A critical review of standardized measures of hypnotic suggestibility

David J. Acunzo[1], & Devin B. Terhune[2]

[1] *CIMeC-Centre for Mind/Brain Sciences, University of Trento, Trento, Italy*

[2] *Department of Psychology, Goldsmiths, University of London, London, UK*

Corresponding author:
Devin B. Terhune
Department of Psychology
Goldsmiths, University of London
New Cross, London, UK SE14 6NW
Tel: +44 (0) 207 078 5148
Email: d.terhune@gold.ac.uk

Abstract

The most well-established finding gleaned from decades of experimental hypnosis research is that individuals display marked variability in responsiveness to hypnotic suggestions. Insofar as this variability impacts both treatment outcome in therapeutic applications of hypnosis as well as responsiveness to suggestions in experimental contexts, it is imperative that clinicians and researchers use robust measures of hypnotic suggestibility. The current paper critically evaluates contemporary measures of hypnotic suggestibility. After reviewing the most widely used measures, we identify multiple properties of these instruments that result in the loss of valuable information, including binary scoring and single-trial sampling, and hinder their utility, such as the inclusion of sub-optimal suggestion content. The scales are not well-suited for contemporary research questions and have outlived their usefulness. We conclude by outlining ways in which the measurement of hypnotic suggestibility can be advanced.

Keywords: hypnosis; hypnotizability; measurement; psychometrics; suggestion

Operationalization plays an essential role in the study of psychological phenomena and in turns shapes the ways different psychological functions are conceptualized, studied and modelled. Since the advent of experimental hypnosis research, the measurement of responsiveness to hypnosis has fundamentally influenced theory and research (Hilgard, 1965; Woody & Barnier, 2008). The present paper reviews contemporary scales of hypnotic responding with a view to the development of the next generation of measures. After outlining a lexicon of terms, we first consider what is needed from hypnosis scales in terms of how they are currently used. Taking into consideration these application requirements, we critically evaluate scales of hypnotic suggestibility and highlight an array of limitations that greatly reduce their efficacy. We conclude by proposing multiple ways by which the next generation of scales can circumvent these limitations and improve the measurement of hypnotic suggestibility.

Lexicon

The study of hypnosis has involved the use of many terms that are often used in confusing, and arguably incoherent, ways. A set of precise terms is necessary to critically evaluate standardized hypnosis scales. Here we briefly define the different terms that will be used throughout. A *suggestion* refers to an invitation to perform an action or experience a cognition or percept in such a way that it is experienced in an involuntary manner and evokes conviction as to the reality of the experience (Barnier et al., 2008); see also (Kirsch, 1999). Suggestions are typically structured as *happenings* (e.g., "when *X* happens, you will find that *Y*"), that is, events that *happen* to a person rather than *doings* (e.g., "when *X* happens, you will do *Y*")—actions that one *performs* (Spanos & Gorassini, 1984). The experience that one's response to a suggestion is outside one's control (i.e., a distortion in the sense of agency; (Haggard, 2017; Moore, 2016) is the hallmark phenomenological feature of response to suggestion (Bowers, 1981; Bowers et al., 1988; Polito et al., 2014) and widely referred to as the *classic suggestion effect* (Weitzenhoffer, 1974).

Researchers have not yet developed a consensus definition of hypnosis despite numerous attempts (Elkins et al., 2015; Wagstaff, 2014). We maintain that the best way to circumvent this impasse is to adopt a theory-neutral procedural definition that aligns different methods based on their family

resemblance (Terhune, 2014). According to this approach, *hypnosis* consists of a set of procedures involving an induction, intended to minimize awareness of one's environment and mental activity, focus attention on the words of the experimenter, and enhance responsiveness to suggestions for alterations of affect, cognition, motor control, or perception. This definition is neutral regarding the impact of inductions on response to suggestion, the evidence for which is far more mixed than many assume (Terhune & Cardeña, 2016), obviates any need to refer to states of consciousness, trance, or other loaded concepts, and permits nuance in the way in which hypnosis is conceptualized.

