more likely to be untrustworthy. This finding is inconsistent with an opportunistic-deception account (Olekals & Smith, 2009), which implies that people who expect to be trusted would be likely to exploit that trust rather than comply with it.

An essentialist-impressions account would not necessarily imply our mediation results. Such an account emphasizes the genetic, inherent correspondence between facial features and behavior rather than the role of targets' expectations posited by the internalized-impressions perspective. It might be possible to further discriminate between essentialist-impressions and internalized-impressions accounts by testing for the causal order that we posit (e.g., manipulating targets' expectations about being trusted, and examining their behavioral trustworthiness).

Features of our new paradigm may have allowed us to observe predictive accuracy of face-based judgments that was not apparent in past work (e.g., Rule et al., 2013). In this study, targets repeatedly confronted a basic but

constrained question: "Do I act untrustworthy toward this person for a better chance to win this game?" Placing targets in identical, constrained, highly social contexts, with repeated observations, is likely to increase the robustness of any judgment-behavior link, but that constraint also limits generalizability. The type of photograph-based judgments we used might not predict cheating on a test, for example, but might predict behavior in other mixed-motive social contexts.

In sum, the consensus that people achieve in rating targets' faces corresponds to how people interact with those targets. Targets are aware of the expectations implicit in that consensus. We propose that targets come to internalize such expectations, acting in accordance with them, and thus those initial judgments, over time, become accurate judgments.

## Author Contributions

Both authors contributed to designing the study, analyzing the data, and writing the manuscript.

## Acknowledgments

We thank Patrick Bergemann, Amie Blocker, Ashli Carter, Jae Cho, Jinseok Chun, Drew Jacoby-Senghor, Alice Lee, Ashley Martin, Pilar Opazo, Anastasia Usova, Stacey Sasaki and Abbie Wazlawek for their assistance in the current work. We also thank the Social Cognitive & Neural Sciences Lab at New York University for helpful feedback on this work, and Rodney Atkins and Nicholas Rule for helpful comments on an early version of the manuscript.

## Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## Open Practices



All data and materials have been made publicly available via Open Science Framework. Data can be accessed at <https://osf.io/kfdca/>, and materials can be accessed at <https://osf.io/bnqd5>. The complete Open Practices Disclosure for this article can be found at <http://pss.sagepub.com/content/by/supplemental-data>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/> and <http://pss.sagepub.com/content/25/1/3.full>.

## Notes

1. We also measured participants' predictions of how much they would trust their fellow participants, how often their fellow participants would act trustworthy, and how accurately they would judge their fellow participants. However, these judgments would be made by the participant in the role of

counterpart (rather than target) and are thus outside the scope of the current investigation (which focuses on targets' behaviors, not counterparts' judgments). We also asked participants, in their role as targets, to predict how accurately they would be judged by their counterparts, yet this measure of participants' perceived transparency does not distinguish between behavior as trustworthy or untrustworthy and is thus also outside the scope of the current investigation.

2. Targets' expectations predicted their behavior. Thus, one might wonder whether individual differences in meta-accuracy (i.e., the *difference* between the amount of trust participants expected to receive and the amount that they actually received) would predict behavior. However, meta-accuracy did not correspond to how often targets told the truth,  $b = -0.120$ ,  $SE = 0.185$ , 95% CI = [-0.487, 0.248],  $t(88) = -0.647$ ,  $p = .519$  (nor did the absolute value of the difference,  $b = 0.103$ ,  $SE = 0.265$ , 95% CI = [-0.424, 0.629],  $t(88) = 0.387$ ,  $p = .700$ ).

## References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using 'Eigen' and S4 (Version 1.1-7) [Computer software]. Retrieved from <http://cran.r-project.org/package=lme4>
- Cooley, C. H. (1902). *Human nature and the social order*. New York, NY: Charles Scribner's Sons.
- Hehman, E., Flake, J. K., & Freeman, J. B. (2015). Static and dynamic facial cues differentially impact the consistency of social evaluations. *Personality & Social Psychology Bulletin*, *41*, 1123–1134.
- Kohlberg, E., & Mertens, J.-F. (1986). On the strategic stability of equilibria. *Econometrica*, *54*, 1003–1037.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests in linear mixed effects models (Version 2.0-11) [Computer software]. Retrieved from <http://cran.r-project.org/web/packages/lmerTest/>
- Olekalns, M., & Smith, P. L. (2009). Mutually dependent: Power, trust, affect and the use of deception in negotiation. *Journal of Business Ethics*, *85*, 347–365.
- R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, *3*, 176–179.
- Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology*, *104*, 409–426.
- Sacco, D. F., & Hugenberg, K. (2009). The look of fear and anger: Facial maturity modulates recognition of fearful and angry expressions. *Emotion*, *9*, 39–49.
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*, 349–354.
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*, 208–224.
- Todorov, A., Olivola, C., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545.
- Todorov, A., & Porter, J. M. (2014). Misleading first impressions different for different facial images of the same person. *Psychological Science*, *25*, 1404–1417.
- Zebrowitz, L. A. (2011). Ecological and social approaches to face perception. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The Oxford handbook of face perception* (pp. 31–50). Oxford, England: Oxford University Press.

# Exploring the Secrecy Burden: Secrets, Preoccupation, and Perceptual Judgments

Michael L. Slepian  
Columbia University

Nicholas P. Camp  
Stanford University

E. J. Masicampo  
Wake Forest University

Recent work suggests that secrecy is perceived as burdensome. A secrecy–burden relationship would have a number of consequences for cognitive, perceptual, social, and health psychology, but the reliability of these influences, and potential mechanisms that support such influences are unknown. Across 4 studies, the current work examines both the reliability of, and mechanisms that support, the influence of secrecy processes upon a judgment that varies with diminished resources (i.e., judgments of hill slant). The current work finds that a manipulation of secret “size” fails to reliably predict judged hill slant, whereas measurement and manipulation of preoccupation with a secret does reliably predict judged hill slant. Moreover, these effects are found to be mediated by judged effort to keep the secret, consistent with a resource-based mechanism of the burdens of secrecy.

*Keywords:* secrecy, perception, coping, resources

Secrecy involves the active concealment of information from others. Active inhibition is depleting (Cricher & Ferguson, 2014), and individuals describe keeping a secret as being burdensome, or physically weighing one down (e.g., “carrying a secret”; Slepian, Masicampo, Toosi, & Ambady, 2012). These processes might have consequential outcomes for well-being. Indeed, keeping secrets brings negative consequences for mental and physical well-being (Cole, Kemeny, Taylor, & Visscher, 1996; Cole, Kemeny, Taylor, Visscher, & Fahey, 1996; Rodriguez & Kelly, 2006). Given close links of secrecy with physical and mental processes (see Pennebaker, 1989), one topic that has received recent attention is that of the bodily based perceptual consequences of secrecy and guilt. For example, secrecy elicits guilt (Frijns & Finkenauer, 2009; Vangelisti, 1994), and a growing body of literature demonstrates that people link guilt with the sensation of carrying weight (Day & Bobocel, 2013; Kouchaki, Gino, & Jami, 2014). For example, sensations of weight enhance feelings of guilt (Kouchaki et al., 2014), and feelings of guilt enhance sensations of weight (Day & Bobocel, 2013). Being forgiven, however, reduces feelings of burden, as demonstrated by reduced judgments of hill slant (Zheng, Fehr, Tai, Narayanan, & Gelfand, 2014), as does feelings of being understood (Oishi, Schiller, & Gross, 2013), and being sup-

ported (Schnall, Harber, Stefanucci, & Proffitt, 2008), or being affirmed (reducing judgments of distance; Shea & Masicampo, 2014). As these examples suggest, carrying information alone, without others’ support and understanding, (i.e., secrecy) can influence perceived burden. Holding secrets can affect people in the same way that carrying physical weight does—people who carry consequential secrets act and make judgments as if they are physically burdened (Slepian et al., 2012).

Despite this growing body of literature, which suggests links between physical and mental processes in the domain of secrecy, a number of important questions remain. Given that secrecy has been associated with diminished physical health (Cole, Kemeny, Taylor, & Visscher, 1996; Cole, Kemeny, Taylor, Visscher, & Fahey, 1996), it is critical to understand how secrecy might actually influence perceived burden to shed light on its health consequences. Additionally, it is important to know how robust prior demonstrated links are. Demonstrating that secrecy can lead to perceived burden has a number of consequences for cognitive, perceptual, social, and health psychology (e.g., with implications for judgments of physical space, social cognition and interpersonal relations, and coping with secrecy). It is thus important to understand how reliable the links between secrecy and experienced burden might be, and how best to operationalize secrecy in a manner that allows for insights into how secrecy influences bodily based perceptual judgments. The current work focuses on judgments of hill slant to examine these questions as such judgments have been reliably linked to experiences of burden.

One method used in prior work to examine whether secrecy can be burdensome has asked participants to recall “big” versus “small” secrets (based on random assignment), and then provide a number of judgments, including the judgment of hill slant.

---

This article was published Online First February 9, 2015.

Michael L. Slepian, Management Division, Columbia University; Nicholas P. Camp, Department of Psychology, Stanford University; E. J. Masicampo, Department of Psychology, Wake Forest University.

Correspondence concerning this article should be addressed to Michael L. Slepian, Management Division, Columbia University, 3022 Broadway, Uris Hall, New York, NY 10027. E-mail: [michael.slepian@columbia.edu](mailto:michael.slepian@columbia.edu)

Judgments of hill slant have been reliably linked to experiences of physical burden in prior work (for a review see Proffitt, 2006). That is, judgments are scaled to the ability to interact with the external environment (Cole & Balciotis, 2013; Eves, 2014; Proffitt, 2006; Witt, Proffitt, & Epstein, 2004). Thus, if one feels that he or she has fewer resources available to act upon the external environment, that environment is judged as more challenging. Judging the environment as more challenging leads to judgments of that environment as being more forbidding and extreme (e.g., judging hill slant as steeper with fewer perceived resources to scale the hill).<sup>1</sup> In one of four studies, Slepian and colleagues (2012) found that recalling “big,” relative to “small,” secrets led to steeper judgments of hill slant, consistent with the hypothesis that secrecy is perceived as physically burdensome. A recent article, however, failed to find an influence of a manipulation that asks participants to recall “big,” relative to “small,” secrets on judgments of hill slant (LeBel & Wilbur, 2014). The goal of the current work was to examine how reliably secrecy processes influence judgments that vary with perceived burden. To accomplish this goal, the current authors examined the replicability of the original work, with a special focus on understanding if, when, and how secrecy can be burdensome.

