Bayes to the rescue: does the type of hypnotic induction matter?

Article (Accepted Version)

Bayes to the rescue: Does the type of hypnotic induction matter?

Running head: Bayes and Hypnotic inductions

Jean-Rémy Martin

University of Sussex

Zoltan Dienes

University of Sussex

Sackler Center for Consciousness Studies

Corresponding Authors:

Martin Jean-Rémy & Zoltan Dienes

School of Psychology

Pevensey

Building

University of

Sussex Falmer

BN1 9QH

jm643@sussex.ac.uk

dienes@sussex.ac.uk

Abstract

Studies comparing different forms of hypnotic induction (e.g., indirect *versus* direct induction) on responsiveness to suggestion have typically found no significant difference between induction types. However, no firm conclusion can be drawn from a non-significant result. In contrast, *Bayes factors* (Jeffreys, 1939) indicate whether evidence favors the null hypothesis (H0) and against the alternative hypothesis (H1), or whether data are simply insensitive. Here we apply Bayes factors to those non-significant results in order to decide: does the form of hypnotic induction really not matter, does it matter, or should we suspend judgment? As the claim that different inductions are differentially effective comes mostly from clinicians, we based the Bayes factors on hypnotherapists' judgments of expected differences between inductions. In addition, we also used empirical differences between induction *versus* no-induction as an estimate of the order of size of effect that could be expected between different inductions, independent of clinical judgment. As a whole, the Bayesian re-analysis of the present evidence supports the claim that additional research should be done on the influence of the induction procedure on hypnotic responsiveness (at least with regard to the inductions considered in the present study), with several exceptions.

Keywords. Hypnosis; Hypnotic Induction; Bayesian Statistics; Bayes Factors

Introduction

In clinical practice, it is often assumed that the depth of the hypnotic state may vary according to the type of hypnotic induction in use (Bandler & Grinder 1975; Erickson & Rossi, 1979; see Terhune and Cardena's 2016 recommendation to optimize inductions; for a skeptical review see Weitzenhoffer, 2000, Chapter 4). The hypnotic proceedings typically include an induction that presents or defines the context as hypnosis and sets the stage for imaginative suggestions that follow (e.g., a hand levitation suggestions to imagine the hand rising off the resting surface), which call for changes in behaviors as well as an array of subjective experiences including alterations in sensations, perceptions, memories, and the experience of involuntariness or nonvolition that often accompanies hypnotic suggestions.

A strong version of state theories of hypnosis argues that hypnosis instantiates a state of consciousness that differs radically from nonhypnotic states of consciousness and that the putative altered state associated with hypnosis directly facilitates responses to hypnotic suggestions (i.e., suggestions for changes in behavior, experiences of volition and reality; Bower, 1966; Fromm, 1992; Jamieson, 2016; Nash, 199; Orne, 1959; Tart, 1983; Woody & Bowers, 1994; for a review of types of altered state theories, see Kirsch, 2011). Consequently, according to strong versions of altered state theories, the effectiveness of different types of hypnotic inductions should be reflected in subjects' responsiveness to suggestions. In contrast, non-state theorists (e.g., T. Barber, 1969; Kirsch, 1985; Lynn et al., 2015; Sarbin, 1950; Spanos, 1986) argue that hypnosis does not produce a specific state of consciousness that, in itself, facilitates response to hypnotic suggestions. Therefore, the type of induction procedure may affect suggestibility indirectly via mundane mechanisms of changes in beliefs, expectations, or motivations to respond to hypnotic suggestions (e.g., Wickless & Kirsch, 1989, JPSP; Benham et al., 1998, JPSP). For current purposes, we can group together with the non-state position those state theories that acknowledge that hypnosis produces alterations in consciousness in keeping with suggestions and perhaps even a more general altered state of consciousness, but do not regard such a state in a causal sense as facilitating other hypnotic responses (cf Kirsch, 2011).

In standard scales of hypnotizability (e.g., Shor & Orne, 1963; Weitzenhoffer & Hilgard, 1962) the induction procedure commonly involves suggestions for relaxation (e.g., "You are becoming more and more deeply and comfortably relaxed") and the wording is phrased in a direct style (e.g., "You feel pleasantly, deeply relaxed and very comfortable as you continue to hear my voice"). We will refer to this type of induction as the *standard induction*. Studies comparing the effectiveness of different forms of induction have usually contrasted the standard induction with other forms of induction that diverge in important ways from the standard induction.

For example, research has contrasted the standard induction with an *active-alert* form of induction in which every suggestion for relaxation in the original induction is replaced by suggestions for alertness (e.g., the suggestion "you become more and more relax" becomes "you become more and more alert"). In addition, in the active-alert induction the participant rides a stationary bicycle during the induction phase, although the wording of the induction is still worded in a direct style (Bányai & Hilgard, 1976; Malott, 1984; Miller, Barabasz, & Barabasz, 1991). The comparison of the active alert with the traditional induction allows testing claims of the relevance of relaxation, drowsiness, or eye closure, for instance, to hypnotic responsiveness (Bányai & Hilgard, 1976).

Researchers have also contrasted the standard induction with a *double induction* procedure (Matthews, Kirsch, & Mosher, 1985) developed by Bandler and Grinder (1975). Matthews et al. (1985), for example, used the double induction procedure as follows: The subject was delivered a syntactically complex hand-levitation induction into the ear contralateral to the dominant hemisphere and simple (childlike) messages for relaxation into the ear contralateral to the non-dominant hemisphere. According to Bandler and Grinder (1975), the complex induction is meant to overload the dominant hemisphere, which is putatively responsible for normal conscious thinking. In this theoretical context, as the dominant hemisphere is overloaded, the non-dominant hemisphere, which is the putative neural basis of unconscious processes, can process the childlike messages more easily, arguably closer to its own language.

Lynn, Neufeld, and Matyi (1987) contrasted the standard induction with an *indirect* form of the induction. In contrast to the former, indirect inductions are worded in an indirect and permissive way (for example “Isn’t it nice to close your eyes at the end of a day? … maybe your eye lids will feel heavy, or maybe they won’t ..”). Studies have yielded conflicting results with respect to the superiority of indirect induction (or suggestions) compared with the standard induction (J. Barber, 1977; Van Gorp, Meyer & Dunbar, 1985). Lynn et al., (1987) were the first to manipulate the wording (i.e., indirect *versus* direct) of inductions and suggestions independently to make possible a direct comparison of the effect of standard *versus* indirect induction on hypnotizability scores. Finally, Van Der Does, Van Dyck, Spinhoven, & Kloosman (1989) contrasted the standard induction with an individualized or *idiosyncratic* procedure. In this case, the hypnotist is free to use any procedure (relaxation or not) and wording (direct, indirect and so on) that seems appropriate for a particular individual.