The most well established fact about hypnosis is that there are marked individual differences in responsiveness to hypnotic suggestions (Hilgard, 1965; Laurence et al., 2008; Woody & Barnier, 2008). This ability is typically referred to as *hypnotizability* (Elkins et al., 2015), but we maintain that the term *hypnotic suggestibility* is more suitable because it is the most theory-neutral descriptive term for what is being measured by hypnosis scales, namely responsiveness to direct verbal suggestions (suggestibility) following an induction. Although this term has negative connotations, it makes no assumptions regarding an underlying ability, whether someone is or is not, or can or cannot be, hypnotized, and avoids other conceptual confusions and the circularity of the term *hypnotizability* (Braffman & Kirsch, 1999; Kirsch, 1997; Lynn et al., 2015; Terhune, 2014). Finally, in acknowledging responsiveness as a form of suggestibility, it allows for greater integration between hypnosis research and germane fields. This is in keeping with the reliable finding that when measured with robust scales, non-hypnotic suggestibility is a strong predictor of hypnotic suggestibility (Braffman & Kirsch, 1999; Wieder & Terhune, 2019).

Taxonomy and phenomenology of hypnotic suggestions

Hypnotic suggestions can be used to target an array of psychological functions and are typically demarcated into six categories based on two dimensions: the *type* of suggestion and the *psychological function* being targeted (Woody & Barnier, 2008). A fourth function (affect) is omitted from this scheme, but has been the subject of research (e.g., (Ludwig et al., 2013; Wheatley & Haidt, 2005), thereby making up eight types of suggestions (see **Table 1**). In all cases, suggestions are for some type of response, such

as a motor movement (e.g., an anarchic hand movement; Haggard et al., 2004), a phenomenal experience that lacks an external stimulus (e.g., hallucination; (Szechtman et al., 1998)), or changes in performance on a cognitive task (Oakley & Halligan, 2009). An individual is scored as "passing" a suggestion if they display the corresponding response, according to a specific criterion (Hilgard, 1965).

[Table 1 about here]

Embedded within an individual's behavioural response to a suggestion are two phenomenological dimensions pertaining to the experiential concomitants of the response (i.e. attendant *qualia*). In addition to the dimension of *involuntariness* described (Bowers, 1981), responses can differ in *verisimilitude*, the perceived reality of the suggested experience (Terhune et al., 2017; Woody & Szechtman, 2007). For instance, a suggestion for an auditory hallucination may or may not generate an experience as real as an auditory stimulus. The extent to which these dimensions are distinct remains poorly understood and it may be better to think of certain suggestions as evoking one type of experience more than the other (Barnier et al., 2008; Polito et al., 2014).

Standardized scales of hypnotic suggestibility

The measurement of hypnotic suggestibility has been closely intertwined with the development of experimental hypnosis research: reliable, standardized measurement of suggested phenomena provided a solid foundation for the emergence of experimental hypnosis research in the 20th century (Hilgard et al., 1958). We concern ourselves with scales currently in use, which we consider to be those developed in the late 1950s onwards (for reviews of earlier scales, see (Balthazard, 1993; Barnier & McConkey, 2004; Council, 2002; Hilgard et al., 1958). **Figure 1** presents modern hypnotic suggestibility scales in chronological order. We have omitted general-purpose scales that tap broader phenomenological features of hypnosis (Field, 1965b; Pekala, 1991), are not specific to suggestion (Polito et al., 2013), and/or are measures of non-hypnotic suggestibility (Wieder & Terhune, 2019).