Comparing any replication attempt to the original work, considerations can be made at two different levels of analysis: the *operationalization* level and the *phenomenon* level. These different levels of analysis are content-general (i.e., can apply to any domain of interest), but we discuss them here in the domain of the burdens of secrecy.

Beginning first with the lower level, the operationalization level, we can wonder whether a failure to replicate the effect of recalling “big” and “small” secrets on judgments of hill slant simply suggests that the relationship between this exact manipulation and this exact dependent measure is nonexistent or weak. That is, perhaps secrets are indeed burdensome, but the precise manner in which secrets are being manipulated, or burden is measured, is suboptimal—the operationalizations are imprecise. If the link between the specific operationalization of burdensome secrets and the specific operationalization of burden-like outcomes is nonexistent, we should not expect an exact replication of the original study to be successful (i.e., we would see a failure to replicate the effect of the “big” vs. “small” secret recall manipulation on judgments of hill slant). Use of other, more precise operationalizations, however, may yield support for the notion that secrets produce burden-like outcomes.

Moving to the phenomenon level, we can wonder whether a failure to replicate indicates that the broader phenomenon (i.e., of secrecy leading to perceived burden) does not exist. If so, then secrets will not be burdensome regardless of the particular methods used (e.g., the exact manipulation, the exact dependent measure, or both).

With respect to the current topic, a failure to replicate one operationalization of the hypothesis that secrecy is burdensome (LeBel & Wilbur, 2014) presents an important question: Does a failure to replicate the effect of recalling “big” versus “small” secrets on judgments of hill slant question the replicability of this particular methodology (i.e., does it speak to replicability at the operationalization level), or does it question whether secrecy pro-

cesses, more broadly, influence judgments of hill slant (i.e., does it speak to replicability at the phenomenon level)?

As reviewed above, a growing body of work that uses a diverse array of dependent measures (judgments of hill slant, judgments of distance, jumping height, judgments of the weight of objects, judgments of body weight) suggests that the burden of secrets leads to outcomes that resemble physical burden (Day & Bobocel, 2013; Slepian et al., 2012; Slepian, Masicampo, & Ambady, 2014; Susewind, Christandl, & Hoelzl, 2013; see also Critcher & Ferguson, 2014; Kouchaki et al., 2014). In contrast, elements that relieve the burden of secrecy, such as feeling understood, affirmed, supported, in control, or forgiven, reduce burden-consistent outcomes (reducing judgments of hill slant, distance, and felt weight; Lee & Schnall, 2014; Oishi et al., 2013; Schnall et al., 2008; Shea & Masicampo, 2014; Zheng et al., 2014).

Thus, a converging body of evidences demonstrates that having to carry information alone, without support, is burdening, whereas being relieved (e.g., by being understood, affirmed, supported, or forgiven) is unburdening. In addition to the reliability of conceptually similar measures, it is also important to know the reliability of specific dependent measures. We focus on one such dependent measure in the current work, judgments of hill slant. Thus, more specifically we can wonder whether a failure to find an influence of recalling “big” versus “small” secrets on judgments of hill slant suggests that this specific operationalization is suboptimal, or whether secrecy processes, more generally, do not reliably influence judgments of hill slant (a measure that tracks the experience of diminished resources). A simple way to test these interpretations against one another is to conduct an exact replication of the original study, thereby capturing the exact methodology to examine its replicability, but within the same study, add an alternative operationalization of the phenomenon. This would provide a simultaneous test, within the same participant sample, of two operationalizations of the more general hypothesis. Bonferroni correction can then be applied to account for the fact that examining

<sup>1</sup> *Judgments* of the external environment reflect one’s perceived capability to interact with the environment (see Witt, 2011). Whether *visual perception* of the environment can be altered by these processes, however, has been debated (cf. Firestone, 2013; Proffitt, 2013). A full discussion of the debate regarding judgments versus perception is outside of the scope of the current paper, but we wish to mention that the focus of the current work is on judgments of hill slant. We do not make claims about the *visual perception* of hill slant. Additionally, researchers (cf. Durgin et al., 2009; Proffitt, 2009) have debated the processes by which, specifically, wearing a heavy backpack influences hill slant judgments. Yet, this debate is confined to the backpack manipulation and judgments of hill slant (other work also debates the effect of throwing a heavy object; Woods, Philbeck, & Danoff, 2009). A variety of other physical burden variables (reduced action potential from advanced age, from fatigue, from low fitness, from pain) also predict judgments of hill slant (and distance), and these have not been suggested to be consequences of demand characteristics (Cole & Balciotis, 2013; Eves, 2014; Sugovic & Witt, 2013; Witt, Proffitt, & Epstein, 2004; Witt et al., 2009; see Proffitt, 2006). Thus, the debate surrounding the mechanisms by which heavy backpacks influence judgments of hill slant does not cast doubt on the relationship between physical burden and judgments, more generally, only specifically studies that use the backpack manipulation. Thus, this particular debate (which is far from settled; cf. Durgin, Klein, Spiegel, Strawser, & Williams, 2012; Proffitt, 2013; Witt & Sugovic, 2013) does not have any a priori relevance to the current work given that heavy backpacks are not used as manipulations in the current studies, nor did any participants intuit the experimental hypotheses or present any indication of demand effects in funneled debriefing.

replicability at the operationalization and phenomenon level within the same sample provides two opportunities to find a significant effect.

### Burdensome Secrets: “Size” Versus Preoccupation

The current work takes a new approach to studying the secrecy burden, examining the role of how preoccupied one is by a secret, as well as how much effort one believes is needed to keep a secret. This approach, combined with examining the reliability of prior operationalizations, sheds light on the mechanisms of the burdens of secrecy, extending current theory on how secrecy processes impinge on bodily based responses. First, if secrecy does influence judgments of hill slant, then the specific operationalization of burdensome secrets as “big” rather than “small” in a recent failed replication (LeBel & Wilbur, 2014) may be suboptimal. The assumption underlying this operationalization is that “big” secrets are more personally relevant and influential than “small” ones and hence should be more burdensome (Slepian et al., 2012). However, we suspect that the “big” versus “small” distinction does not precisely capture the difference between personally influential and noninfluential secrets as was originally presupposed. That is, these terms may be underspecified. For example, consider the secret of infidelity. If one has come to terms with an infidelity, is accepting of it, and thus no longer upset by it, it may not be a subjectively influential secret, but one may nevertheless describe it as a “big” secret as it is conventionally treated as such. Indeed, simply committing infidelity is not burdensome, but rather it is the extent to which one treats one’s infidelity as consequential and spends time thinking about it that determines how burdensome it is (Slepian et al., 2012).

Thus, it might be that *preoccupation* with a secret is a more reliable indication of a secret’s gravity. Indeed, research on concealable stigma supports this suggestion. Among individuals who possess concealable stigmatized identities (e.g., sexual orientation, mental illness), the importance of their identity, the frequency with which they think about the identity, and the anticipated negative consequences of disclosing that identity mediate the effects of these identities on psychological distress (Quinn & Chaudoir, 2009). Additionally, the more individuals who committed an infidelity thought about their infidelity and were bothered by it, the more they perceived physical tasks as requiring effort (Slepian et al., 2012), suggesting that it is not the norm-based “size” of a secret that determines burden, but rather it might be the extent to which one is preoccupied with one’s secret that determines perceived burden. Asking participants to recall “big” versus “small” secrets, therefore, might not systematically lead to the recall of secrets that are personally influential or not, respectively. In other words, these terms may only weakly elicit influential and noninfluential secrets, thereby leading such a manipulation to sometimes succeed, but not always. We therefore examined whether more specific (and more carefully defined) measures and manipulations of the recall of preoccupying secrets would produce differences in judgments of hill slant, as this operationalization more closely captures the phenomenon of interest. We thus measured how preoccupied participants were by their secret. Preoccupation was assessed by three criteria: a) the amount a person thinks about his or her secret, b) how much one feels that the secret personally affects him or her, and c) how much it

bothers him or her. That is, preoccupation is not defined in this framework as mere cognitive accessibility (cf. Lane & Wegner, 1995), but rather represents the subjective motivational engagement one has with the secret.

### Role of Effort

A body of evidence suggests that preoccupation as defined above should deplete resources for acting upon the environment (see Cole & Balcetis, 2013). That is, when preoccupied by a secret, one is devoting personal resources toward that secret. This increased preoccupation with the secret might suggest to the secret holder that increased effort is needed to keep the secret (and thus less effort is available for other pursuits). In other words, just as physical burdens lead to more extreme judgments of the environment by increasing the effort required to interact with it (e.g., Cole & Balcetis, 2013; Eves, 2014; Proffitt, 2006; Schnall, Zadra, & Proffitt, 2010; Witt et al., 2004), perhaps so does preoccupation with secrets.

This notion suggests a natural mediator to test between preoccupations with one’s secret and judgments of the environment as more forbidding and extreme, increased effort needed to keep the secret. When sizing up how steep a staircase is, for example, it will seem steeper if one has to carry a bag of groceries up the stairs, rather than an empty bag. The increased effort required by walking upstairs with something heavy, will lead the staircase to seem more forbidding, that is, steep (see Cole & Balcetis, 2013). In other words, the effort required to walk up a staircase in this example, mediates the relationship between how heavy the object one must carry is, and how steep the staircase looks. Likewise, we hypothesized that the effort required to keep a secret will mediate the relationship between how preoccupied one is by the secret and how steep a hill is judged. Specifically, we tested the hypothesis that preoccupying secrets (relative to nonpreoccupying secrets) will be judged as more effortful to keep, which will predict the judgment that the environment is more forbidding (i.e., a hill is more steep).

### The Present Studies

Across four studies, we test both the reliability of, and mechanisms that support, the influence of secrecy upon judgments of hill slant. First, we test whether a previously used manipulation of secret “size” can reliably predict judged hill slant, or whether measurement of preoccupation with a secret is superior in that regard (Studies 1 and 2). We also tested whether a manipulation of preoccupation with a secret would show evidence of a causal link between secrecy preoccupation and burden-consistent outcomes, and whether this causal link is mediated by judged effort to keep the secret, consistent with a resource-based mechanism of the burdens of secrecy (Studies 3 and 4).

#### Study 1

In Study 1 we ask participants to recall “big” and “small” secrets as in Slepian and colleagues (2012, Study 1). We also measure a variable that we hypothesize more closely tracks the participant’s subjective sense of the gravity of the secret (i.e., how personally influential the secret is). We test this by measuring how preoccupied participants are by their secrets.