The outcomes of studies evaluating the effectiveness of different forms of induction procedures have largely been non-significant. In other words, the active-alert, double, indirect, and idiosyncratic inductions did not produce significant differences on hypnotizability scores when compared with the standard induction. A tempting conclusion from these studies is that the specific form of induction exerts no effect on hypnotic responsiveness; what really matters may simply be that subjects understand the situation as one in which they should use their hypnotic skills (e.g. Gibbons & Lynn, 2010; contrast, Malott, 1984).

Nonetheless, we cannot draw any conclusions about the status of theories from a non- significant result in itself. A non-significant result may mean that there is some evidence for the null-hypothesis (H0) and against the alternative hypothesis (H1); or it may mean that data are simply insensitive in discriminating H0 from H1(Dienes, 2014). The latter case may arise even when the evidence favors the alternative hypothesis somewhat more than the null hypothesis. Therefore, no substantive conclusion follows from a non-significant result. By contrast, *Bayes factors* can indicate whether data provide relative evidence for the null (H0), for the alternative (H1), or whether data are simply insensitive in distinguishing the hypotheses (e.g. Jeffreys, 1939; Morey, Romeijn, & Rouder, in press; Verhagen, & Wagenmakers, 2014). In the present article, we apply Bayes factors to these non-significant results in order to decide: Does the form of hypnotic induction really not matter, does it matter, or should we suspend judgment given the current state of the evidence?

In order to know how strong the evidence is for the H1 that an induction is especially effective, versus H0, the predictions of H1 need to be specified (i.e. the approximate expected effect size is needed; Dienes, 2015). Evidence is always relative, depending on how well one model (e.g. an H1) better predicts the data compared with another model (e.g. H0). For example, data may not discriminate very well between H0 and an H1 that only predicted small effect sizes, yet strongly favor H0 over an H1 that allowed large effect sizes. Thus, Bayes factors are, and should be, sensitive to how H1 is modeled (Lindley, 1993).

The claim that different inductions are differentially effective comes essentially from clinicians (Bányai & Hilgard, 1976; Lynn et al., 1987; Malott, 1984; Matthews et al., 1985; Miller et al., 1991; Van Der Does et al., 1989). Therefore, their judgments about what size of difference could be expected constitute a key relevant consideration in assessing the effectiveness of different inductions. We therefore designed a questionnaire aimed at professional hypnotherapists in which we asked them –based on their intuitions and experience– to predict what size of difference could be expected between the standard induction and the other forms of induction described above: indirect, idiosyncratic, double, and active-alert inductions. We used the estimations of professionals to compute Bayes factors. However, we also based the Bayes factors on empirical results that investigated the effect of induction *versus* no-induction on hypnotic responsiveness as an estimate of the order of size effect that could be expected between different inductions, independent of clinical judgment. In a recent review, Terhune and Cardeña (2016) refer to Braffman and Kirsch (1999) and Hilgard and Tart, (1966) for the effect of an induction *versus* no-induction on measurements of suggestibility; in addition we used Bányai & Hilgard, (1976) and Weitzenhoffer and Sjoberg (1961) studies.

In sum, we will use Bayes factors to evaluate the effectiveness of different inductions compared to the standard induction: Specifically, we will investigate the key evidence for the active-alert induction, double induction, the indirect induction and the idiosyncratic (or individualized) induction. The evidence for each against the standard induction depends on how more effective these inductions could be relative to the standard induction. We will be clear how we formulate predictions, and the different bases for formulating predictions.

Method

Analyses

Bayes Factors. Bayes factors (*B*) indicate the relative strength of evidence for two hypotheses (H0 and H1, Dienes, 2014) and are based on the principle that evidence supports the hypothesis that most strongly predicted it. Specifically, *B* is how much more probable the pattern of data are with respect to H1 rather than H0, so a BF of 3, for instance, means that the data are 3 times more probable assuming H1 rather than H0. Consequently, we need to determine what the hypothesis predicts. Here H0 is the prediction of a single population value, no difference between conditions. The predictions for H1 can be based, *inter alia*, on results from previous studies, on other conditions included in the current experiment, or on expert judgments (see Dienes, 2014, 2015, for different examples). *B* varies between 0 and ∞. A *B* of "1" indicates that the data do not favor one theory over the other, whereas values greater than 1 indicate increasing evidence for the alternative over the null hypothesis, and values less than 1 indicate increasing evidence for the null over the alternative hypothesis. But we need to know how far from 1 and how close to 0 the BF has to be in order to draw specific conclusions. Using conventional criteria (Jeffreys, 1939; Lee & Wagenmakers, 2013), $1/3 \leq B \leq 3$ means that data are insensitive in distinguishing the hypotheses, $B > 3$ indicates "substantial" (Jeffreys, 1939; "moderate", Lee & Wagenmakers, 2013) evidence for H1 over H0, and $B < 1/3$ indicates substantial (or moderate) evidence for H0 over H1 (for an online Bayes factor calculator:

http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm).

These criteria are not arbitrary; 3 corresponds to the level of evidence we have been using for a century as a community, in that if the effect size is about that expected in the model of H1, a *p* of about .05 corresponds to a *B* of about 3 (though there is no monotonic relation between p's and B's; e.g. Lindley, 1957, Dienes 2014).

We modelled the different alternative hypotheses (H1s) on two main sources: First, we asked to a panel of professional hypnotherapists, by means of a questionnaire (see the following section), to predict what size of difference could be expected between the standard induction and the other forms of induction described in the introduction. Second, we used the empirical effect of a standard induction *versus* no-induction/no instruction at all (Braffman & Kirsch, 1999; Bányai & Hilgard, 1976; Hilgard & Tart, 1966; Weitzenhoffer & Sjoberg, 1961). We reasoned that the difference between the standard induction and the other forms of procedure could be expected to be of the magnitude of the difference of size between standard *versus* no- induction. If an induction changes suggestibility by x, then a very poor induction, amounting to no induction, would of course differ from the standard induction by that amount x. Using symmetry, a simple prediction is that a better induction may differ from the standard by roughly the same scale of effect, x, in the opposite direction. Note these expectations are explicitly modeled as imprecise; the scale of the effect is just that, a rough scale. Note also that to get the Bayes factors to give meaningful answers we just need a rough scale, not an exact effect size. The standard induction increasing performance by 5% versus 25% would affect how effective we think inductions in general could be and hence by how much they might differ between themselves (in the same way as baby shoe sizes differ less between themselves than adult shoe sizes).

In combining the four empirical studies (by a fixed effects meta-analysis, studies inversely weighted by squared standard error (i.e., by Bayesian updating based on assuming the uncertainty in each mean was normally distributed), we found that the standard induction increased performance by 1.46 suggestions out of 10 (SE = 0.20). (For simplicity, we averaged over whether a no induction came before an induction or vice versa, where order was counterbalanced.)