The early Stanford scales (Weitzenhoffer & Hilgard, 1959, 1962) and all that followed consist of a relaxation-based hypnotic induction followed by a series of suggestions. Aside from a few exceptions

(Elkins, 2014; Spiegel & Spiegel, 1978; Weitzenhoffer & Hilgard, 1967), responses to suggestions are dichotomously-scored (pass or fail) by an experimenter or the participant. Most scales, except the single-item *Phenomenology of Consciousness Inventory-Hypnotic Assessment Procedure* (Pekala et al., 2010) (for a critique, see (Terhune & Cardeña, 2010b)) include multiple suggestions. Early scales, such as the *Harvard Group Scale of Hypnotic Susceptibility, Form A* (HGSHS:A; Shor & Orne, 1962), were primarily comprised of motor suggestions, which are inferior to cognitive-perceptual suggestions in the identification of highly suggestible individuals (Laurence et al., 2008; Woody & Barnier, 2008). To remedy this limitation, researchers subsequently developed the *Stanford Hypnotic Susceptibility Scale, Form C* (SHSS:C; Weitzenhoffer & Hilgard, 1962), which has a greater representation of items. This orientation was further expanded with the *Revised Stanford Profile Scales* (RSPSs; Weitzenhoffer & Hilgard, 1967), which are comprised entirely of cognitive-perceptual suggestions, although the RSPSs are infrequently used (Moran et al., 2002; Terhune et al., 2011).

[Figure 1 about here]

Further measurement advances include the development of scales that can be quickly administered in clinical and experimental contexts (Elkins, 2014; Hilgard, Crawford, & Wert, 1979; Lush et al., 2018; Pekala et al., 2010; Spiegel & Spiegel, 1978) or online (Palfi et al., 2019; Wieder & Terhune, 2019). However, these and derivative scales have not involved any substantive changes to the original scales. Perhaps the most significant development in recent decades is the introduction of experiential scales focusing on subjective, rather than behavioural responses to suggestion (Bowers, 1981; Kirsch et al., 1990; Kirsch et al., 1998; Polito et al., 2014). These scales, however, are not widely used, despite research showing that they can be useful in correcting for compliance (Bowers et al., 1988) and are often more sensitive than behavioural scales (Cardeña & Terhune, 2014; Polczyk & Pasek, 2006). Notably, all recent scales include experiential measures (Elkins, 2014; Lush et al., 2018; Wieder & Terhune, 2019).

A review of **Figure 1** reveals that a great deal of activity was devoted to scale development in the 1960s. Despite the continued progress on research on basic mechanisms of hypnosis, its clinical application, and exploration of its features from a cognitive neuroscience perspective (Oakley & Halligan,

2009, 2013), there has not been reciprocal development in the area of measurement (Terhune & Cohen Kadosh, 2012; Woody & Barnier, 2008). Thus, the most recent introduction of a novel, standalone measure of hypnotic suggestibility was the SHSS:C nearly 60 years ago, which is widely referred to as the "gold standard" in the field (Kihlstrom, 2008). Henceforth, we refer to the HGSHS:A and SHSS:C, the two most widely used scales (Barnier & McConkey, 2004), and subsequent variations thereof, as the *standard scales*.

Uses of standardized scales of hypnotic suggestibility

Hypnotic suggestibility measurement has had a paramount influence on both the experimental use of hypnosis and its clinical application. Among researchers, hypnotic suggestibility measurement is nearly universally accepted to be necessary and some form of measurement is incorporated in nearly all experimental studies. In turn, it constrains the way hypnosis is studied and the way individual differences in hypnotic suggestibility and hypnosis more broadly are conceptualized and modelled.

The most common use of the standard scales is the screening of participants to identify individuals of varying levels of hypnotic suggestibility. This allows experimental designs involving the stratification and comparison of participants on an independent task or a set of suggestions. Scales are also used as screening tools for specific abilities. It is widely established that highly suggestible individuals do not exhibit uniform patterns of hypnotic responding (McConkey & Barnier, 2004; Terhune, 2015). Accordingly, for experiments on specific phenomena (e.g. suggestions for the modulation of pain) it is advisable, and indeed often necessary, to screen participants for both general hypnotic suggestibility and responsiveness to a specific suggestion. In such cases, researchers might use either the tailored SHSS:C (Hilgard, Crawford, Bowers, et al., 1979), in which relevant suggestions are substituted for unnecessary suggestions, or measure responsiveness to the specific suggestion of interest (Szechtman et al., 1998).