## Method

Study 1 served as an exact replication of Slepian and colleagues' (2012) Study 1, with the inclusion of the additional aforementioned measurement of how preoccupied participants are by their recalled secrets. This study was preregistered on the Open Science Framework.<sup>2</sup> Thus, the methods, procedure, sample size, and analysis plan, including rules for data exclusions, were all committed to in advance of data collection (see Brandt et al., 2014). Sample size was determined by a recent proposal that replication attempts should recruit 2.5 times the sample size as the original study (Simonsohn, 2013). One hundred participants were thus recruited for each study in the current work, utilizing Mechanical Turk, which allowed for anonymous recall of secrets. All data exclusions, all manipulations, and all measures are reported for each study (see Simmons, Nelson, & Simonsohn, 2012).

Participants ( $M_{\text{age}} = 30.44$  years, 54% female) were recruited for a study ostensibly on judgments about the workplace, and were given instructions for the study upon agreeing to participate (in lieu of a consent form). Participants read, "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." On the next line, participants read, "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." Participants were reminded that they could write as much as they would like and that responses were completely anonymous. Participants who wrote that they did not have a secret that met the qualifications of the prompt were excluded (as decided prior to data collection). Next, participants completed a measure of how preoccupied they were by their secret. Participants answered, "How much do you think about your secret?", "How much does it affect you?", and "How much does it bother you?" (from 1-*not at all* to 7-*very much*).

Next, in an ostensibly separate study, participants judged a series of control items (the sturdiness of a table, the durability of a water bottle, the temperature in degrees Fahrenheit of a park), and the critical dependent measure, the slant of a pictured hill. The former three items were standardized, and an average was taken as a measure of control numerical estimation. Participants were reminded that 0 degrees is a flat surface, while 90 degrees is a vertical surface, and therefore their estimation should be in between those two numbers. Participants whose responses did not fall in this range of values were also excluded from analysis (as decided prior to data collection).

In all studies, a JavaScript code was embedded into the Mechanical Turk recruitment, preventing individual Mechanical Turk users from participating in a study if they had previously participated in a secrecy study conducted by the authors. This code thus prevents repeat participants within the current work (across studies), and also across previous studies on the burdens of secrecy (i.e., those previously conducted by the current authors).

## Results

Two participants in the "big" secret condition indicated that they did not have a secret that fit the prompt, and these participants were therefore excluded. The predetermined analysis plan was to conduct a 2 (condition)  $\times$  2 (judgment type) ANOVA as in

Slepian and colleagues (2012, Study 1), as well as parallel analyses that examined whether the preoccupation measure predicted hill slant judgments, and control judgments.

**"Big" versus "small" secrets.** The 2 (condition: big secret, small secret)  $\times$  2 (judgment type: hill slant, control estimates) ANOVA was conducted on standardized measures of hill slant and the control numerical estimation index (for ease of interpretation, untransformed slant estimates are presented in text; see Figure 1 for standardized means). This analysis revealed no main effect of condition ( $n_{\text{big}} = 47$ ,  $n_{\text{small}} = 51$ ),  $F(1, 96) = 1.90$ ,  $p = .17$ ,  $\eta^2 = .02$ , no main effect of judgment-type,  $F(1, 96) < 0.01$ ,  $p > .99$ ,  $\eta^2 < .01$ , and no interaction,  $F(1, 96) = 1.33$ ,  $p = .25$ ,  $\eta^2 = .01$ . Thus, recalling "big" versus "small" secrets did not differentially influence judgments of hill slant ( $M_{\text{big}} = 39.72^\circ$ ,  $SD = 18.42$ ;  $M_{\text{small}} = 45.29^\circ$ ,  $SD = 17.30$ ).

**Preoccupation.** Regressions were conducted to examine whether the preoccupation measure ( $\alpha = .83$ ) predicted hill slant judgments, and the control numerical estimation index. Increased preoccupation with the recalled secret predicted increased hill slant judgments,  $b = 2.28$ ,  $t(96) = 2.27$ ,  $p = .026$ , but did not predict the control numerical estimation index,  $b = .40$ ,  $t(96) = 1.00$ ,  $p = .32$ .

**Bonferroni correction.** One could argue that because the current study is examining replicability at the operationalization level and the phenomenon level within the same sample of participants, this provides "two chances" to find a significant effect. A conservative answer to this concern would be to apply a Bonferroni correction, which requires taking the alpha-level used to assess significance and dividing it across the two tests, thereby requiring each test to meet the criterion of  $\alpha = .025$  to be considered significant. The  $p$  value for the relationship between the preoccupation measure of secrecy and judgments of hill slant was  $p = .026$ . Thus, in the most conservative test, the link between secrecy preoccupation and judgments of hill slant was only marginally significant.

## Discussion

Asking participants to recall "big" or "small" secrets, based on random assignment, did not influence judgments of hill slant. Yet a measure of how preoccupied participants were by their secret did show some evidence of predicting judgments of hill slant. The former finding presents a demonstration of nonreplicability at the operationalization level, whereby using the same variables from Slepian and colleagues (2012, Study 1), we did not find evidence for a relationship between secrecy and judgments of hill slant. The latter finding presents a demonstration of replicability at the phenomenon level, whereby a different operationalization of the gravity of a secret (how preoccupied one was by the secret) did predict hill slant judgments, lending support to the hypothesized phenomenon.

These findings provide initial evidence that perceived burden is related to secrecy processes, but not necessarily related to the manipulation of recalling "big" and "small" secrets. The link between secrecy preoccupation and hill slant judgments was perhaps only marginally significant. The  $p$  value for the relationship

<sup>2</sup> Retrieved from [https://osf.io/7v5fd/?view\\_only=cfa19efafa1b45b2b5e957cbddcfea13](https://osf.io/7v5fd/?view_only=cfa19efafa1b45b2b5e957cbddcfea13)

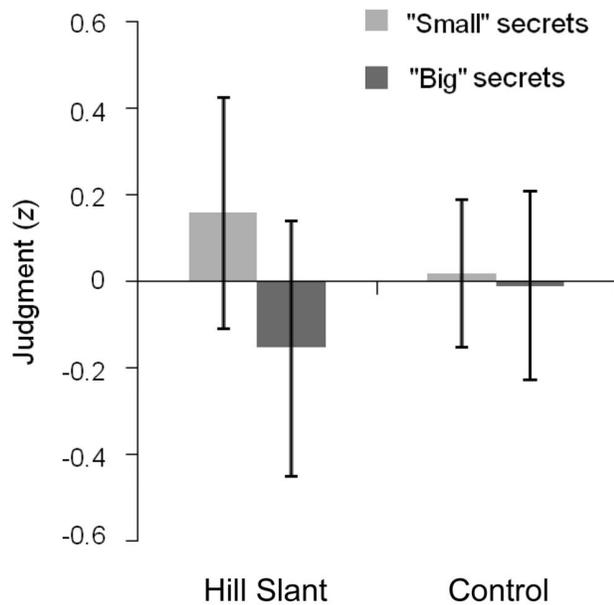


Figure 1. Standardized means of hill slant and control numerical judgments as a function of recalling "small" and "big" secrets in Study 1. Error bars denote 95% Confidence Intervals of the Mean.

between the preoccupation measure of secrecy and judgments of hill slant was  $p = .026$ . This could be considered "significant" or not, depending on which test one considers to be most appropriate (i.e., at the more stringent Bonferroni-corrected  $\alpha$ -threshold of .025, it is as significant as is  $p = .052$  when the  $\alpha$ -level is set to .05). Given that the finding did not meet the strictest alpha-level, this suggested to us the importance of conducting an exact replication of the current Study 1 to examine the reliability of the current pattern of results.

## Study 2

### Method

Study 2 was an exact replication of Study 1, and thus utilized the same preregistered protocol, whereby the same methods, procedure, sample size ( $N = 100$ ;  $M_{\text{age}} = 29.48$  years, 55% male), and analysis plan, including rules for data exclusions, from Study 1 were used, and were committed to in advance of data collection.

### Results

Two participants in the "small" secret condition, and one participant in the "big" secret condition, indicated that they did not have a secret that fit the prompt, and these participants were therefore excluded. The predetermined analysis plan was identical to Study 1: conducting a 2 (condition)  $\times$  2 (judgment type) ANOVA as in Slepian and colleagues (2012, Study 1), as well as parallel analyses that examined whether the preoccupation measure predicted hill slant judgments, and control judgments.

**"Big" versus "small" secrets.** As in Study 1, the 2 (condition: big secret, small secret)  $\times$  2 (judgment type: hill slant, control

estimates) ANOVA was conducted on standardized measures of hill slant and the control numerical estimation index (for ease of interpretation, untransformed slant estimates are presented in text; see Figure 2 for standardized means). This analysis revealed a main effect of condition ( $n_{\text{big}} = 49$ ,  $n_{\text{small}} = 48$ ),  $F(1, 95) = 4.54$ ,  $p = .04$ ,  $\eta^2 = .05$ , but no main effect of judgment-type,  $F(1, 95) = 0.17$ ,  $p = .68$ ,  $\eta^2 = .002$ . These effects were qualified, however, by a significant interaction,  $F(1, 95) = 7.78$ ,  $p = .006$ ,  $\eta^2 = .08$ .

To examine the nature of this interaction, follow-up tests examined the influence of condition on each variable. Asking participants to recall "big" secrets led them to make steeper hill slant judgments ( $M = 43.45^\circ$ ,  $SD = 18.04$ ) than participants asked to recall "small" secrets ( $M = 34.02^\circ$ ,  $SD = 14.34$ ),  $t(95) = 2.85$ ,  $p = .005$ ,  $r = .28$ . There was no difference, however, in the control numerical estimation index ( $M_{\text{big}} = -.005$ ,  $SD = .597$ ;  $M_{\text{small}} = .010$ ,  $SD = .605$ ),  $t(95) = 0.12$ ,  $p = .91$ ,  $r = .01$ .

**Preoccupation.** Regressions were conducted to examine whether the preoccupation measure ( $\alpha = .84$ ) predicted hill slant judgments, and the control numerical estimation index. Increased preoccupation with the recalled secret predicted increased hill slant judgments,  $b = 4.30$ ,  $t(95) = 5.31$ ,  $p < .0001$ , but did not predict the control numerical estimation index,  $b = .06$ ,  $t(95) = 1.76$ ,  $p = .08$ .