Participants

To obtain participants, we implemented the questionnaire by means of Google Form and Qualtrics (Qualtrics, Provo, UT) (the questionnaire is available as Supplementary Material). The questionnaire was distributed by means of the mailing list *Hypnosis Listserv* of the International Society of Hypnosis (http://www.ishhypnosis.org/membership/hypnosis-listserv/), which only accepts clinicians who are credentialed, and to different UK professionals by using directories found on Google by entering the keyword *professional hypnotherapists United Kingdom*. We also sent the questionnaire to the members of the Society for Clinical and Experimental Hypnosis (SCEH), which requires the highest degree in one's field (PhD/PsyD or MA/RN for nurses). We thus have two samples of subjects, as described below.

The questionnaire asked professionals to predict what difference in hypnotizability scores (such as they are calculated in classical standardized hypnotizability scales) they would expect between the standard induction and the other forms of hypnotic induction described in the introduction. More specifically, we instructed participants to imagine that we tested the hypnotizability score of a group of 100 individuals (undergraduates who volunteer to take part in hypnosis research) –using the standard induction– and that we obtained a group average hypnotizability score of 5 (out of 10). Then professionals were asked to predict the effect of using the other induction procedures, rather than the standard induction, on that average score of 5. They should indicate, according to their intuition or experience, 1) the most likely average new score (the average of all subjects) the new induction (e.g., indirect induction) would produce; 2) the maximum score for the average of all subjects that is just plausible for the new induction; and 3) the minimum score for the average of all subjects that is just plausible for the new induction. As an illustration, in the case of the standard *versus* indirect induction, professionals had the following statement:

> An indirect induction here is defined as an induction that uses relaxation suggestions but that is worded in an indirect and permissive way rather than in a direct and non-permissive way. Indicate, according to your intuition or experience, the minimum plausible average, the most likely average, and the maximum plausible average hypnotizability score, respectively, that using an indirect induction would produce, given that an average score of 5 is obtained with a standard direct relaxation induction. Note that each value should be larger than the preceding one (see Supplementary Material).

Five participants filled out the questionnaire from the first sample (four of them came from the list *Hypnosis lListserv*) and 28 from the second sample. One participant was removed from the first sample as he/she had missing data. Ten participants from the second sample filled out the questionnaire completely and no participants from the remainder replied to either one of the test questions. We thus have 10 participants from the second sample for subsequent analyses. We do not have demographic data for the first sample, but we do have such data for the second sample (see Table_1). As can be seen from Figure 1, clinicians are not of one mind; there is a range of beliefs among hypnotherapists. Inspecting the graphs, clinicians almost span the full range from 0 to 10 (given that, we tested another model, a uniform from 0 to 10, see supplementary materials). For simplicity, we merged the two samples for analysis (see Figure 2) (in supplementary material we give Bayes factors for the two samples separately). We fitted a normal distribution by taking the mean most plausible average as the mean of the normal, and the minimum and maximum average as being two standard deviations from the mean. We asked people to respond only if they were not aware of the laboratory work on the effectiveness of different inductions. If clinicians based their intuitions on the very studies we are testing, the test would be circular and invalid. We cannot guarantee that the intuitions of the responding clinicians were not, albeit indirectly, informed by the papers we analyze below. The analyses based on the empirically determined difference between induction and no induction does not have this weakness.

	SCEH (*N* = 10)
University Degree	*N* (out of 10)
Ph.D/Dr./Psy.D	5
Prof.	1
M.A (Master of Art)	0
M.A/RN (for nurses)	1
Other	3: 1 MSW, 1 M.D Development- Behavioural Paediatrician, 1 LCSW-ACP, BCD
Profession	*N* (out of 10)
Clinician	4
Researcher	1
Both	5
Hypnotherapy years of experience (for clinicians, *N*= 9)	Mean = 22.10 SD = 13.95
Academic and/or clinical publications and/or presentations on hypnosis?	*N* (out of 10)
Yes	8
No	2

Table 1. Demographic data of participants.

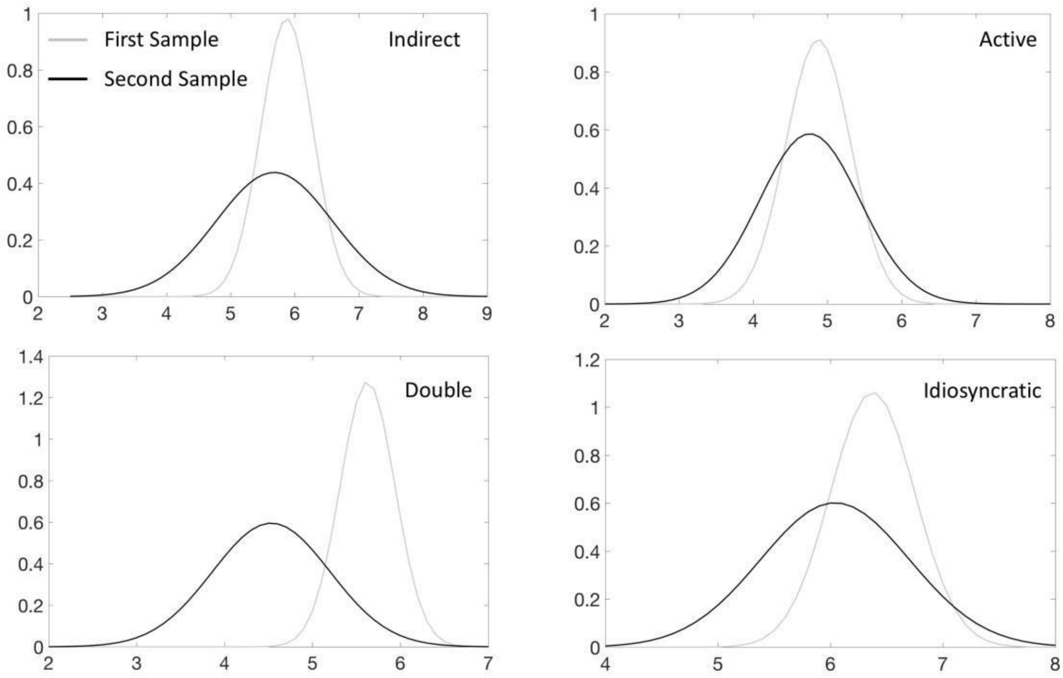


Fig 1. Differences between the first and second sample of participants. The figure shows the differences between the first and second sample of participants in their evaluations of the different types of inductions.