The standard scales are also widely used as dependent measures. Nearly all studies that have attempted to modify hypnotic suggestibility using cognitive training (Gorassini, 2004), non-invasive brain stimulation (Coltheart et al., 2018; Dienes & Hutton, 2013), or pharmacological agents (Whalley &

Brooks, 2009) have used these scales or abbreviated versions thereof. They have also been used by all studies investigating the heritability, and genetic basis, of hypnotic suggestibility (Morgan, 1973; Rominger et al., 2014) as well as its cognitive (Parris, 2017; Terhune et al., 2017) and personality correlates (Cardeña & Terhune, 2014), as well as numerous theoretical questions, such as the relation between hypnotic and non-hypnotic suggestibility (Kirsch & Braffman, 2001). Such usage occasionally involves analysis of individual hypnotic suggestions (e.g., (Bryant et al., 2001; Whalley & Brooks, 2009), although these scales were not developed to provide item-level measurement.

Limitations of contemporary measures of hypnotic suggestibility

The standard scales have many desirable properties including their reliability, their ease of administration, and their construct validity. For the time period in which they were developed, they have a number of excellent structural properties and they have served the field well. The very fact that little deviation has been made since their introduction in the late 1950s clearly speaks to their utility. Although these scales are widely extolled, multiple studies have shown that specific suggestions have poor psychometric properties with detrimental effects for the reliable and valid assessment of hypnotic suggestibility. Posthypnotic motor movement and amnesia suggestions on the HGSHS:A were found to be prone to pseudo-guessing, such that low suggestible participants respond to these items at a much higher rate than would be expected, leading to likely false positive responses (Sadler & Woody, 2004), while the utility of posthypnotic amnesia suggestions appears to come into doubt when appropriate controls are included (Freedman, 2012). An analysis of HGSHS:A items found that self- and observer- scoring significantly differed for the majority of suggestions (Varga et al., 2012) (see also (Younger et al., 2005). Finally, specific suggestions from different scales (finger lock, arm rigidity, hallucinations, posthypnotic suggestion) are reliably associated with more compliant responding (Anllo et al., 2017; Barnes et al., 2009; Bowers et al., 1988). Moving beyond shortcomings pertaining to specific items in the standard scales, in what follows we highlight a number of fundamental and general limitations of these scales, many of which have implications for our understanding of hypnosis.

Response distributions and intra-individual variability

A number of the critiques we outline below rest on the assumption that responsiveness to a particular suggestion is not a deterministic set value and is best understood and modelled as a continuous random variable with an underlying probability distribution. Here we briefly justify this assumption.

Humans display intra-individual variability in all psychological functions. The characteristics of this variability and our capacity to capture it depends in part on the precision of our measures. An individual will be considered better at a task at one time point than another, as a function of endogenous fluctuations in brain states and cognitive abilities as well as variability in exogenous factors (e.g., measurement precision). As a consequence, an individual's ability is better represented as a distribution than as a single value (**Figure 2a**). The central tendency of the distribution reflects one's average ability, whereas the width of the distribution reflects its variability. This account also applies to hypnotic suggestibility. An individual's responsiveness to specific suggestions will vary slightly over time, such that each measurement corresponds to a single sample from their ability distribution. The sample position will be determined on the basis of an interaction between the foregoing *endogenous* and *exogenous* factors. This means that a single sample from one's distribution will not necessarily be representative of one's ability. Rather, reliably estimating the modal response requires obtaining a sufficiently large number of precise samples of the distribution. One of the consequences of this approach is that variability will differ within and between individuals for specific suggestions and across suggestions. The features of these distributions can thus provide useful insights into the characteristics of an individual's ability.

[Figure 2 about here]

Measurement precision

Multiple features of the standard scales attenuate measurement precision with consequences for their utility in basic and applied contexts. Here we describe three salient features.

Binary scoring

In nearly all of the standard scales, responsiveness to each suggestion is scored dichotomously. This scoring format assumes that hypnotic responding is an all-or-nothing phenomenon—that one is either responsive or not. It further assumes that the scoring criterion is optimal—namely, that the threshold set by the scale for passing the suggestion is placed at the optimal position between the two polar categories of complete non-responsiveness (fail) and responsiveness (pass) (see **Figure 2)**.

