Consciousness and Cognition 18 (2009) 762–765

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Short Communication

Inverse effectiveness, multisensory integration, and the bodily self: Some statistical considerations

Nicholas P. Holmes *

Department of Psychology, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel

ARTICLE INFO

Article history:
Received 15 December 2008
Available online 24 May 2009

Keywords:
Body image
Body schema
Rubber hand illusion
Regression towards the mean
Principle of inverse effectiveness
Inverse efficiency

ABSTRACT

A recent report in *Consciousness and Cognition* provided evidence from a study of the rubber hand illusion (RHI) that supports the multisensory principle of inverse effectiveness (PoIE). I describe two methods of assessing the principle of inverse effectiveness ('a priori' and 'post-hoc'), and discuss how the post-hoc method is affected by the statistical artefact of 'regression towards the mean'. I identify several cases where this artefact may have affected particular conclusions about the PoIE, and relate these to the historical origins of 'regression towards the mean'. Although the conclusions of the recent report may not have been grossly affected, some of the inferential statistics were almost certainly biased by the methods used. I conclude that, unless such artefacts are fully dealt with in the future, and unless the statistical methods for assessing the PoIE evolve, strong evidence in support of the PoIE will remain lacking.

1. Introduction

A recent report (Longo, Cardozo, & Haggard, 2008) showed that tactile perception during the 'rubber hand illusion' (RHI) followed the 'principle of inverse efficiency'.[1] The RHI is a bodily illusion in which, following a short period of synchronous multisensory stimulation, participants feel that a visible dummy hand is 'their own' hand, and that touches applied to their real hands are felt 'on' the dummy hand (for review, see Makin, Holmes, & Ehrsson, 2008). This important illusion has been enormously influential in understanding how visible body parts (or their substitutes) may be perceived as belonging to the bodily self. The principle of inverse effectiveness (*sic*, PoIE), states that multisensory stimuli are most likely, robustly, or strongly integrated when the response elicited under the most effective unisensory condition is weak (Meredith & Stein, 1983). Historically, the PoIE has also been enormously influential in the field of multisensory integration, particularly in single-unit recording studies of the mammalian superior colliculus. The PoIE is one of three fundamental scientific principles of multisensory integration. The PoIE has often been applied to behavioural and human studies of multisensory integration, and also appears to have great explanatory power in these domains. The methodology surrounding the PoIE has recently become somewhat controversial, however (e.g., Foxe, 2008; Holmes, 2007), and a number of related issues will be discussed here. The present commentary provides a brief history and perspective on how and why 'regression towards the mean' may affect the interpretation of the PoIE, and offers a few critical comments and numerical simulations concerning Longo and colleagues' recent report. Some

* Fax: +972 25825659.
E-mail address: npholmes@neurobiography.info
URL: http://www.neurobiography.info

[1] This principle is usually termed 'inverse effectiveness'. It is unrelated to, and not to be confused with 'inverse efficiency' calculations, used in assessing speed-accuracy trade-offs (e.g., Holmes, Sanabria, Calvert, & Spence, 2006a).

1053-8100/$ - see front matter
doi:10.1016/j.concog.2009.04.009

closely-related issues and methods were addressed recently (Holmes, 2007), however, due to some potential misinterpretations of these earlier comments, an additional discussion is justified.

2. Regression towards the mean

'Regression towards the mean' occurs when data are split into groups or sorted according to one criterion, and then compared according to another criterion, if those two criteria are not perfectly correlated (i.e., $r < 1$). Galton, 1877 noted the phenomenon of 'reversion towards mediocrity' in 1877. He weighed hundreds of sweet-pea seeds across several generations, and found that the mean weight of the offspring seeds was consistently and lawfully closer to the population mean than the parent mean. The offspring weights 'reverted' towards the mean, even though the overall distribution remained constant across generations. The equations and concepts of regression originated from these observations.

Extending this principle to humans, Galton later noted (e.g., Galton, 1886), that the adult children of tall parents tended to be shorter than the mean height of their parents (after accounting for sex). Galton initially explained this reversion towards mediocrity in terms of sexual selection and fitness, but later settled for a simpler account, noting that height is determined by numerous different variables (e.g., more than 50 bones, the cartilage between them, the flesh of the scalp and feet, the time of day, and one's posture). These variables are somewhat independent of each other, and may be transmitted with varying fidelity to the next generation. Height is the sum of many variables, some hereditary, some non-hereditary, but all of which contribute some 'signal' and some 'noise' to the height of an individual both within and between generations.