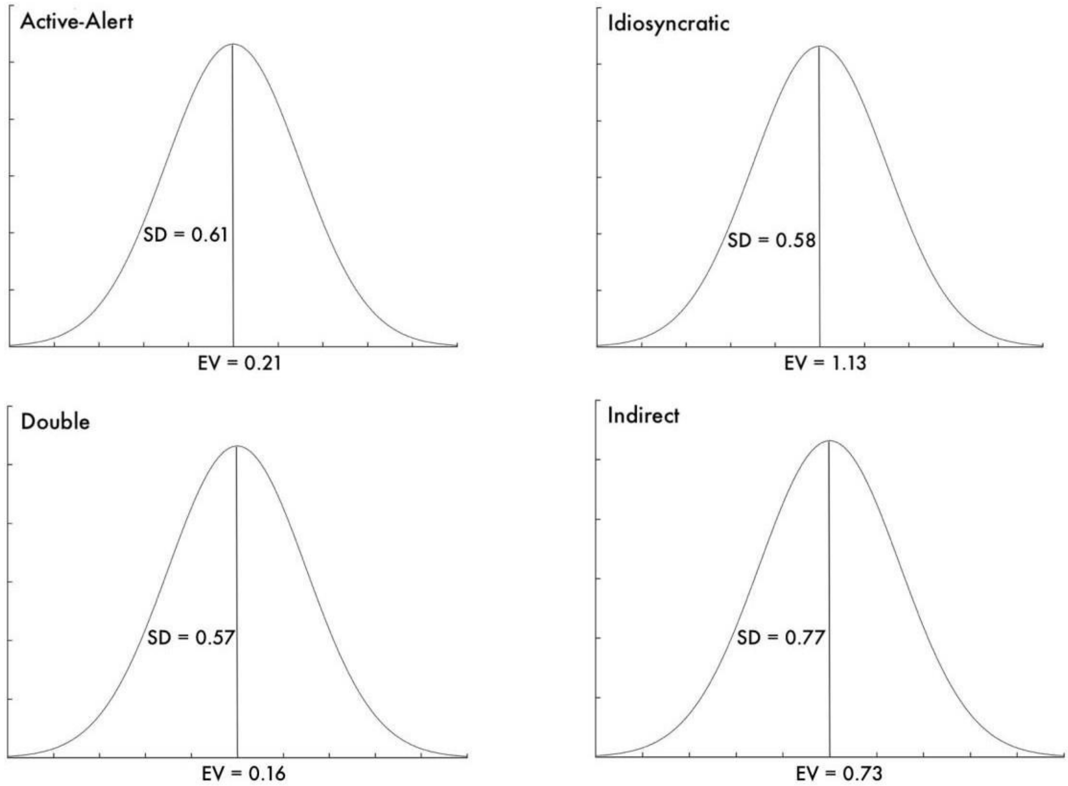


Figure 2. H1s (priors) based on experts' subjective judgments for the different types of induction. EV stands for Expected Value (the most likely average difference the new induction would

produce); SD stands for Standard Deviation.

Results

Indirect. Lynn et al., (1987) found a difference on hypnotizability scores of 0.01 (SE = 0.3) between the indirect and standard procedure, out of 12 suggestions (for comparison with predictions, the difference was rescaled to be out of 10 . Accordingly, here and below the values reported are the rescaled values and original values can be found in supplementary material. . On average professionals estimated a likely difference of 0.73 between indirect and standard (i.e., the indirect induction is considered to be more effective by professionals than the standard induction) with a minimum of -0.68 and a max of 2.37. On average the tail is 1.53 in either direction. So we modelled H1 as a normal distribution with a mean of 0.73 and an SD of 0.77 (see Figure 2). We obtained $B_{N(0.73,\ 0.77)} = 0.25$[1], that is evidence for no effect relative to the effect expected by the professionals.

We also computed a *B*-value with the alternative hypothesis based on the estimated difference between induction and no induction in past key studies, as detailed above. We modelled H1 with a half-normal distribution with a mean of 0 and the SD set to the scale of effect expected, i.e. 1.46. We obtained a $B_{H(0,\ 1.46)} = 0.20^2$.[2] That is, there is substantial evidence for the null hypothesis relative to the size of the effect that could be expected based on the general effectiveness of inductions. We can explore the robustness of this conclusion by means of a Robustness Region.[3] That is, the range of scale factors

[1] For notating how H1 was modelled in calculating *B*, the "N" indicates a Normal distribution was used; the first number in parentheses is the mean and the second the standard deviation (see Dienes 2014, 2015 for this convention). Here we used a normal distribution because it is one of the simplest distributions that could be fit to a given mean about equidistant from a rough minimum and maximum.

[2] For notating how H1 was modelled in calculating B, the "H" indicates a Half-Normal distribution was used; the first number in parentheses is always 0 in this case, indicating the mode of the half-normal; and the second is the standard deviation (see Dienes 2014, 2015, for this convention). Here we used a half- normal distribution as the alternative hypothesis is that the effect of the studied induction on suggestibility scores in comparison with the traditional induction should be positive and expected to be of the magnitude of the size of the effect between induction and no-induction (as argued for by Dienes & Mclatchie, 2018).

[3] Thanks to Balazs Aczel for this suggestion.

(standard deviations of the half-normal) that yield the same conclusion (in this case evidence for H0 relative to H1 with B < 1/3): RR[min, max] where min is the minimum SD that still yields B < 1/3 and max the maximum: RR[0.9, ∞]. Accordingly, the conclusion holds even if one thought the right scale of effect was as low as 0.9 units.

Idiosyncratic. Van Der Does et al. (1989) found a difference of 0.30 (SE = 0.30) between the idiosyncratic and standard procedure out of 5 suggestion (for comparison with predictions, the difference was rescaled to be out of 10). On average, professionals estimated a likely difference of 1.13 between idiosyncratic and standard (i.e., the idiosyncratic induction is considered to be more effective by professionals than the standard induction) with a minimum of -0.06 and a max of 2.26. On average the tail is 1.16 in either direction. So we modelled H1 with a normal distribution with a mean of 1.13 and an SD of 0.58 (see Figure 2). We obtained $B_{N(1.13,\ 0.58)} = 0.34$; that is evidence rather in the direction of no effect relative to the effect expected by the professionals.

As previously, we computed a *B*-value with the alternative hypothesis based on the difference on hypnotizability scores between standard induction and no-induction. We modelled H1 with a half-normal distribution with a mean of 0 and an SD of 1.46, giving a $B_{H(0,\ 1.46)} = 0.54$, that is no substantial evidence in favor of either H0 or H1, RR [0, 2.5], a conclusion robust over reasonable scale factors.

Double induction. Matthews et al., (1985) found a difference of 0.29 (SE = 0.20) between the standard procedure and double procedure out of 12 suggestions (here the standard procedure gave rise to numerically higher hypnotizability scores than the double procedure). For comparison with predictions, the difference was rescaled to be out of 10. On average, professionals estimated a likely difference of -0.16 between double and standard (i.e., the double induction is considered to be less effective by professionals than the standard induction) with a minimum of -0.99 and a max of 1.27. On average the tail is 1.13 in either direction. So we modelled H1 as a normal distribution with a mean of -0.16 and an SD of 0.57 (see Figure 2). The sign of the mean was entered as positive, (i.e., 0.16) because in the study of Matthews et al., (1985), similarly to the expectations of professionals, the standard procedure gave rise to higher hypnotizability scores than the double procedure. We obtained a $B_{N(0.16,\ 0.57)}$ = 0.93; so the evidence is insensitive for discriminating whether the double induction is any different in effectiveness than the standard induction. This conclusion is based on a model of H1 elicited from professionals as a group who do not believe that the double induction would be more effective than the standard induction; if we just consider the H1 derived from the first sample of professionals who believed that the double induction would be more effective, then the Bayes factor provides substantial evidence for no difference (see supplementary materials).