If we accept that responsiveness to a suggestion is best represented by a distribution, then the limitations of binary scoring become readily evident (**Figure 2**). First, if the criterion happens to fall near the middle of a participant's distribution (as will happen for moderately responsive individuals), the measurement error will be more pronounced as they will be coded as non-responsive approximately 50% of the time. A binary criterion may also exaggerate differences between two individuals due to the placement of the criterion (**Figures 2b and 2d**) or conversely fail to discriminate between two individuals who are both coded as responsive yet who clearly have different distributions (**Figure 2c**). Binary scoring may similar fail to fail to distinguish two individuals with different distributions who were sampled at different locations in their respective distributions (**Figure 2e**). The optimality of the criterion thus becomes a very critical issue with significant impact for measurement.

The issue of the optimality of the criterion position also relates to the confounding of content and difficulty in the standard scales (Woody, 1997; Woody & Barnier, 2008). Specifically, it is unclear whether the greater pass rates of motor suggestions than cognitive-perceptual suggestions are due to differences in difficulty or content-specific ability. This limitation, as well as discrepancies between behavioural and experiential scoring (Cunningham & Ramos, 2012) are partly driven by the use of binary scoring. Indeed, binary scoring provides no information regarding the magnitude of item-level responsiveness. Recognition of this limitation is implicit in various experiential measures, which typically involve Likert or continuous scoring (Kirsch et al., 1990; Kirsch et al., 1998)(see also (McConkey et al., 1999). The introduction of Likert or continuous scoring circumvents most of these limitations and largely obviates concerns regarding the criterion optimality. One might counter that summary scores (see below) address these issues but item-level analyses are routinely performed and low item-level resolution will

still reduce summary score precision. Binary scoring could be advantageous in certain contexts, such as in the use of psychophysical methods (see below), but this requires multiple trials to provide data that can be modelled.

Single-trial sampling

Aside from a single, rarely used, scale (Weitzenhoffer & Hilgard, 1967), all of the standard scales measure responsiveness to each suggestion only once. This amounts to a single sample drawn from an individual's response distribution (**Figures 2e and 2f**) to the detriment of precise measurement.

Single-trial sampling assumes that a single measurement is sufficient to obtain a reliable estimate of responsiveness. Although hypnotic suggestibility scales are moderately to highly reliable over extended periods of time (Piccione et al., 1989), this applies only to summary scores: there has been little to no research on reliability at the individual item level. The available evidence demonstrates that single-item assessment is not a robust approximation of proper sampling of an individual's distribution. An illustrative example comes from the RSPS posthypnotic verbal compulsion suggestion (Weitzenhoffer & Hilgard, 1967) to speak a specific word in a compulsive manner upon presentation of different cues on seven trials. The raw data (Lauer, 1966) reveals that 64% of participants responded on 0 trials, 11% on one, 13% on two to four, and 12% on five or more. This clearly demonstrates that responding on a single trial is not a reliable indicator of responsiveness on subsequent trials. There is clear variability in those who respond on one or more trials and this variability is completely lost when only a single trial is administered (**Figure 2f**). This highlights how single-trial sampling misses out on intra- and inter-individual differences in variability (see also McConkey et al., 1999; Terhune, 2015), which are necessary to properly characterize responsiveness to specific suggestions (**Figure 2d**).

Control Conditions

Control conditions represent a basic, but crucial, component of most psychological tasks. Nearly all assessments of suggestions in the standard scales do not include a control condition, for which baseline responsiveness is merely assumed. For most suggestions, this assumption is reasonable. For example, assessment of baseline hallucination-proneness is probably unnecessary for hallucination suggestions in