**Bonferroni correction.** Despite Study 2 being an exact replication of Study 1, the most conservative analysis again requires a lower criterion at which to assess statistical significance. That is, examining replicability at the operationalization and phenomenon level, within the same participant sample, provides two opportunities to find a significant effect. As in Study 1, the most conservative answer to this concern divides the  $\alpha$ -level used to assess significance across the two tests, requiring each test to meet the criterion of  $\alpha = .025$  to be considered significant. The  $p$  value for the secrecy recall manipulation upon judgments of hill slant falls

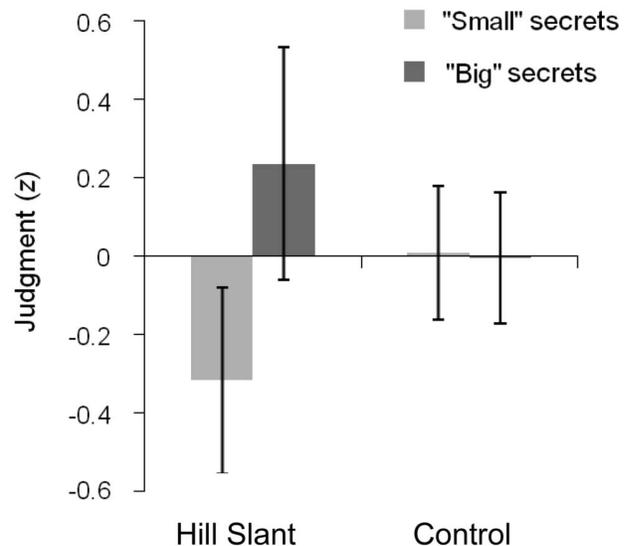


Figure 2. Standardized means of hill slant and control numerical judgments as a function of recalling "small" and "big" secrets in Study 2. Error bars denote 95% Confidence Intervals of the Mean.

below this threshold ( $p = .005$ ) as does the  $p$  value for the relationship between preoccupation and judgments of hill slant ( $p < .0001$ ). In contrast, the parallel  $p$  values for control numerical judgments do not fall below the more stringent  $\alpha = .025$  threshold ( $p = .91$ , and  $p = .08$ , respectively).

## Discussion

Comparing Study 2 to the original Slepian and colleagues (2012) work, we found evidence for replicability at both the operationalization level and the phenomenon level. That is, both the original variable (recalling “big” vs. “small” secrets) as well as the new secrecy variable (how preoccupied one was by the secret) predicted hill slant judgments.

Thus in both Studies 1 and 2, the *preoccupation* measure of secrecy predicted hill slant judgments, whereas only in Study 2 did the recall of “big” and “small” secrets predict hill slant judgments. To examine the reliability of each relationship, we conducted a meta-analysis of the two studies, following procedures outlined in Rosenthal (1991) for combining effect sizes. This analysis revealed an overall effect size of  $r = .07$  for the effect of recalling “big” and “small” secrets upon judgments of hill slant, with an associated  $p$  value of  $p = .37$ .

The overall effect size for the relationship between preoccupation with one’s secret and judgments of hill slant was  $r = .38$ , with an associated  $p$  value of  $p < .001$ . The preoccupation relationship (but not the “big” vs. “small” manipulation) is significant at below the  $\alpha = .025$  threshold that might be reasonably adopted to answer the criticism that including both measures gives two opportunities to find a significant effect.

The failure of recalling “big” versus “small” secrets to produce a consistent effect upon judgments of hill slant is consistent with the notion that secret “size” is an imprecise manipulation of the subjective importance of a secret. Secrets categorized as “big” could, on average, tend to weigh more heavily on the mind of the secret keeper than secrets categorized as “small,” just as big objects tend to weigh more than small objects. But these standard conventions should only matter to the extent that they correlate with a participant’s own subjective view of the secret (just as the size of objects only matters in determining their weight to the extent their size correlates with their mass). How preoccupied a participant is with their secret should determine whether it hampers their ability to navigate their external world, just as the weight of an object one must carry determines whether it hampers one’s movement in the external world. If this is the case, then we would expect a manipulation of subjective preoccupation to more reliably influence our dependent measure, judgments of hill slant.

## Study 3

Preoccupation with a secret may be a more reliable indication of a secret’s gravity than whether the secret is conventionally treated as “big” or “small.” Indeed, in Studies 1 and 2, there was a reliable relationship between preoccupation with one’s secret and a variable that varies with physical burden, judgments of hill slant. However, the link between preoccupation and burden-consistent outcomes has thus far been correlational, and so whether preoccupation with secrets can induce burden-consistent outcomes remains an open question. Study 3, therefore, examined whether a

manipulation of the recall of preoccupying versus nonpreoccupying secrets would produce differences in judgments of hill slant, thereby suggesting a causal link between preoccupation with secrets and burdensomeness.

Study 3 also had a second goal, which was to seek a better understanding of the mechanisms of the burdens of secrecy. Replication attempts are most informative when they extend existing theory (see Cesario, 2014). Absent in prior work on the burdens of secrecy (Slepian et al., 2012; LeBel & Wilbur, 2014) has been theorizing behind how secrecy could influence burden-consistent outcomes. Following recent theory on the economy of action and perceptual judgments (Cole & Balci, 2013; Proffitt, 2006; Schnall et al., 2010), we suggest that the influence of secrecy on exaggerated perceptual judgments of the physical world (e.g., increased judgments of hill slant) should be interpreted as secrecy leading to judgments of the environment as more forbidding and extreme. That is, when preoccupied by a secret, one is devoting personal resources toward that secret. This increased preoccupation with the secret might suggest to the secret holder that increased effort is needed to keep the secret (and thus less effort is available for other pursuits). In other words, just as physical burdens lead to more extreme judgments of the external environment by increasing the effort required to interact with it (Cole & Balci, 2013; Eves, 2014; Proffitt, 2006; Schnall et al., 2010; Witt et al., 2004), perhaps so does preoccupation with secrets.

We hypothesized that increased effort needed to keep a secret mediates the link between preoccupations with one’s secret and judgments of the environment as more forbidding and extreme. Specifically, we predict that preoccupying secrets (relative to nonpreoccupying secrets) will be judged as more effortful to keep, which will predict judgments of the environment as more forbidding (i.e., a hill as more steep).

## Method

As with the prior studies, this study was preregistered on the Open Science Framework.<sup>3</sup> Thus, the methods, procedure, sample size, and analysis plan, including rules for data exclusions, were all committed to in advance of data collection. These predetermined details were identical to Studies 1 and 2, with the exception of a change in the manipulation, and the addition of the new effort measure to explore a potential mediator of the burdens of secrecy. In place of being asked to recall “big” or “small” personal secrets participants were now asked (by random assignment) to recall a secret that met a set of criteria. In the preoccupied condition, participants were asked to recall a secret that met each of three criteria: a) “You think about it reasonably often,” b) “It really affects you,” and c) “It really bothers you.” In the nonpreoccupied condition the criteria were a) “You almost never think about it,” b) “It doesn’t really affect you,” and c) “You feel okay about it.” As is standard, between this manipulation and the dependent measures (from the prior studies), we measured the proposed mediator by asking, “How much effort does it take for you to keep your secret?” from 1-*not at all effortful* to 9-*extremely effortful*. Participants ( $N = 100$ ,  $M_{age} = 34.84$  years, 62% female) then rated the same dependent measures as in Studies 1 and 2. Lastly, as a

<sup>3</sup> Retrieved from [https://osf.io/ntwig/?view\\_only=8484617c288a422bb89e20eb2f8850a2](https://osf.io/ntwig/?view_only=8484617c288a422bb89e20eb2f8850a2)

manipulation check, we measured how preoccupied participants were by their recalled secret, utilizing the same measure from the prior studies.

## Results

Two participants in the nonpreoccupied secret condition, and one participant in the preoccupied secret condition, indicated that they did not have a secret that fit the prompt, and these participants were therefore excluded. The predetermined analysis plan was identical to Study 1: conducting a 2 (condition)  $\times$  2 (judgment type) ANOVA as well as parallel analyses that examined whether the effort measure predicted hill slant judgments, and control judgments. We first confirm, however, whether the new manipulation was successful as measured by the final manipulation check.

**Manipulation check.** We compared participants' responses on the preoccupation measure ( $\alpha = .94$ ), measured last as a manipulation check. Participants asked to recall preoccupying secrets indeed indicated being more preoccupied by those secrets ( $M = 4.94$ ,  $SD = 1.79$ ) than participants asked to recall nonpreoccupying secrets ( $M = 2.23$ ,  $SD = 1.44$ ),  $t(92.87) = 8.24$ ,  $p < .0001$ ,  $r = .64$ .<sup>4</sup>

**Preoccupying versus nonpreoccupying secrets.** As in the prior studies, the 2 (condition: preoccupied, nonpreoccupied)  $\times$  2 (judgment type: hill slant, control estimates) ANOVA was conducted on standardized measures of hill slant and the control numerical estimation index (for ease of interpretation, untransformed slant estimates are presented in text; see Figure 3 for standardized means). This analysis revealed no main effect of condition ( $n_{preoccupied} = 50$ ,  $n_{nonpreoccupied} = 47$ ),  $F(1, 95) = 1.64$ ,  $p = .20$ ,  $\eta^2 = .02$ , and no main effect of judgment-type,  $F(1, 95) = 0.01$ ,  $p = .94$ ,  $\eta^2 < .01$ . There was, however, a significant interaction,  $F(1, 95) = 6.60$ ,  $p = .01$ ,  $\eta^2 = .06$ .

To examine the nature of this interaction, follow-up tests examined the influence of condition on each variable. Asking participants to recall preoccupying secrets led them to judge hill slant as steeper ( $M = 46.24^\circ$ ,  $SD = 14.94$ ) than asking participants to recall nonpreoccupying secrets ( $M = 38.63^\circ$ ,  $SD = 17.17$ ),  $t(95) = 2.33$ ,  $p = .02$ ,  $r = .23$ . There was no difference, however, in the control numerical estimation index ( $M_{preoccupied} = -.076$ ,  $SD =$

$.707$ ;  $M_{nonpreoccupied} = .081$ ,  $SD = .622$ ),  $t(95) = 1.15$ ,  $p = .25$ ,  $r = .12$ .

**Effort.** Regressions were conducted to examine whether the new effort measure predicted hill slant judgments, and the control numerical estimation index. Increased effort judged to keep the secret predicted increased judgments of hill slant,  $b = 1.71$ ,  $t(95) = 2.36$ ,  $p = .02$ , but did not predict the control numerical estimation index,  $b = -.02$ ,  $t(95) = -0.65$ ,  $p = .52$ .