As in the two previous cases we computed a *B*-value with an alternative hypothesis based difference on hypnotizability scores between standard induction and no-induction, in this case considering the H1 that the double induction would be more effective than the standard induction, $B_{H(0,\ 1.46)}$ = 0.06; substantial evidence for that the double induction procedure is no more effective than the standard induction, assuming the effect is of the same order as the difference between induction and no induction; RR[0.23, ∞], robust over any plausible value of the scale factor.

Active-alert. Bányai and Hilgard (1976) found a difference in hypnotizability scores of 0.27 (SE = 0.25) between the standard procedure and the active-alert procedure out of 8 suggestions (here the standard procedure gave rise to numerically higher hypnotizability scores than the active alert). For comparison with predictions, the difference was rescaled to be out of 10. On average professionals estimated a difference of -0.21 between active-alert and standard (i.e., the standard induction is considered to be more effective by professionals than the active-alert induction) with a minimum of -1.26 and a max of 1.18. On average the tail is 1.22 in either direction. So we modelled H1 with a normal distribution with a mean of 0.21 and an SD of 0.61 (changing the sign of the mean (-0.21) so that positive is in the theoretically expected direction) (see Figure 2). We obtained a $B_{N(0.21,\ 0.61)}$ = 0.68; that is, the evidence is insensitive for discriminating whether the active induction is any more effective than the standard induction. We also computed a *B*-value with an alternative hypothesis based difference on hypnotizability scores between standard induction and no induction, $B_{H(0, 1.46)}$ = 0.51, that is no substantial evidence for H0 or H1, RR[0, 2.3].

Malott (1984) found a difference of 0.05 (SE = 0.84) between standard and active-alert out of 8 suggestions (here the standard procedure gave rise to numerically higher hypnotizability scores than the active alert). For comparison with predictions, the difference was rescaled to be out of 10. Using the same normal distribution as above to model H1, $B_{N(0.21, 0.61)}$ = 0.80; no substantial evidence for no effect relative to the effects that the professionals expected. Using the alternative hypothesis based difference on hypnotizability scores between a standard induction and no-induction, $B_{H(0, 1.46)}$ = 0.52, no substantial evidence for H0 or H1, RR[0, 2.5].

Discussion

The present statistical revaluation of the effectiveness of different forms of induction on hypnotic responsiveness indicates that more research should be done in order to determine the influence of different forms of induction procedure on hypnotic responsiveness (Terhune & Cardeña, 2016). Bayes factors mostly indicated no substantial evidence for H0 or H1, no matter whether the alternative hypotheses (H1s) were based on expert judgments or an empirically informed alternative hypothesis (using the effect of induction *versus* no-induction on hypnotic responsiveness as an estimate). This is true for the active-alert and idiosyncratic induction. The exceptions concern the indirect *versus* direct induction for which the Bayes factors indicate evidence that the former is no more effective than the latter; for the double induction, where there is evidence against the double induction being better than the standard (a hypothesis that the clinicians as a group did not share).

To summarize our strategy in modeling H1, the model (often called a prior) needs to reflect what knowledge we have and otherwise represent the remaining uncertainty; so long as the model does that, it allows for a computation of the evidence for H1 vs H0 given existing knowledge. As the change in suggestibility given a standard induction is reasonably small, our model based on that change predicts the difference between different inductions is reasonably small; and thus, the analyses mostly led to the conclusion that we have insufficient evidence to date for drawing any conclusion for several of the inductions. Now compare that to using a uniform over the full range as a model; that indicates we already have evidence for the equivalence of different inductions. We think the former conclusion is more reasonable, because it is scientifically informed. In other words, we should take into account the effectiveness of a standard induction, because in the absence of other relevant knowledge, it sets a reasonable scale of effect.

Our conclusions concerning there being insufficient evidence for differential effectiveness of inductions are restricted to the inductions examined in the present paper and to behavioral scores, as subjective scores have not been studied. Our findings do not necessarily generalize to other types of inductions; however, we have included the main inductions considered in the literature. Also, our conclusions are restricted to comparisons of specific inductions and do not speak to the impact of inductions more broadly (Braffman & Kirsch, 1999). Here we restricted our analysis to types of induction that we deemed particularly relevant to clinicians or that were developed by clinicians (Bányai & Hilgard, 1976; Lynn et al., 1987;Matthews et al., 1985; Van Der Does et al., 1989) and that we therefore expected clinicians would possess intuitions regarding their relative effectiveness. In addition, we restricted our analyses to studies that used suggestibility scores as dependent variables. Therefore, we excluded studies that compared the effectiveness of one of the different induction procedures we evaluated (indirect, double, idiosyncratic, active-alert) that used a standard induction but used a different dependent variable to assess hypnotic responsiveness. For example, Miller et al., (1991) found that suggestions for analgesia were not significantly different after following an active-alert versus following a standard induction.

Other types of inductions than the ones studied in our research have been tested. For example, Glass and Barber (1961) substituted the standard induction with the administration of a placebo described to subjects as a powerful hypnotic drug.[44] They

[4] Descriptives not given, so a Bayes factor could not be calculated.

found that the pill was non-significantly different from a classical induction in increasing suggestibility (see also, Baker & Kirsch, 1993 who used the placebo induction procedure in the context of analgesia). Council, Kirsch, Vickery, and Carlson (1983) compared the effectiveness of a cognitive-behavioural skill induction to a standard induction procedure. The former instructs subjects that suggestions are goal-directed imaginings and experimenters "[describe] cognitive strategies that the subject could use to experience hypnotic suggestions" (p. 434). The researchers found that the skill-based induction increased suggestibility scores to a similar level as the standard induction compared with a no induction baseline condition.[55]

Malott (1984) provided one finding that goes against the pattern of non-significant results across inductions. He found that while the active-alert and traditional inductions did not differ significantly, each was superior to an alert induction that did not use pedaling.[66] Thus, he postulated a U-shape relation between arousal and the effectiveness of a hypnotic induction. An important methodological point for future studies is to determine the changes an induction causes to expectations, motivations, and beliefs about hypnosis or hypnotic responding (Braffman & Kirsch, 1994). In this case, procedures that have greatest effects on participants (e.g. making them extra relaxed or extra alert) may change expectations about being in a special state to the greatest extent, and hence function as especially good inductions. This hypothesis, however, remains a conjecture.