A sample of extreme values selected post-hoc from a distribution of measurements like height or performance will be biased, containing extremes of random variability along with the real underlying values. If this extreme sample is retested, the underlying values may be preserved (i.e., the first and second samples are correlated), but the extremes of variability selected in the first sample will likely diminish, and the second sample will likely be closer to the mean (if the between-sample correlation is <1). Tall parents are likely to have shorter children and short parents are likely to have taller children. To rephrase this in terms of multisensory integration and the PoIE: cells or subjects with low responses or poor performance in a 'unisensory' condition are likely to have higher responses or better performance in a multisensory condition, even if no 'real' multisensory integration occurs.

3. Combining the RHI and the PoIE

Longo and colleagues (2008) combined the experimental techniques of the RHI with the analytical techniques of the PoIE. They assessed whether participants who were worst at a tactile orientation discrimination task also showed the greatest improvement in performance when a visible dummy hand was felt as their own (i.e., during the RHI). The participants discriminated between two orientations of different width gratings (0.75, 1.00, 1.25, and 1.50 mm), applied to their fingertip, following three kinds of stimulation: (1) synchronous visual-tactile stroking, which induces the RHI in susceptible participants; (2) asynchronous visual-tactile stroking, which typically does not induce the RHI, and; (3) a novel control condition with no stroking.

This unisensory tactile detection task performed under several different conditions or illusion states is quite different from the tasks traditionally used to assess the principles of multisensory integration (i.e., passive stimulation with single or pairs of discrete sensory stimuli such as airpuffs, tones, or flashes). It is therefore unclear whether one should expect the data to follow the PoIE at all. Indeed, it might be preferable to think of this kind of task in terms of multisensory 'interplay' rather than 'integration', since no single unified multisensory percept of 'orientation' or 'my hand' can be assumed (Driver & Noesselt, 2008). It remains an open question whether the principles of multisensory integration originally derived from studies on the cat superior colliculus will apply to all such forms of multisensory interplay, and numerous experimental findings are now known not to obey the classic principles of multisensory integration (Foxe, 2008).

The participants in Longo and colleagues' study performed quite poorly on their task, with a mean of 62.3% correct. Indeed, 9, 10, and 13 of the 22 participants in the three conditions performed at chance levels (<60% correct, binomial test), while 8 performed at chance overall (<55%). Interpreting data from this level of performance is justified by the fact that near-threshold performance is thought to be important to test the PoIE. The authors first assessed whether the PoIE significantly characterised their data on a between-participants basis. They found that it did – a negative correlation was found between overall performance and differences in performance between the synchronous condition and the other two conditions (i.e., the visual enhancement of touch effect, VET). The VET was largest when overall performance on the task was worst, consistent with predictions based on the PoIE.

4. Two kinds of inverse effectiveness

The PoIE is typically assessed using one of two methods, which may be termed 'a priori' and 'post-hoc', based, respectively, upon the stimuli presented and the responses obtained. The a priori method assesses the PoIE using two or more levels of task difficulty (or stimulus intensity) under discrete experimental conditions, and may be analysed on a balanced within-participants basis (e.g., Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). The post-hoc method does not necessarily design conditions differing in task difficulty or stimulus intensity, but rather sorts and divides the data post-hoc with respect to the obtained responses. The post-hoc method typically assesses the PoIE on a between-participants basis (e.g., Serino, Farnè, Rinaldesi, Haggard, & Làdavas, 2007). This method may, however, be problematic (Holmes, 2007).

Longo and colleagues designed an a priori PoIE experiment, with four stimuli of monotonically increasing grating widths (i.e., increasing task difficulty), but only reported PoIE analyses according to the post-hoc method. This raises two questions: (1) Why was the problematic post-hoc method relied upon, when the authors' elegant parametric design enabled a within-subjects a priori assessment, and (2) How did the authors account for the inherent statistical biases associated with the post-hoc method (Holmes, 2007).

In response to the first question (Longo, personal communication), there was no significant interaction between the grating width (task difficulty) and the RHI condition. According to the a priori method, then, the data did *not* follow the PoIE. In response to the second question, the authors noted the following:

> "Serino et al. (2007) divided their participants into high- and low-performance groups on the basis of performance in the baseline condition only. In the present study, participants were divided on the basis of overall performance across all experimental conditions. This eliminates the possibility that differences between groups could be a statistical artefact of regression to the mean (cf. Holmes, 2007)." (Longo et al., 2008, footnote 1).