**Indirect effect of preoccupation on judgments of hill slant through effort.** Consistent with recent theory on the economy of action and perceptual judgments, which demonstrates that physical burdens lead to more extreme judgments of the external environment by increasing the effort required to interact with it (Cole & Balcetis, 2013; Eves, 2014; Proffitt, 2006; Schnall et al., 2010; Witt et al., 2004), we predicted that judged effort needed to keep the secret might mediate the influence of recalling preoccupying (vs. nonpreoccupying) secrets upon judgments of hill slant. We used a bootstrapping technique to estimate the indirect effect of the preoccupation manipulation on judgments of hill slant through judgments of effort needed to keep the secret (Hayes, 2009; Preacher & Hayes, 2004). This technique generates an empirical representation of the distribution of the sample by repeatedly resampling it with replacement, producing 5,000 estimates of the indirect effect. The size of the indirect effect is estimated by examining the 95% bias-corrected and accelerated bootstrap confidence interval of these estimates (which corrects for any bias or skew in the distribution; Hayes, 2009). This produces a confidence interval of the indirect path from the preoccupation manipulation to increased judgments of hill slant, through increased judgments of effort needed to keep the secret.

The 95% confidence interval (CI) for the predicted indirect unstandardized path coefficient ( $M = 2.0319$ ,  $SE = 1.4904$ ) ranged from .0423 to 6.1586, which as it does not include zero, demonstrates a significant indirect pathway, wherein preoccupation predicted increased judgments of hill slant through increased judged effort to keep the secret. The same mediation model with the control numerical estimation index as the dependent variable ( $M = -.0132$ ,  $SE = .0488$ ) did not produce a significant indirect path (-.1184, .0823), indicating that the results were specific to judgments of hill slant.

## Discussion

A manipulation that asked participants to recall preoccupying versus nonpreoccupying secrets produced differences in judgments of hill slant. We propose that the success of this manipulation (relative to the "big" vs. "small" manipulation) speaks to the mechanisms of the burdens of secrecy. That is, it is important to separate phenomenon-based hypotheses from specific study methodology in evaluating replication attempts. For example, Slepian and colleagues (2012) proposed that consistent with language used to describe secrets (e.g., being "burdened" or "weighed" down), secrecy processes might lead to judgment outcomes that vary with physical burden, suggesting that the psychological burden from

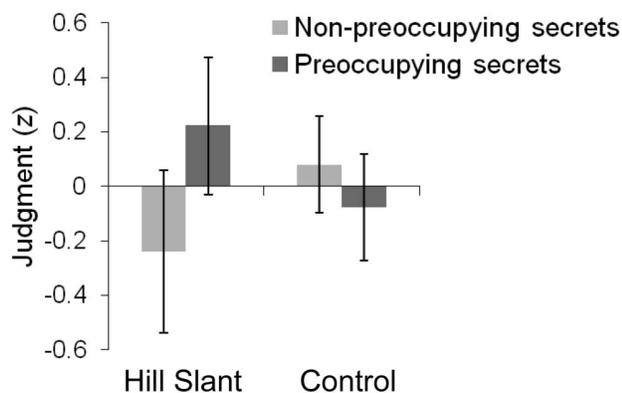


Figure 3. Standardized means of hill slant and control numerical judgments as a function of recalling preoccupying and nonpreoccupying secrets in Study 3. Error bars denote 95% Confidence Intervals of the Mean.

<sup>4</sup> Levene's Test for Equality of Variances,  $F = 3.97$ ,  $p = .05$ , revealed that variances significantly differed, and thus a correction factor was used, which does not alter the significance of the results (without the correction factor:  $t(95) = 8.18$ ,  $p < .0001$ ).

secrecy could lead to outcomes similar to physical burden (indeed, a wealth of evidence supports a tight coupling between such mental and physical depletion, as well as a coupling between mental and physical effort; Baumeister, Muraven, & Tice, 2000; Preston & Wegner, 2009).

One method used to test secrecy-burden hypothesis was asking participants to recall “big” versus “small” personal secrets, relying upon the assumption that secrets recalled in the former condition should be more psychologically burdensome than those recalled in the latter condition. However, the veracity of that assumption rests upon participants recalling secrets that are more or less psychologically burdensome in their respective conditions. They might sometimes, however, recall secrets that fit conventions of “big” or “small,” without actually recalling secrets that are personally influential, or uninfluential, respectively.

Instead, a more precise manipulation of asking for influential or uninfluential secrets might serve to better manipulate the phenomenon of interest. Indeed, initial evidence supports this hypothesis. Recall that in Study 3, participants in the preoccupied condition recalled more preoccupying secrets than those in the nonpreoccupied condition [ $t(92.87) = 8.24, p < .0001$ ]. If recalling “big” and “small” secrets consistently produces similar effects (i.e., recalling personally influential and noninfluential secrets, respectively), then participants should consistently differ with respect to the measure of preoccupation. While in the expected direction, they did not significantly differ on this measure, however [Study 1:  $M_{\text{“big”}} = 4.14, SD = 1.70$ ;  $M_{\text{“small”}} = 3.75, SD = 1.84, t(96) = 1.11, p = .27$ ; Study 2:  $M_{\text{“big”}} = 4.16, SD = 1.95$ ;  $M_{\text{“small”}} = 3.76, SD = 1.80, t(95) = 1.05, p = .30$ ], suggesting perhaps some participants are recalling influential and noninfluential secrets, respectively, but not all. Thus it seems that the “big” versus “small” secret recall manipulation only weakly correlates with a measure of how personally influential a secret is to the secret keeper, and thus might only sometimes produce downstream burden outcomes.

The new successful secret preoccupation manipulation (based upon the reliable relationship found between measured preoccupation and judgments of hill slant in Studies 1 and 2) also suggested the measurement of a potential mediator to the burdens of secrecy, that of judged effort to keep the secret. That is, prior work examining mechanisms of the influence of physical burden on perceptual judgments demonstrates a key role for effort. One way in which the experience of physical burden leads to judgments of the environment as forbidding or extreme is through increased effort judged to interact with external environment while retaining the physical burden (Cole & Balcetis, 2013; Eves, 2014; Proffitt, 2006; Schnall et al., 2010; Witt et al., 2004; but see Woods, Philbeck, & Danoff, 2009). In other words, physical burden compromises one’s ability to interact with external environment, which leads the environment to be judged as requiring more effort to manage, which thereby leads the environment to seem more forbidding. Indeed, we found evidence that effort judged to keep a secret played a similar meditational role for secrecy.

One possibility is that effort does not serve this meditational role, but rather is redundant with how preoccupied participants are with their secret. Although a regression reveals that preoccupation with secrets (as measured by the manipulation check) and judged effort to keep secrets are related to one another,  $b = .61, p < .001$ , the  $R^2$  for this regression reveals that the preoccupation measure

(i.e., the manipulation check) explains only 32% of the variance of judged effort to keep the secret. It should be noted that measurement of the effort mediator can only reveal statistical mediation. That is, we can only infer, not demonstrate, a causal role for this mediator. That said, asking specifically for how much effort participants judge as needed to conceal their secret brings the advantage of measuring a variable explicitly linked to burden-consistent outcomes in previous work (Cole & Balcetis, 2013; Eves, 2014; Proffitt, 2006; Schnall et al., 2010; Witt et al., 2004). Moreover, judged effort needed to keep a secret is theoretically distinct from how much one is preoccupied by that secret (e.g., we argue that it is easy to imagine a secret that is easy to keep, but preoccupying nonetheless).

Consistent with the resource-based theory of perceptual judgments, we propose that one way in which secrecy might lead to judgments of the environment as forbidding or extreme is through increased effort judged to interact with external environment while retaining the secret. Given the novelty of this prediction, we sought to examine whether this meditational pathway would replicate, conducting an exact replication of Study 3.

## Study 4

Given that Study 3 presents a new paradigm to manipulate the burdens of secrecy, it is important to examine the replicability of the new manipulation, as well as the meditational pathway found in Study 3. We thus conducted an exact replication of Study 3, and examined whether the new manipulation would again influence judgments of hill slant. Critically, we also tested whether the manipulation influenced judgments of effort to keep the secret, and if the meditational pathway found in Study 3 replicated. This latter meditational path begins to suggest a mechanism behind the burdens of secrecy, and thus is particularly important to examine.

## Method

Study 4 employed the same procedure as Study 3. To determine a sample size for this replication study, we calculated the sample size needed to find an effect size of  $r = .15$  (equivalent to Cohen’s  $d = .30$ ). We chose this effect size because we considered it be the smallest effect size we considered meaningful and practical for the current context (see Fritz, Morris, & Richler, 2012); one that is small-to-medium, slightly on the smaller size according to standard conventions (small:  $r = .10, d = .20$ ; medium:  $r = .30, d = .50$ ). As our results suggest thus far, secrecy processes can be difficult to manipulate and measure. Thus, we propose that even a small effect size could be considered meaningful, which is why the chosen effect size is on the smaller end of the small-to-medium continuum, but not so small as to be practically difficult to uncover. An a priori sample size calculator for an independent samples  $t$  test determined that a sample size of 352 participants would be needed to find such an effect size (entering  $d = .30$ , power = .80,  $\alpha = .05$ ; Soper, 2014), and thus we recruited this number of participants ( $N = 352, M_{\text{age}} = 31.92$  years, 55% male).

## Results

Sixteen participants indicated not having a secret that fit the prompt (preoccupying secret  $n = 11$ , nonpreoccupying secret  $n =$

5), and one participant indicated the hill was 90° steep. As in the earlier studies, these participants were excluded. Study 4's results paralleled those of Study 3. The manipulation check was successful, whereby participants indicated being more preoccupied by preoccupying secrets ( $M = 4.96, SD = 1.47$ ) than by nonpreoccupying secrets ( $M = 2.32, SD = 1.10$ ),  $t(325.44) = 19.39, p < .001, r = .73$ .<sup>5</sup>

Recalling preoccupying secrets led to steeper judgments of hill slant ( $M = 43.11^\circ; SD = 17.07$ ) than did recalling nonpreoccupying secrets ( $M = 39.62^\circ; SD = 14.81$ ),  $t(332.995) = 2.001, p = .046, r = .11$ .<sup>6</sup> Additionally, the former group judged more effort was needed to keep their secret ( $M = 4.49, SD = 2.37$ ) than did the latter group ( $M = 2.74, SD = 1.85$ ),  $t(329.20) = 7.43, p < .001, r = .38$ .<sup>7</sup> Judged effort needed to keep the secret predicted judgments of hill slant,  $b = 1.14, t(333) = 3.01, p = .003$ .