The conclusions about the judgments of clinicians are relative to the clinicians we obtained responses from. Our sample was small, and clinicians with different views of hypnosis are likely to make different judgements. We do not have sufficient data to test predictors of these speculations. Fortunately, there is converging evidence from our analyses of empirically derived effect sizes for potential differences between the standard induction and no induction. We can take into account the large variability

[5] Using the alternative hypothesis based difference on hypnotisability scores between a standard induction and no-induction, $B_{H(0, 1.46)} = 0.16$, that is substantial evidence in favour of H0; standard *versus* skill procedures are not differentially effective.

[6] *Bs* here confirms H1s: Active-alert *versus alert* ($B\text{n}_{(0.21, 0.61)} = 9.7$; $B_{H(0, 1.46)} = 51$); Standard *versus* alert ($B_{H(0, 1.46)} = 60.8$).

among clinicians in their judgments of the differential effectiveness of different inductions by allowing large differences among inductions. The evidence indicates if we expect that inductions differ greatly in their impact on hypnotizability, the evidence is against them being differentially effective. On the other hand, if we base our judgments on the small difference that arises between the standard induction and no induction, then the current evidence leaves open the possibility that different inductions may produce small differences amongst themselves in hypnotizability, on a similar scale as the difference between induction and no induction.

Clinical implications follow from our findings (for a similar view see Lynn, Green, Polizzi, Gautum, et al, in press). “Clinicians need not be unduly concerned” (Lynn et al, p xx) about the exact nature of the induction; but working with clients in ways that suit their preferences may well enhance the experience of hypnosis, even if it produces only a minimal difference in hypnotic responsiveness.

Theoretically, the evidence against large effects places pressure on theories that regard hypnotic response as caused by a special altered state of consciousness. It seems gratuitous to introduce an as yet undefined altered state to explain a small difference that might be explained by expectations or motivations enhancing the (as yet unknown) processes that can produce a response without an altered state (see also Lynn et al).

In sum, the lack of substantial evidence *pro* or *contra* the differential effectiveness of different forms of induction procedure invites additional research (Terhune & Cardeña, 2016), and we provide a methodology for determining when the evidence accumulated is good enough for determining differences between inductions.

References

Baker, S. L., & Kirsch, I. (1993). Hypnotic and placebo analgesia: Order effects and the placebo label. *Contemporary Hypnosis*, *10*(3), 117‐ 126.

Bandler, R., & Grinder, J. (1975). *Patterns of the hypnotic techniques of Milton H. Erickson, M.D. (Vol. 1)*. Cupertino, CA: Meta Publications.

Bányai, E. I., & Hilgard, E. R. (1976). A comparison of active-alert hypnotic induction with traditional relaxation induction. *Journal of Abnormal Psychology*, *85*(2), 218‑ 224.

Barber, J. (1977). Rapid induction analgesia: A clinical report. *American Journal of Clinical Hypnosis, 19,* 138-147.

Barber, T. X. (1969). Hypnosis: A scientific approach. Consulté à l’adresse http://psycnet.apa.org/psycinfo/1970-19931-000

Bowers, K. S. (1966). Hypnotic behavior: The differeniation of trance and demand characteristic variables. *Journal of Abnormal Psychology,* 77,42-51.

Braffman, W., & Kirsch, I. (1999). Imaginative suggestibility and hypnotizability: An empirical analysis. *Journal of Personality and Social Psychology*, *77*(3), 578‑ 587. http://doi.org/10.1037/0022-3514.77.3.578

Council, J. R., Kirsch, I., Vickery, A. R., & Carlson, D. (1983). « Trance » versus « skill » hypnotic inductions: The effects of credibility, expectancy, and experimenter modeling. *Journal of Consulting and Clinical Psychology*, *51*(3), 432‑ 440. http://doi.org/10.1037/0022-006X.51.3.432

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Quantitative Psychology and Measurement*, *5*, 781. http://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2015) *How Bayesian statistics are needed to determine whether mental states are unconscious.* In:Overgaard, Morten (ed.) Behavioural methods in consciousness research. Oxford University Press, Oxford, pp. 199-220.

Elkins, G. R., Barabasz, A. F., Council, J. R., & Spiegel, D. (2015). Advancing Research and Practice: *The Revised APA Division 30 Definition of*

Hypnosis.

International Journal of Clinical and Experimental Hypnosis, *63*(1), 1‐ 9. http://doi.org/10.1080/00207144.2014.961870

Erickson, M . H ., & Rossi, E. (1979). *Hypnotherapy: An exploration case book.* New York: Irvington.

Fromm, E. (1992). An ego-psychological theory of hypnosis. In E. Fromm & M. Nash (Eds.), *Contemporary hypnosis research* (pp. 131-148). New York: Guilford Press.

Gandhi, B., & Oakley, D. A. (2005). Does ‘hypnosis’ by any other name smell as sweet? The efficacy of ‘hypnotic’ inductions depends on the label ‘hypnosis’.

Consciousness and Cognition, *14*(2), 304‐ 315. http://doi.org/10.1016/j.concog.2004.12.004

Gibbons, D. E., & Lynn, S. J. (2010). Hypnotic inductions: A primer. In S. J. Lynn, J. W. Rhue, & I. Kirsch (Eds.), *Handbook of clinical hypnosis* (pp. 267-291). Washington, DC, US: American Psychological Association.

Glass, L. B., & Barber, T. X. (1961). A note on hypnotic behavoir, the definition of the situation and the placebo effect. *Journal of Nervous and Mental Disease, 132*, 539-541.

Hilgard, E. R., & Tart, C. T. (1966). Responsiveness to suggestions following waking and imagination instructions and following induction of hypnosis. *Journal of*

Abnormal Psychology, 71(3), 196‐ 208. http://doi.org/10.1037/h0023323

Jeffreys, H. (1939/1961). *The Theory of Probability*, 1st/3rd Edn. Oxford, England: Oxford University Press.

Jamieson, G. A. (2016). A unified theory of hypnosis and meditation states: the interoceptive predictive coding approach. In A. Raz & M. Lifshitz (Eds.), *Destructive organizational communication: Processes,* *consequences, and constructive ways of organizing* (pp. 53–73). New York, NY: Taylor & Francis.

Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist*, *40*(11), 1189.

Kirsch, I., Wickless, C., & Moffitt, K. H. (1999). Expectancy and suggestibility: are the effects of environmental enhancement due to detection? *The International*

Journal of Clinical and Experimental Hypnosis, *47*(1), 40‑ 45. http://doi.org/10.1080/00207149908410021

Kirsch, I. (2011). The altered state issue: Dead or alive?. *International Journal of Clinical and Experimental Hypnosis*, *59*(3), 350-362.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.

Lindley, D. V. (1993). The analysis of experimental data: The appreciation of tea and wine. *Teaching Statistics*, *15*, 22–25.