Unfortunately, contrary to one possible but incorrect interpretation of an earlier statement (Holmes, 2007, p. 3342), this procedure may reduce, but almost certainly does *not* eliminate the possibility that regression towards the mean may have affected the validity of the statistical tests performed. The performance of participants in the three conditions reported by Longo and colleagues were highly, but not perfectly, correlated with the criterion that was used to assess the PoIE (the overall mean performance was correlated with the conditions: synchronous, $r = .912$, asynchronous, $r = .960$, and no stroking, $r = .945$). Although these r-values are very high, and the differences between conditions small, *any* between-condition r-value lower than 1 will result in some regression towards the mean when assessing the PoIE post-hoc. Since performance in the synchronous condition was correlated least with overall performance ($r = .912$), this condition should show the greatest regression towards the mean (see Longo et al., 2008, Fig. 3). The between-groups difference in performance for the synchronous (12.6%) was smaller than for the asynchronous (20.8%) and no stroking conditions (16.4%). These differences rank order identically to the above r-values (synchronous < no stroking < asynchronous), consistent with regression towards the mean.

5. Assessing inverse effectiveness post-hoc: simulation

To determine the magnitude of any such artefact in Longo and colleague's data, I repeated the analytical and simulation procedures reported previously (Holmes, 2007). Briefly, 10,000 experiments were simulated in which 22 participants completed two blocks of 40 trials under three conditions. Percentage correct performance was randomly generated by sampling from a normal distribution. Performance on the three conditions was then generated by adding a variable amount of noise per condition, rounding to the nearest number of trials, and correcting for ceiling and floor effects. The parameters of the condition-specific noise were adjusted to match the r-values relating performance in each condition to overall performance. A mixed two-way ANOVA[2] was conducted on each experiment, with the variables group (low-performance, high-performance), and condition (synchronous, asynchronous, no stroking).

As expected, the post-hoc PoIE method biased the ANOVA and increased the false-positive rate for the critical interaction between group and condition, from the conventional level of 5% to approximately 14%. For this interaction, the simulated minimum F-value required for a statistically significant effect at the standard alpha level of 0.05 was $F(2, 40) = 5.78$. The F-value reported by Longo and colleagues, $F(2, 40) = 3.99$, translates to a simulated p-value of .102. As shown in Fig. 1, the observed means in the real data fell within the 95% range of the simulated means for all of the experimental conditions, suggesting that the observed values are generally what one would expect based upon the means and inter-condition correlations of the original dataset.

Despite these negative conclusions, the simulated correlations between overall mean performance and the *differences* between the synchronous condition and the other two conditions (i.e., the VET effects) were actually shifted towards positive values (mean r-values = .22, .17), and opposite in direction to those reported by Longo and colleagues (−.52, −.37). This simulation therefore supports the conclusion that being in the RHI selectively improves the performance of participants who otherwise performed at or below chance, however the reported p-values obtained in the ANOVA are likely to be misleading.

It is crucial to note that the post-hoc method used to search for evidence of the PoIE (e.g., Longo et al., 2008; Serino et al., 2007) *does not eliminate* the possibility that regression towards the mean may occur. It is an inherent complication of assessing the PoIE post-hoc. Future assessments of the PoIE should always and explicitly take into account the statistical consequences of the search strategies used, and the statistical dependencies among the variables tested. With the post-hoc method of assessing PoIE, it is almost never a question of *if*, but rather of *how much* and in *which direction* regression towards the mean has biased the statistical tests. Ideally, evidence for the PoIE should be derived from experiments and analyses

[2] Although the two-alternative forced choice response required from participants would result in binomially-distributed mean responses across participants, the number of trials per condition (80), and the number of participants tested (22) means that n was sufficiently large for a normal approximation to the binomial distribution, hence justifying the use of ANOVA.