Given this pattern of findings, we again examined the proposed meditational pathway using the bootstrapping technique from Study 3. First, the proposed mediator, effort, seemed distinct to preoccupation (as measured by the manipulation check). That is, while the two were related to each other,  $b = .74, p < .001$ , preoccupation only explained 37% of the variance of judged effort needed to keep the secret, giving us confidence that a) these theoretically distinct constructs were b) not too highly interrelated to test our proposed meditational analysis.

Again, recalling preoccupying (vs. nonpreoccupying) secrets increased judgments of hill slant, through increased judgments of effort needed to keep the secret ( $M = 1.7347, SE = .7489$ ), 95% CI (.4706, 3.4452). Critically, a parallel mediation analysis with the control numerical estimation index [ $M = .0152, SE = .0277$ ], 95% CI (-.0374, .0744)] revealed that these effects were specific to judgments of hill slant.

## Discussion

These results demonstrate that, as in Study 3, recalling preoccupying (vs. nonpreoccupying) secrets increased judgments of hill slant, mediated by increased judgments of effort needed to keep the secret. This pattern of results is consistent with the hypothesis that by preoccupying one's resources, a secret seems to require more effort to keep. By requiring more effort to keep the secret, less effort should be available for other pursuits, which is known to lead to more extreme judgments of the environment (Cole & Balcetis, 2013; Eves, 2014; Proffitt, 2006; Schnall et al., 2010; Witt et al., 2004).

### General Discussion

The hypothesis of secrecy processes influencing burden-consistent outcomes promises integrative insights across a number of psychological disciplines (social, cognitive, perceptual and health psychology) by speaking to disparate areas of study (e.g., judgments of physical space, social cognition, coping with secrecy). Yet to examine these intersections, the mechanisms and reliability of such effects need to be examined, and this was the goal of the current work.

The current work demonstrates that secrecy reliably leads to burden-consistent outcomes, but that this process is based in internal, subjective experiences of preoccupation, that is, how personally influential or weighty the secret is (rather than external,

standard conventions of a secret's "size"). Measurement or manipulation of preoccupation with one's recalled secret is reliably related to judgments of hill slant, and this relationship is mediated by judgments of effort needed to keep the secret.

The current work examined the mechanisms of the burdens of secrecy as well as the replicability of the burdens of secrecy at two different levels, the operationalization level, and the phenomenon level. The current findings suggest that one methodology does not reliably produce burden-consistent effects, whereas across each study, evidence was found for the burdens of secrecy with a different operationalization. A meta-analysis of the first two studies demonstrated that the manipulation of asking participants to recall "big" versus "small" secrets did not reliably produce differences in judgments of hill slant, but still a measure of how preoccupied participants were with their recalled secret did reliably predict hill slant judgments (even when dividing the alpha-level used to assess significance across the two tests, thereby requiring each test to meet the criterion of  $\alpha = .025$  to be considered significant).

Studies 3 and 4 then revealed that it is not that secrecy manipulations or dichotomous variables, in general, fail to influence burden outcomes, but rather that the particular manipulation used in Studies 1 and 2 might not be well suited to examining the burdens of secrecy. Participants asked to recall "big" or "small" personal secrets might sometimes produce secrets that seem "big" or "small" according to standard conventions regardless of how influential those secrets are at a personal level. For example, a person who has come to feel accepting of a prior infidelity may nevertheless consider it to be a "big" secret because infidelity is generally treated as such. Therefore, a manipulation that assigns participants to recall "big" or "small" personal secrets may be a rather imprecise manipulation for inducing burdening and nonburdening secrets, respectively. Indeed, a recent paper (LeBel & Wilbur, 2014) documents two failures of a manipulation that asks participants to recall "big," relative to "small," personal secrets on judgments of hill slant, as does one study (of two studies) in the current work. These study failures contribute an important advancement to the study of the burdens of secrecy, by helping advance the notion that a manipulation of secret "size" is too imprecise.

That is, such study failures may be due in part to an overreliance on asking for differently "sized" secrets (wherein participants were not told what is meant by a "big" or "small" secret). Given the demonstrated ambiguity of these terms, participants may have recalled secrets based on criteria that only weakly corresponded with self-perceived personal importance, thereby yielding a rather imprecise manipulation as the critical variable with which to predict experiences of burden (initial evidence supports this hypothesis; see discussion of preoccupation in Study 3's Discus-

<sup>5</sup> Levene's Test,  $F = 15.36, p < .001$ , revealed that variances significantly differed; a correction factor was used that did not alter the significance of the results (without correction factor:  $t(333) = 19.01, p < .0001$ ).

<sup>6</sup> Levene's Test,  $F = 4.49, p = .035$ , revealed that variances significantly differed; a correction factor was used that did not alter the significance of the results (without correction factor:  $t(333) = 1.98, p = .048$ ).

<sup>7</sup> Levene's Test,  $F = 24.07, p < .001$ , revealed that variances significantly differed; a correction factor was used that did not alter the significance of the results (without correction factor:  $t(333) = 7.43, p < .001$ ).

sion). Instead, Studies 1 and 2 found that how preoccupied participants were with their secret positively predicted judgments of hill slant, consistent with the burdens of secrecy hypothesis. Importantly, Studies 3 and 4 also demonstrated this process experimentally, whereby those asked to recall preoccupying secrets, versus nonpreoccupying secrets, judged hill slant to be steeper.

Moreover, the present studies suggested a mechanism by which consequential secrets lead to burden-consistent judgments: increased effort needed to keep the secret. The current work suggests that, just as physical burdens lead to more extreme judgments of the external environment by increasing the effort required to interact with it (Cole & Balcetis, 2013; Eves, 2014; Proffitt, 2006; Schnall et al., 2010; Witt et al., 2004), so too do secrets.

Beyond both providing clarification for a mechanism behind the burdens of secrecy, and the replicability of the influence of recalling secrets on judgments of hill slant at both the operationalization and phenomenon levels, the current work offers implications for secrecy more broadly. The demonstration of the importance of how preoccupied one is by a secret points toward potential avenues for intervention, with potential improvement for health. Without others to discuss the secret with, one only has the option to internally ruminate upon the secret (or attempt to suppress thoughts of the secret, which could also lead to rumination; Lane & Wegner, 1995). Increased preoccupation with one's own secret is indeed associated with decreased well-being (Maas, Wismeijer, van Assen, & Aquarius, 2012), suggesting one possible negative health consequence for the burdens of secrecy.

Recent work suggests that one way to relieve the burdens of secrecy is simply to reveal the secret (Slepian, Masicampo, & Ambady, 2014). In that work, participants revealed (or merely thought about) secrets anonymously over the Internet. Thinking about secrets led to increased judgments of hill slant and distance, relative to a control condition where no mention was made of secrets. Explicitly revealing secrets anonymously over the Internet, in contrast, led to judgments that were no different from the control condition. Thus, revealing secrets, at least temporarily, in an anonymous environment can lift the burden. In those studies, participants were asked to recall, or reveal, "big" secrets (but there was no "small" secrets comparison). Perhaps the methods in those studies could also be improved, with consequent improvement in participants' coping. Given that being asked to recall a "big" secret might not consistently lead to recalling secrets that are most personally influential to participants, then being asked instead to reveal a preoccupying secret might lift the burden even more than being asked to reveal a "big" secret.

It is important to note that revealing secrets to people (nonanonymously) could prove even more beneficial if those others are accepting. Revealing secrets to the wrong person, however, could do more harm than good (by increasing distress; Kelly & Yip, 2006; Rodriguez & Kelly, 2006). Thus when revealing a secret is not an option, targeting preoccupation with one's secret could prove to be a beneficial intervention to the burdens of secrecy. Expressive writing about personal trauma has been shown to have positive benefits for physical and psychological health, through giving individuals enhanced insights into those traumas (Pennebaker, 1989; Smyth, 1998). Perhaps an intervention such as this can also have positive benefits for secrecy by reducing preoccupation with one's secret and thereby reducing the burdens of secrecy.

Finally, Studies 3 and 4 linked consequences of psychological processes upon sensorimotor processes, through action-regulation mechanisms. That is, we linked these two processes by drawing upon an established mechanism of influence, here, the role of effort (for a distinct but related account between mental and physical effort, see Preston & Wegner, 2009). By framing the influence of having a personally influential and meaningful secret as a process by which one can become preoccupied by that secret, we identified a possible mechanism of influence through which such preoccupation could lead to burden-consistent outcomes. Indeed, in both Studies 3 and 4, manipulating the recall of preoccupying (vs. nonpreoccupying) secrets increased judgments of hill slant, mediated by increased judgments of effort needed to keep the secret, consistent with the notion that psychological processes that make interacting with the environment more difficult and effortful lead the environment to be judged as more forbidding and extreme (Cole & Balcetis, 2013; Eves, 2014; Proffitt, 2006; Schnall et al., 2010; Witt et al., 2004).

The current work demonstrates that measurement and manipulation of the recall of preoccupying secrets reliably influences judgments of hill slant, whereas the underspecified manipulation of recalling different "sizes" of secrets does not. Additionally, recalling preoccupying secrets, relative to nonpreoccupying secrets, increased judgments of hill slant, mediated by increased judgments of effort needed to keep the secret, suggesting a potential mechanism for the burdens of secrecy. In sum, the current work helps illuminate how secrecy can prove burdensome, and suggests potential points of intervention to relieve such burden.