Lynn, S. J., Green, J. P., Kirsch, I., Capafons, A., Lilienfeld, S. O., Laurence, J.-R., & Montgomery, G. H. (2015). Grounding Hypnosis in Science: The « New » APA Division 30 Definition of Hypnosis as a Step Backward. *The American*

Journal of Clinical Hypnosis, *57*(4), 390‑ 401. http://doi.org/10.1080/00029157.2015.1011472

Lynn, S. J., Neufeld, V., & Matyi, C. L. (1987). Inductions versus suggestions: Effects of direct and indirect wording on hypnotic responding and experience. *Journal of Abnormal Psychology*, *96*(1), 76‑ 79. http://doi.org/10.1037/0021-

843X.96.1.76

Lynn, S. J., Green, J. P., Polizzi, C., Ellenberg, S., Gautam, A. & Aksen, D. (*in* *press*). Hypnosis, Hypnotic Phenomena, and Hypnotic Responsiveness: Clinical and Research Foundations--A 40 Year Perspective. *International Journal of Clinical & Experimental Hypnosis.*

Malott, J. M. (1984). Active–alert hypnosis: Replication and extension of previous research. *Journal of Abnormal Psychology*, *93*(2), 246‑249. http://doi.org/10.1037/0021-843X.93.2.246

Matthews, W. J., Kirsch, I., & Mosher, D. (1985). Double hypnotic induction: An initial empirical test. *Journal of Abnormal Psychology*, *94*(1), 92‑ 95. http://doi.org/10.1037/0021-843X.94.1.92

Miller, M. F., Barabasz, A. F., & Barabasz, M. (1991). Effects of active alert and relaxation hypnotic inductions on cold pressor pain. *Journal of Abnormal*

Psychology, *100*(2), 223‑ 226.

Morey, R. D., Romeijn, J. W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, *72*, 6-18.

Nash, M. R. (1991). Hypnosis as a special case of psychological regression. In S. J. Lynn & J. W Rhue (Eds.), *Theories of hypnosis: Current models and perspectives* (pp. 171-194). New York: Guilford Press.

Orne, M. T. (1959). The nature of hypnosis: Artifact and essence.

Journal of Abnormal Psychology, 58, 277-299.

Sarbin, T. R. (1950). Contributions to role-taking theory: I. Hypnotic behavior. *Psychological Review*, *57*(5), 255‑ 270. http://doi.org/10.1037/h0062218 Shor, R. E., & Orne, E. C. (1963). Norms on the Harvard Group Scale of Hypnotic

Susceptibility, Form A. *International Journal of Clinical and Experimental Hypnosis,* *11*(1), 39‑ 47.

http://doi.org/10.1080/00207146308409226

Spanos, N. P. (1986). Hypnotic behavior: A social-psychological interpretation of amnesia, analgesia, and « trance logic ». *Behavioral and Brain Sciences*, *9*(03), 449–467. http://doi.org/10.1017/S0140525X00046537

Terhune, D. B., & Cardeña, E. (2016). Nuances and uncertainties regarding hypnotic inductions: toward a theoretically informed praxis. *American Journal of Clinical Hypnosis*, 59(2), 155-174.

Tart, C. T. (1983). In R. Harre & R. Lamb (Eds.), *The encyclopedic dictionary of psychology* (pp. 19-20). Cambridge, MA: MIT Press.

Van Der Does, A. J. W., Van Dyck, R., Spinhoven, P., & Kloosman, A. (1989). The effectiveness of standardized versus individualized hypnotic suggestions: a brief communication. *International Journal of Clinical and Experimental Hypnosis*, *37*(1), 1‐ 5.

Van Gorp, W. G., Meyer, R. G., & Dunbar, K. D. (1985). The efficacy of direct versus indirect hypnotic induction techniques on reduction of experimental pain. *International Journal of Clinical and Experimental Hypnosis*, *33*, 319- 328.

Verhagen, J., & Wagenmakers, E. J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, *143*(4), 1457.

Weitzenhoffer, A. M., & Hilgard, E. R. (1962). *Stanford hypnotic susceptibility scale, form C* (Vol. 27). Palo Alto, CA: Consulting Psychologists Press. Consulté à l'adresse http://ist-socrates.berkeley.edu/~kihlstrm/PDFfiles/Hypnotizability/SHSSC%20Script.p df

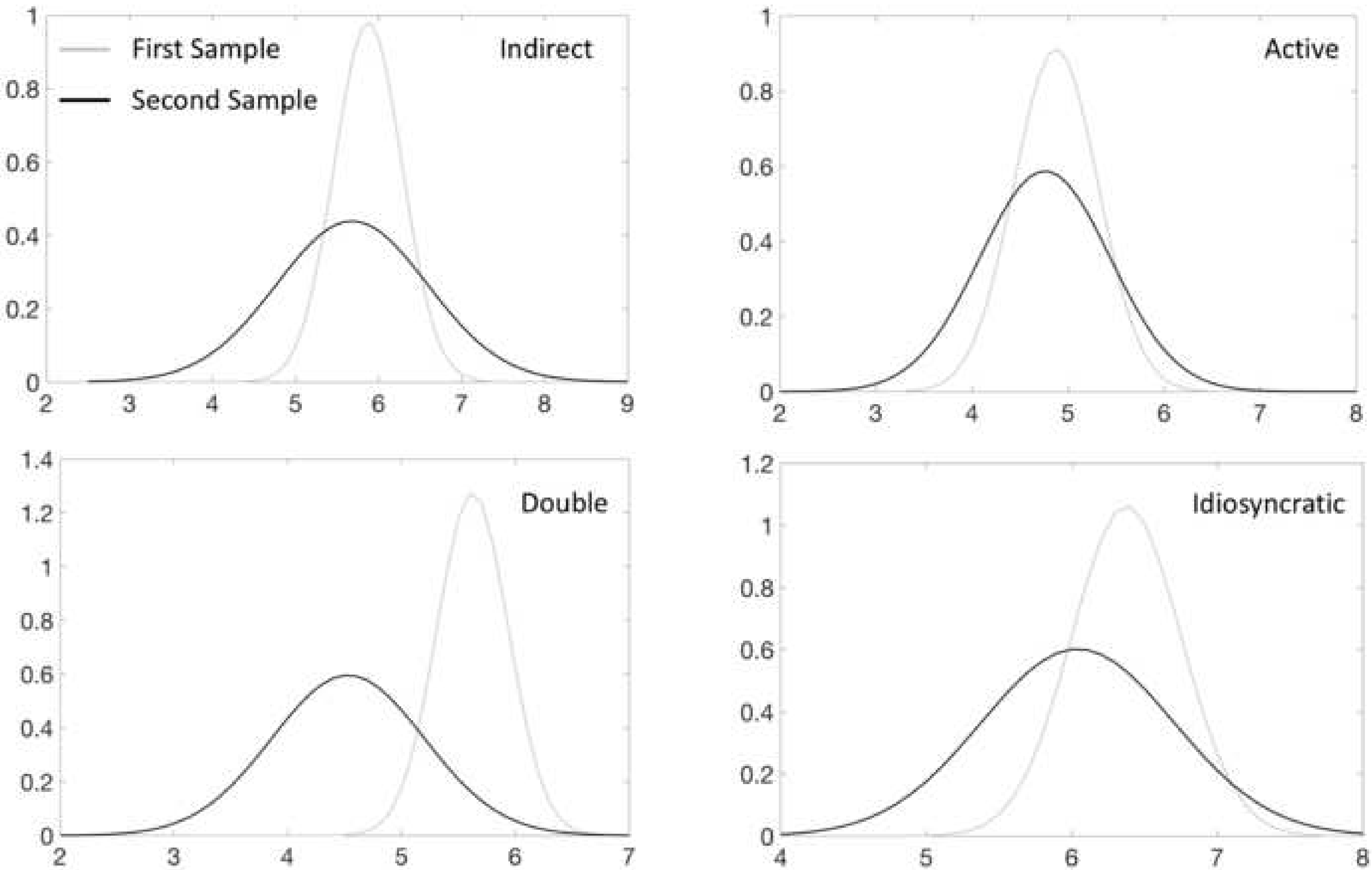
Weitzenhoffer, A. M., & Sjoberg, B. M. (1961). Suggestibility with and without « induction of hypnosis ». *The Journal of Nervous and Mental Disease*, *132*, 204‐ 220.