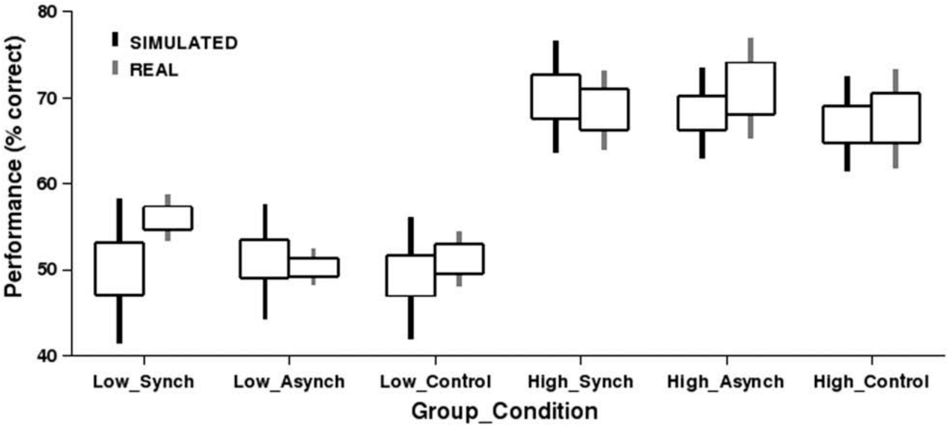

Fig. 1. Simulated versus real datasets. The six conditions along the category axis refer to data for the low accuracy group (Low, leftmost three conditions), and the high accuracy group (High, rightmost three). Each group performed synchronous (Synch), asynchronous (Asynch), and control (Control) conditions. Within each condition, the left box and dark whiskers show simulated data, and the right boxes and light whiskers show the real data. The simulated data boxes show the mean of means ± the mean of standard errors, and the whiskers show the 2.5–97.5% range of the simulated means. The real data boxes show the mean ± standard error of the mean, and the whiskers show the 95% confidence intervals for the means. Only the Low_Synch condition approached statistical significance, but the real mean of this condition was still within the expected 95% limits of the simulated distributions of means, making this result marginal.

conducted using the a priori method, with systematic variations in stimulus intensity or task difficulty. Strong evidence in support of the PoIE will be found in significant interactions between stimulus intensity and the type of sensory stimulation (unisensory *vs.* multisensory).

6. Advances in understanding the bodily self

Aside from the above considerations, the paper by Longo and colleagues provides several important advances: First, a crucial new control condition for the RHI ('no stroking') was introduced alongside the standard conditions. It was previously unknown to what extent the various subcomponents of the RHI, such as the visual capture of touch, body ownership, and proprioceptive drift are *caused by* synchronous multisensory stimulation or rather *prevented by* asynchronous stimulation, since previous studies have focussed only on differences between these two conditions. Longo and colleagues thus provide the first evidence showing that ownership over the dummy hand is stronger for synchronous stimulation than for mere visual exposure. It is not yet known to what extent the visual capture of touch or the proprioceptive drift components are dependent upon a sustained period of correlated multisensory stimulation or rather on simply viewing the rubber hand. Visual capture of touch tends to be an integral part of the RHI, and is typically rated as high as or higher than the ownership components on questionnaires. By contrast, proprioceptive drifts are known to occur in the absence of strong illusions of ownership (Holmes, Snijders, & Spence, 2006b). Second, Longo and his colleagues continue to provide insight into the bodily self by using principal components analysis on RHI questionnaire data (see Longo, Schüür, Kammers, Tsakiris, & Haggard, 2008), improving upon the previous somewhat arbitrary analysis of particular questions, and some other potentially misleading statistical methods (e.g., Holmes et al., 2006b).

References

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron, 57*, 11–23.

Foxe, J. J. (2008). Toward the end of a "principled" era in multisensory science. *Brain Research, 1242*, 1–3.

Galton, F. (1877). Typical laws of heredity. *Nature, 15*, 492–495. 512–514, 532–533.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute, 15*, 246–263.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia, 45*, 3340–3345.

Holmes, N. P., Sanabria, D., Calvert, G. A., & Spence, C. (2006a). Multisensory interactions follow the hands across the midline: Evidence from a non-spatial visual-tactile congruency task. *Brain Research, 1077*, 108–115.

Holmes, N. P., Snijders, H. J., & Spence, C. (2006b). Reaching with alien limbs: Visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership. *Perception & Psychophysics, 68*, 685–701.

Longo, M. R., Cardozo, S., & Haggard, P. (2008). Visual enhancement of touch and the bodily self. *Consciousness and Cognition, 17*, 1181–1191.

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition, 107*, 978–998.

Makin, T. R., Holmes, N. P., & Ehrsson, H. H. (2008). On the other hand: Dummy hands and peripersonal space. *Behavioural Brain Research, 191*, 1–10.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*, 389–391.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*, 1147–1153.

Serino, A., Farnè, A., Rinaldesi, M. L., Haggard, P., & Làdavas, E. (2007). *Can vision of the body ameliorate impaired somatosensory function? Neuropsychologia, 45*, 1101–1107.