most cases. However, it represents a significant limitation for other suggestions, such as the arm heaviness suggestion, present on many standard scales. This requires participants to hold their left arm straight out; they subsequently receive a suggestion for arm heaviness and responses are scored as a pass if their arm lowers by a specific amount. Without a control condition, it is assumed that participants' arms would not normally fall under such circumstances. However, the general population varies in how long they can perform this action, due to age and physical fitness, and some participants' arms begin to fall prior to the administration of the suggestion. This can also promote confusion when participants estimate their own responsiveness. Similarly, participants are given the suggestion to have a dream in multiple standard scales, without a control condition. This omission is problematic because many participants experience vivid spontaneous imagery following an induction (Cardeña, 2005; Pekala & Kumar, 2007; Terhune & Cardeña, 2010a). Anecdotally, we can state that many participants report that this suggestion was easy because they were already experiencing spontaneous imagery. The absence of a control condition in these and other suggestions (e.g., posthypnotic amnesia; (Freedman, 2012)) affects the precision of response estimates—in these cases by introducing a bias towards higher measured responsiveness—and thereby reduces the reliability and validity of the measurement of these suggestions.

Optimization and standardization

Independent of the issues raised above that directly pertain to measurement precision, the standard scales include multiple features that are suboptimal and thereby contribute to the further attenuation of the precision and utility of these measures.

Suboptimal inductions

A salient limitation of the standard scales is the lack of optimization of hypnotic inductions (Terhune & Cardeña, 2016). The available evidence indicates that inductions effect a small, but significant, increase in suggestibility (Braffman & Kirsch, 1999; Tart & Hilgard, 1966). Although the essential features of an induction and its mechanism of action remain poorly understood, standard scale inductions may impede robust measurement. Some are excessively long (e.g., (Shor & Orne, 1962) although there is no evidence

that this is beneficial. They are also relaxation-based, while empirical evidence suggests that inductions encouraging absorption or reduced critical thought are more effective (Brown et al., 2001). Furthermore, most standard scale inductions include suggestions (e.g., bodily heaviness) that are unnecessary (Bányai & Hilgard, 1976) and make reference to connoted or ill-defined concepts (e.g., sleep, hypnotic depth). These features may influence response to subsequent suggestions (e.g., arm paralysis, dreams) or generate unwanted sequelae, thereby introducing further noise into estimates of hypnotic suggestibility.

Suboptimal suggestion content

Insofar as the standard scales are widely used as screening instruments for identifying individuals with relevant abilities, the content of these scales has considerable importance. In particular, they should include suggestions that reliably discriminate individuals of different levels of hypnotic suggestibility and provide information regarding intra-individual variability and variability within levels of hypnotic suggestibility. This is currently not the case. An example comes from a study that applied latent profile analysis to the response patterns of subscales from the *Stanford Profile Scales* (SPS; (Weitzenhoffer & Hilgard, 1963) in medium and highly suggestible individuals (Terhune, 2015), which revealed four distinct classes of participants (or patterns of responding). Crucially, the subscale comprising dreams and regressions was unable to discriminate between the three classes of highly suggestible participants, one of which was mostly comprised of hypnotic virtuosos. By contrast, the two best subscales for class discrimination were the inhibitory cognitive and posthypnotic suggestions. These results are salient because dream and regression suggestions are well-represented on the standard scales whereas inhibitory cognitive suggestions are absent other than posthypnotic amnesia, which has poor psychometric properties. Adding to this, another issue alluded to above is that the standard scales include multiple suggestions that are characterized by high levels of compliant responding (e.g., Bowers et al., 1988), which further contaminates hypnotic suggestibility estimates. It follows that the content of the standard scales is not optimized for the precise measurement of hypnotic suggestibility.

Suggestion coherence

The wording of suggestions in the standard scales varies considerably across items, introducing further noise into response measurement. The impact on reliability and validity of the scales is potentially minor in scope or not immediately clear. Nevertheless, this represents a limitation that could be easily rectified and thus should be addressed in future measures of hypnotic suggestibility.

The standard scales routinely conflate behavioural and experiential scoring as the scales are identified as behavioural, but include multiple suggestions that are scored on purely experiential terms (e.g., dreams, hallucinations). The use of different scoring metrics for different suggestions increases measurement noise. Moreover, the notion of behavioural scoring on the HGSHS:A is misleading because it still relies on participants' estimates of their own responses (Varga et al., 2012). Experiential scoring bears considerable utility (Kihlstrom, 2008; Woody & Szechtman, 2007) but there are meaningful dissociations between behavioural and experiential responses (e.g., Cardeña & Terhune, 2014) that are potentially lost when the two are conflated.