Weitzenhoffer, A. M. (2000). The practice of hypnotism (2nd ed.). New York, NY: Wiley.

Woody, E. Z., & Bowers, K. S. (1994). A frontal assault on dissociated control. In S. J. Lynn & J. W. Rhue, (Eds.), *Dissociation: Clinical, theoretical and research perspectives* (pp. 52-79). New York: Guilford Press.

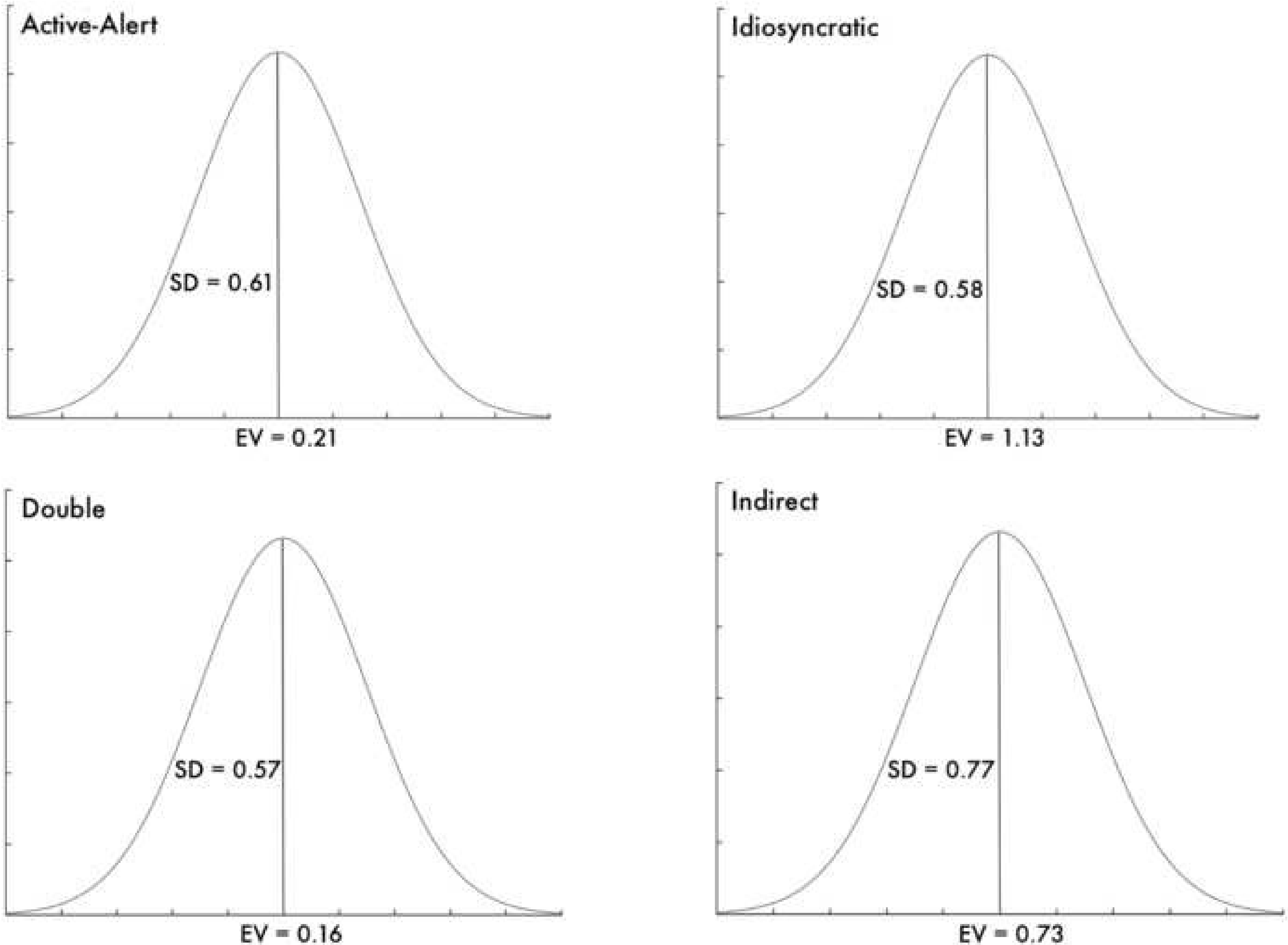
Click here to access/download
Supplemental Materials
Supp_Mat_Questionnaire.docx

Figure



First Sample
Second Sample
Indirect
Active
Double
Idiosyncratic

Figure


Active-Alert
SD = 0.61
EV = 0.21
Idiosyncratic
SD = 0.58
EV = 1.13
Double
SD = 0.57
EV = 0.16
Indirect
SD = 0.77
EV = 0.73

Table

	SCEH (*N* = 10)
University Degree	*N* (out of 10)
Ph.D/Dr./Psy.D	5
Prof.	1
M.A (Master of Art)	0
M.A/RN (for nurses)	1
Other	3: 1 MSW, 1 M.D Development-Behavioural Paediatrician, 1 LCSW-ACP, BCD
Profession	*N* (out of 10)
Clinician	4
Researcher	1
Both	5
Hypnotherapy years of experience (for clinicians, *N* = 9)	Mean = 22.10 SD = 13.95
Academic and/or clinical publications and/presentations on hypnosis?	*N* (out of 10)
Yes	8
No	2