## References

- Baumeister, R. F., Muraven, M., & Tice, D. M. (2000). Ego depletion: A resource model of volition, self-regulation, and controlled processing. *Social Cognition, 18*, 130–150.
- 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.
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science, 9*, 40–48. <http://dx.doi.org/10.1177/1745691613513470>
- Cole, S., & Balcetis, E. (2013). Sources of resources: Bioenergetic and psychoenergetic resources influence distance perception. *Social Cognition, 31*, 721–732. <http://dx.doi.org/10.1521/soco.2013.31.6.721>
- Cole, S. W., Kemeny, M. E., Taylor, S. E., & Visscher, B. R. (1996). Elevated physical health risk among gay men who conceal their homosexual identity. *Health Psychology, 15*, 243–251. <http://dx.doi.org/10.1037/0278-6133.15.4.243>
- Cole, S. W., Kemeny, M. E., Taylor, S. E., Visscher, B. R., & Fahey, J. L. (1996). Accelerated course of human immunodeficiency virus infection in gay men who conceal their homosexual identity. *Psychosomatic Medicine, 58*, 219–231. <http://dx.doi.org/10.1097/00006842-199605000-00005>
- Critcher, C. R., & Ferguson, M. J. (2014). The cost of keeping it hidden: Decomposing concealment reveals what makes it depleting. *Journal of Experimental Psychology: General, 143*, 721–735. <http://dx.doi.org/10.1037/a0033468>
- Day, M. V., & Bobocel, D. R. (2013). The weight of a guilty conscience: Subjective body weight as an embodiment of guilt. *PLoS ONE, 8*(7), e69546. <http://dx.doi.org/10.1371/journal.pone.0069546>
- 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 Experimental Psychology: Human Perception and Performance*, 38, 1582–1595. <http://dx.doi.org/10.1037/a0027805>
- Eves, F. F. (2014). Is there any Proffitt in stair climbing? A headcount of studies testing for demographic differences in choice of stairs. *Psychonomic Bulletin & Review*, 21, 71–77. <http://dx.doi.org/10.3758/s13423-013-0463-7>
- 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>
- Frijns, T., & Finkenauer, C. (2009). Longitudinal associations between keeping a secret and psychosocial adjustment in adolescence. *International Journal of Behavioral Development*, 33, 145–154. <http://dx.doi.org/10.1177/0165025408098020>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2–18. <http://dx.doi.org/10.1037/a0024338>
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76, 408–420. <http://dx.doi.org/10.1080/03637750903310360>
- Kelly, A. E., & Yip, J. J. (2006). Is keeping a secret or being a secretive person linked to psychological symptoms? *Journal of Personality*, 74, 1349–1370. <http://dx.doi.org/10.1111/j.1467-6494.2006.00413.x>
- Kouchaki, M., Gino, F., & Jami, A. (2014). The burden of guilt: Heavy backpacks, light snacks, and enhanced morality. *Journal of Experimental Psychology: General*, 143, 414–424. <http://dx.doi.org/10.1037/a0031769>
- Lane, J. D., & Wegner, D. M. (1995). The cognitive consequences of secrecy. *Journal of Personality and Social Psychology*, 69, 237–253. <http://dx.doi.org/10.1037/0022-3514.69.2.237>
- 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>
- Lee, E. H., & Schnall, S. (2014). The influence of social power on weight perception. *Journal of Experimental Psychology: General*, 143, 1719–1725.
- Maas, J., Wismeijer, A. A. J., van Assen, M. A. L. M., & Aquarius, A. E. A. M. (2012). Is it bad to have secrets? Cognitive preoccupation as a toxic element of secrecy. *International Journal of Clinical and Health Psychology*, 12, 23–37.
- Oishi, S., Schiller, J., & Gross, E. B. (2013). Felt understanding and misunderstanding affect the perception of pain, slant, and distance. *Social Psychological and Personality Science*, 4, 259–266.
- Pennebaker, J. W. (1989). Confession, inhibition, and disease. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 211–244). New York, NY: Academic Press.
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36, 717–731. <http://dx.doi.org/10.3758/BF03206553>
- Preston, J., & Wegner, D. M. (2009). Elbow grease: When action feels like work. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 569–586). New York, NY: Oxford University Press.
- Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*, 1, 110–122. <http://dx.doi.org/10.1111/j.1745-6916.2006.00008.x>
- Proffitt, D. R. (2009). Affordances matter in geographical slant perception. *Psychonomic Bulletin and Review*, 16, 970–972. <http://dx.doi.org/10.3758/PBR.16.5.970>
- Proffitt, D. R. (2013). An embodied approach to perception: By what units are visual perceptions scaled? *Perspectives on Psychological Science*, 8, 474–483. <http://dx.doi.org/10.1177/1745691613489837>
- Quinn, D. M., & Chaudoir, S. R. (2009). Living with a concealable stigmatized identity: The impact of anticipated stigma, centrality, salience, and cultural stigma on psychological distress and health. *Journal of Personality and Social Psychology*, 97, 634–651. <http://dx.doi.org/10.1037/a0015815>
- Rodriguez, R. R., & Kelly, A. E. (2006). Health effects of disclosing personal secrets to imagined accepting versus non-accepting confidants. *Journal of Social and Clinical Psychology*, 25, 1023–1047. <http://dx.doi.org/10.1521/jscp.2006.25.9.1023>
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Schnall, S., Harber, K. D., Stefanucci, J. K., & Proffitt, D. R. (2008). Social support and the perception of geographical slant. *Journal of Experimental Social Psychology*, 44, 1246–1255.
- Schnall, S., Zadra, J. R., & Proffitt, D. R. (2010). Direct evidence for the economy of action: Glucose and the perception of geographical slant. *Perception*, 39, 464–482. <http://dx.doi.org/10.1068/p6445>
- Shea, L., & Masicampo, E. J. (2014). Self-affirmation counteracts the effects of burdens on judgments of distance. *Journal of Experimental Social Psychology*, 50, 105–108.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. Retrieved from SSRN: <http://ssrn.com/abstract=2160588>
- Simonsohn, U. (2013). *Evaluating replication results*. Retrieved from SSRN: <http://ssrn.com/abstract=2259879>. <http://dx.doi.org/10.2139/ssrn.2259879>
- 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 & 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>
- Smyth, J. M. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology*, 66, 174–184. <http://dx.doi.org/10.1037/0022-006X.66.1.174>
- Soper, D. S. (2014). A-priori sample size calculator for Student t-tests [Software]. Retrieved from <http://www.danielsoper.com/statcalc>
- Sugovic, M., & Witt, J. K. (2013). An older view on distance perception: Older adults perceive walkable extents as farther. *Experimental Brain Research*, 226, 383–391. <http://dx.doi.org/10.1007/s00221-013-3447-y>
- Susewind, M., Christandl, F., & Hoelzl, E. (2013). Feel the moral weight on your shoulders: How material objects are experienced as heavier or lighter through moral meaning. In M. Susewind (Eds.), *Subjective morality: Empirical studies on how people balance their own interests with the interests of others and experience moral meaning*. (Unpublished doctoral dissertation). University of Cologne, Cologne, Germany.
- Vangelisti, A. (1994). Family secrets: Forms, functions, and correlates. *Journal of Social and Personal Relationships*, 11, 113–135. <http://dx.doi.org/10.1177/0265407594111007>
- Witt, J. K. (2011). Action’s effect on perception. *Current Directions in Psychological Science*, 20, 201–206. <http://dx.doi.org/10.1177/0963721411408770>
- Witt, J. K., Linkenauger, S. A., Bakdash, J. Z., Augustyn, J. S., Cook, A., & Proffitt, D. R. (2009). The long road of pain: Chronic pain increases perceived distance. *Experimental Brain Research*, 192, 145–148. <http://dx.doi.org/10.1007/s00221-008-1594-3>

- Witt, J. K., Proffitt, D. R., & Epstein, W. (2004). Perceiving distance: A role of effort and intent. *Perception*, *33*, 577–590. <http://dx.doi.org/10.1068/p5090>
- Witt, J. K., & Sugovic, M. (2013). Response bias cannot explain action-specific effects: Evidence from compliant and non-compliant participants. *Perception*, *42*, 138–152. <http://dx.doi.org/10.1068/p7367>
- 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>
- Zheng, M., Fehr, R., Tai, K., Narayanan, J., & Gelfand, M. (2014). The unburdening effects of forgiveness: Effects on slant perception and jumping height. *Social Psychological and Personality Science*. Advance Online Publication. <http://dx.doi.org/10.1177/1948550614564222>

Received November 1, 2014

Revision received December 22, 2014

Accepted December 24, 2014 ■

# Quality of Professional Players' Poker Hands Is Perceived Accurately From Arm Motions

Michael L. Slepian<sup>1</sup>, Steven G. Young<sup>2</sup>, Abraham M. Rutchick<sup>3</sup>, and Nalini Ambady<sup>4</sup>

<sup>1</sup>Tufts University; <sup>2</sup>Fairleigh Dickinson University; <sup>3</sup>California State University, Northridge; and <sup>4</sup>Stanford University

Received 11/1/12; Revision accepted 4/2/13

In the card game of poker, players attempt to disguise cues to the quality of their hand, either by concealment (e.g., adopting the well-known, expressionless “poker face”) or by deception. Recent work, however, demonstrates that motor actions can sometimes betray intentions. The same action can have different movement dynamics depending on the underlying intention (Becchio, Sartori, & Castiello, 2010), and these subtle differences can be decoded by observers (Becchio, Manera, Sartori, Cavallo, & Castiello, 2012; Sartori, Becchio, & Castiello, 2011). Thus, professional poker players' intentions may be visible from their actions while moving poker chips to place bets. Even though professional players may be able to regulate their facial expressions, their motor actions could betray the quality of their poker hand. In three studies, we tested this hypothesis by examining observers' perceptions of poker-hand quality. We also examined individual differences in sensitivity to nonverbal behavior and potential diagnostic motor behaviors as cues to hand quality.

## Study 1

Twenty brief silent video clips (mean duration = 1.60 s,  $SD = 0.68$  s) of professional poker players placing a bet were extracted from randomly sampled videos of the 2009 World Series of Poker (WSOP) tournament. Three versions of each clip were produced: Unaltered clips showed players' bodies from the table up, face-only clips showed players from the chest up, and arms-only clips showed only players' arms pushing chips into the table. Each player's objective likelihood of winning during the bet was known (WSOP displays these statistics on-screen; however, we kept this information from participants by obscuring part of the screen). The number of chips wagered was not confounded with the likelihood of winning (i.e., chip values varied markedly—no participants were poker experts nor knew chip values; see the

Supplemental Material available online for information about the game of poker, WSOP, and further methodological details).

Seventy-eight undergraduates were divided into three groups based on the type of clip they were shown. Each group viewed the 20 clips in a random order and judged the quality of each poker hand (1 = *very bad*, 7 = *very good*). Next, participants rated their overall confidence in their judgments (1 = *not at all confident*, 7 = *very confident*) and their experience with poker (1 = *none*, 7 = *a lot*). Finally, they completed a measure of nonverbal sensitivity (Bänziger, Scherer, Hall, & Rosenthal, 2011).