Suggestions on the standard scales also frequently, yet non-uniformly, conflate imagination and suggestion (Terhune & Oakley, 2020). In the HGSHS:A, participants are asked to imagine that their eyelids are glued-shut in the eye catalepsy suggestion, but a reference to gluing is absent from other germane suggestions, such as the finger lock suggestion. Inconsistent references to imagery across items reduce our certainty that *only* response to suggestion is being measured, as opposed to imagery ability, thereby introducing further measurement noise. The available evidence indicates that imagination and suggestion are distinct and imagination does not seem to facilitate response to suggestion (Terhune & Oakley, in press) although invocation of imagery may selectively benefit responsiveness in low and medium suggestible individuals (Scacchia & De Pascalis, 2020). To ensure proper standardization of the scales, it is necessary to more closely match different suggestions regarding any references to imagination.

The standard scales possess other limitations pertaining to item coherence. In particular, some scales conflate non-hypnotic and hypnotic suggestibility through the inclusion of non-hypnotic suggestions (e.g., head falling suggestion of the HGSHS:A). Suggestion length and wording (e.g., repetition) also varies

considerably across suggestions. Preliminary evidence implies that repetition enhances response to suggestion (e.g., Braffman & Kirsch, 1999). All of these inter-item differences increase measurement noise. To improve standardization and permit direct contrasts of different suggestions, suggestions should be worded such that they are as similar as possible except for the specific effect being suggested.

Integration with contemporary psychology

One of the most salient consequences of relying on measures developed more than half a century ago is that hypnosis researchers and clinicians are unable to benefit from conceptual, measurement, and empirical advances during this period. As Council (2002) previously argued, "outmoded conceptions of hypnosis and hypnotizability have influenced our measurement methodology" (p. 206). Here, we highlight limitations of the standard scales that relate to a relative failure to address gaps with respect to contemporary measurement methods, knowledge gleaned about hypnosis, and suitability for contemporary uses of hypnosis in research and clinical domains.

Capturing the classic suggestion effect

Although an attenuation in the sense of agency is the core feature of response to suggestion, numerous authors have questioned the extent to which the standard scales, or at least specific items on these scales, reliably capture the classic suggestion effect (Bowers et al., 1988; Cunningham & Ramos, 2012; Weitzenhoffer, 1974, 1980). Despite the wide recognition of unacceptably high levels of compliance on specific suggestions, consensus has not yet emerged among experimental hypnosis researchers regarding how this problem is mitigated (e.g., Balthazard, 1993). As described above, the problem of compliant responding is largely specific to certain suggestions and thus greater effort is required to omit or modify problematic suggestions from the standard scales (e.g., Wieder & Terhune, 2019). Such an effort should be complemented by wider incorporation of experiential scales, which enable greater elucidation of the nuances of involuntariness and compliance in hypnotic responding. Despite the proliferation of such scales, they are not still routinely used and some have been criticized for confounding involuntariness and verisimilitude (Woody & Szechtman, 2007). Multiple researchers have proposed, or presented, methods

for correcting hypnotic suggestibility scores for compliance (Anllo et al., 2017; Bowers et al., 1988; Wieder & Terhune, 2019) but such approaches haven't been subjected to extensive scrutiny and more empirical attention to this problem is necessary to more robustly capture the classic suggestion effect.