Data were analyzed using multilevel linear models with quality ratings of the hand depicted in each clip, nested within participants, predicting objective likelihoods of winning. Specifically, the model included participants' quality ratings at Level 1, a set of dummy codes representing condition at Level 2 (the face-only condition was the reference group because our primary hypothesis concerned a comparison between judgments based on facial expressions vs. arm movements or vs. upper-body movements), and all interactions predicting objective likelihoods of winning. This analysis revealed the predicted interaction between the arms-only (vs. face-only) condition and quality ratings,  $b = 1.68$ ,  $t(1554) = 2.88$ ,  $p = .004$ , such that the arms-only group's ratings significantly predicted likelihoods of winning,  $b = 0.94$ ,  $t(1554) = 2.26$ ,  $p = .02$ , whereas the face-only group's ratings marginally inversely predicted likelihoods of winning,  $b = -0.74$ ,  $t(1554) = -1.81$ ,  $p = .07$ . The interaction between the

### Corresponding Authors:

Michael L. Slepian, Tufts University, Department of Psychology, 490 Boston Ave., Medford, MA 02155  
 E-mail: michael.slepian@tufts.edu

Nalini Ambady, Stanford University, Department of Psychology, 450 Serra Mall, Stanford, CA 94305  
 E-mail: nambady@stanford.edu

Psychological Science  
 24(11) 2335–2338  
 © The Author(s) 2013  
 Reprints and permissions:  
 sagepub.com/journalsPermissions.nav  
 DOI: 10.1177/0956797613487384  
 pss.sagepub.com  


upper-body (vs. face-only) condition and quality ratings was not significant,  $b = 0.95$ ,  $t(1554) = 1.65$ ,  $p = .10$ . Reconducting these analyses with the individual-difference measures entered as predictors revealed no two- or three-way interactions,  $ps > .07$ .<sup>1</sup>

We also examined participants' accuracy scores, which were computed by correlating participants' poker-hand ratings with players' objective likelihoods of winning. If these scores were significantly different from zero, performance was different from chance (Table 1). Correlations between these accuracy scores and participants' nonverbal sensitivity, poker experience, and overall confidence in their judgments were separately explored (Table 1). These analyses also showed that judgments in the face-only group were marginally worse than chance, which suggests that players exhibited deceptive facial cues. When isolating arm movements, however, analyses showed that untrained participants judged the quality of poker hands better than chance, which suggests that perceptions of arm movements exert an independent influence on judgments of poker-hand quality. Judgments made when viewing the players' upper body (arm motions plus the face) were at chance. Additionally, when watching arm motions only, participants' nonverbal sensitivity and poker experience were positively correlated with their accuracy.

## Study 2

In Study 2, we replicated the arms-only accuracy finding from Study 1 with a new set of silent video clips to ensure the generalizability of the effect. Twenty-two new, randomly sampled, chest-down close-ups of players placing bets during the 2009 WSOP were extracted from video clips as in Study 1 (mean duration = 1.54 s,  $SD = 0.74$  s). Again, the number of chips wagered was not confounded

with the likelihood of winning (see the Supplemental Material). Thirty undergraduates judged poker-hand quality from these new clips. As in the previous study, data were analyzed with a multilevel model. Results replicated those of Study 1. When participants viewed arm motions, their judgments again predicted the objective quality of professional poker players' hands,  $b = 1.46$ ,  $t(558) = 2.70$ ,  $p = .004$ . Participants' performance was greater than chance when they judged poker-hand quality from viewing players' arm motions (Table 1).

## Study 3

Players who have strong poker hands should be more confident than players who have weak hands, and perhaps this confidence is expressed in motor actions. To the extent that participants' poker-hand quality ratings were influenced by player confidence, having participants judge player confidence could yield similar results. Previous work demonstrates that anxiety disrupts smoothness of body movement (Beuter & Duda, 1985), which suggests that confidence (i.e., lack of anxiety) might be revealed via smoother actions. Therefore, in Study 3, we had participants in one condition judge player confidence, and in a second condition, they judged how smoothly the chips were pushed into the center of the table. If greater confidence in players relates to smoother motor action, smoothness judgments might also predict likelihoods of winning.

Forty undergraduates viewed the same randomly ordered videos from Study 2, judging player confidence ("How confident does this person seem?") or action smoothness ("How smooth is this person's movement?"; 1 = *not at all*, 7 = *very*). They subsequently completed the measure of nonverbal sensitivity used in Study 1. We ran a multilevel model, including participants' quality

**Table 1.** Mean Accuracy in All Conditions and Correlations Between Accuracy and Individual-Difference Measures

Study and condition	Mean accuracy	Correlations		
		Nonverbal sensitivity	Poker experience	Confidence in judgments
Study 1				
Upper body	.02 [-.06, .09]	.14	.14	.19
Face only	-.07 [-.15, .01]	.17	-.32	-.26
Arms only	.07 [.01, .14]	.40*	.39*	.26
Study 2	.15 [.11, .19]	—	—	—
Study 3				
Player confidence	.15 [.07, .24]	.46*	—	—
Smoothness of movement	.29 [.22, .36]	.14	—	—

Note: Accuracy scores are the correlation of participants' ratings of the quality of poker hands with players' objective likelihoods of winning. Values in brackets are 95% confidence intervals (created using Fisher's transformed  $z$ s and then converted back to  $r$  values). If the 95% confidence interval includes zero, accuracy is at chance.

\* $p < .05$ .

ratings at Level 1, a dummy code representing judgment condition (with the player-confidence condition as the reference group) at Level 2, and the interaction predicting objective likelihoods of winning. Analyses revealed a main effect of participants' quality ratings,  $b = 3.33$ ,  $t(855) = 4.17$ ,  $p < .001$ , but no significant interaction of ratings with judgment condition,  $b = 0.54$ ,  $t(855) = 0.58$ ,  $p = .56$ . Reconducting this analysis with the addition of participants' nonverbal-sensitivity scores and all interactions did not reveal any significant main effects of nonverbal sensitivity or interactions with nonverbal sensitivity and other variables,  $ps < .64$ . Thus, both player confidence and smoothness judgments significantly predicted likelihoods of winning, which suggests that movement smoothness might be a valid cue for assessing poker-hand quality. It is unknown, however, how participants interpreted "smoothness" or whether the players' movements that participants rated as smooth were truly smoother than other players' movements. Other physical factors, such as speed, likely played a role (see Patel, Fleming, & Kilner, 2012).

As in Study 1, we also explored correlations between participants' nonverbal sensitivity and accuracy scores. Participants' nonverbal sensitivity significantly correlated with their accuracy as indexed by ratings of players' confidence, but not with their accuracy as indexed by ratings of players' smoothness of movement (Table 1), which suggests the possibility that individual differences in nonverbal sensitivity can be overcome when participants are explicitly directed to attend to potentially diagnostic motor cues.<sup>2</sup>

## Discussion

In three studies with two unique video sets, observers naive to the quality of professional players' poker hands could judge, better than chance, poker-hand quality from merely observing players' arm actions while placing bets. The accuracy of participants' judgments when viewing players' upper bodies was no different from chance, and when observing players' faces, participants' accuracy was nearly worse than chance, which suggests that players' facial cues were deceptive. Arm motions might provide a more diagnostic cue to poker-hand quality than other nonverbal behaviors. Additionally, correlations between nonverbal sensitivity and accuracy from viewing arm motions suggest a positive relationship between the two (see Table 1), and movement smoothness might be a valid cue for assessing poker-hand quality, although more research is needed to document the moderators of the present effects.

These findings are notable because the players in the stimulus clips were highly expert professionals competing in the high-stakes WSOP tournament. Additionally, judges were untrained observers (cf. Ekman & O'Sullivan,

1991) watching clips on average less than 2 s long (see Ambady & Rosenthal, 1992). Nevertheless, professional poker players' motor actions were revealing, enabling perceivers to decode poker-hand quality from minimal visual information. Even in very restrictive settings, motor actions can yield important diagnostic information.

## Author Contributions

M. L. Slepian, S. G. Young, A. M. Rutchick, and N. Ambady conceived and designed the studies. M. L. Slepian, S. G. Young, and A. M. Rutchick conducted the studies and analyzed the data. All authors wrote the manuscript.

## Acknowledgments

We thank Saheela Mehrotra for assistance in conducting the experiments and Aneeta Rattan, Takuya Sawaoka, and Jessica Salerno for helpful comments on an earlier version of this manuscript.

## Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## Funding

This work was supported in part by National Science Foundation Grant BCS-0435547 to N. Ambady and by a National Science Foundation Graduate Research Fellowship to M. L. Slepian.

## Supplemental Material

Additional supporting information may be found at <http://pss.sagepub.com/content/by/supplemental-data>

## Notes

1. The Quality Rating  $\times$  Upper-Body Condition (vs. Face-Only Condition)  $\times$  Participant Confidence interaction was significant,  $b = -0.71$ ,  $t(1476) = -2.19$ ,  $p = .03$ , but subsequent two-way interactions were nonsignificant,  $ps > .07$ , which makes it difficult to interpret the three-way interaction.
2. Additionally, smoothness judgments yielded larger accuracy than confidence judgments. This is an example of when judgments in a "micro" domain (physical properties of action) may be a more diagnostic cue than judgments in a "molar" domain (the meaning behind an action), whereas the reverse is typically the case (see Weisbuch, Slepian, Clarke, Ambady, & Veenstra-Van der Weele, 2010). Such conclusions about greater accuracy, or higher correlations, in one condition than in the other must be made with caution, however, because neither nonverbal sensitivity nor judgment condition significantly interacted with quality ratings in predicting objective likelihoods to win.

## References

- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274.

- Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011). Introducing the MiniPONS: A short multichannel version of the Profile of Nonverbal Sensitivity (PONS). *Journal of Nonverbal Behavior, 35*, 189–204.
- Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: From thought experiments to empirical evidence. *Frontiers in Human Neuroscience, 6*, 117. Retrieved from [http://www.frontiersin.org/Human\\_Neuroscience/10.3389/fnhum.2012.00117/full](http://www.frontiersin.org/Human_Neuroscience/10.3389/fnhum.2012.00117/full)
- Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science, 19*, 183–188.
- Beuter, A., & Duda, J. L. (1985). Sport psychology analysis of the arousal/motor performance relationship in children using movement kinematics. *Journal of Sport & Exercise Psychology, 7*, 229–243.
- Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913–920.
- Patel, D., Fleming, S. M., & Kilner, J. M. (2012). Inferring subjective states through the observation of actions. *Proceedings of the Royal Society B: Biological Sciences, 279*, 4853–4860.
- Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition, 119*, 242–252.
- Weisbuch, M., Slepian, M. L., Clarke, A., Ambady, N., & Veenstra-Van der Weele, J. (2010). Behavioral stability across time and situations: Nonverbal versus verbal consistency. *Journal of Nonverbal Behavior, 34*, 43–56.