Strategy utilization

Although hypnotic suggestibility is typically viewed as a unidimensional characteristic, one of the key insights from recent decades is that highly suggestible individuals do not constitute a homogeneous group in terms of their response to inductions, the types of suggestions to which they respond, and their cognitive profiles (Carlson & Putnam, 1989; McConkey & Barnier, 2004; Sadler & Woody, in revision; Terhune & Cardeña, 2015). Whether this heterogeneity is best captured by componential or typological models is an ongoing source of debate (Barnier et al., in revision) but a wealth of data indicates that individuals of similar levels of hypnotic suggestibility vary considerably in the cognitive strategies they utilize to facilitate response to suggestions (Brown & Oakley, 2004; Galea et al., 2010; Sheehan & McConkey, 1982; Winkel et al., 2006). Preliminary evidence suggests that these strategies reflect the operation of different cognitive mechanisms (e.g., (King & Council, 1998). Despite these potentially important individual differences, to our knowledge, no attempt has yet been made to formally index strategy utilization during response to suggestion as part of a broader assessment of hypnotic suggestibility. Rigorous assessment of strategy utilization has potentially significant implications for clinical applications, as it may help to better differentiate aptitude and motivational elements of hypnotic responding, which may be characterized by different strategy profiles (Jensen et al., 2017). Incorporating measures of strategy utilization into future scales is likely to further improve screening and elucidate latent dimensions of hypnotic responding.

Summary scores

Following from the previous point that response to suggestion is not uniform is the use of a single summary score in the standard scales. Single summary scores implicitly assume that hypnotic responding is a uniform phenomenon – that is, that response to different suggestions is unidimensional and can be summed across individual suggestions (see also (Balthazard, 1993; Woody & McConkey, 2003).

One of the most significant developments in recent decades is that this unidimensional view does not provide a comprehensive account of individual differences in response to suggestions (Barnier et al., in revision; Sadler & Woody, in revision; Terhune, 2015). Rather, a more well-grounded perspective based on empirical data conceptualizes different hypnotic suggestions as tapping different, but related, componential abilities (Woody & McConkey, 2003). For example, response to posthypnotic amnesia suggestions tends to relate only weakly to responsiveness to other hypnotic suggestions (Woody et al., 2005). The utility of adopting a componential orientation is borne out by the findings that certain componential abilities, as indexed by subsets of suggestions drawn from the standard scales, are better predictors of response to suggestions in experimental studies than summary scores (Woody et al., 2005). Similarly, different suggestion subsets differentially capture the classic suggestion effect (Bowers et al., 1988; Polito et al., 2013). This collective body of evidence highlights the importance of intra-individual variability in response to different suggestions. The use of summary scores masks these important dissociations.

Ecological validity

It is prudent for hypnosis scales to include ecologically valid suggestions that are of relevance to both researchers and clinicians. For example, the relatively poor utility of the standard scales in predicting hypnotic treatment outcome in clinical settings (Montgomery et al., 2011) could be greatly improved by including suggestions germane to clinical applications. Similarly, in experimental studies, the inclusion of relevant suggestions can reduce the need for repeated-screening wherein participants are administered a scale and a specific suggestion related to the respective study. By contrast, the standard scales include numerous suggestions that are relics of earlier pre-1960 scales (Balthazard, 1993; Council, 2002). These items provide basic information regarding response to suggestion but are irrelevant to, or unrepresentative of, contemporary research and clinical applications. Examples of unecological suggestions include arm heaviness, feeling one's hands drawn apart (or together), eye catalepsy, and dreams.

What is especially striking is that in addition to including uninformative suggestions, the standard scales omit suggestions of far greater value. In particular, hypnotic analgesia, arguably the most widely

studied hypnotic phenomenon and one with perhaps the clearest clinical significance, is not represented on the standard scales. Likewise, suggestions that have been the focus of research in recent years such as those modulating attention (Raz et al., 2006), delusions (Connors, 2015) and affect (Ludwig et al., 2013) are absent (see **Table 1**). Developed with these omissions in mind, the *Tailored Stanford Hypnotic Susceptibility Scale: Form C* (TSHSS:C; (Hilgard, Crawford, Bowers, et al., 1979) involves substituting a small pool of irrelevant suggestions for relevant SHSS:C suggestions. This scale is used only infrequently but it underscores the need for both relevant and flexible content on standard scales.

Mental Chronometry and psychophysics

Hypnotic suggestibility scales are unique among psychometric measures as they are work sample instruments that partly resemble both traditional neuropsychological tests as well as personality scales. This is striking as most contemporary hypnosis researchers conceptualize hypnotic suggestibility as a (meta)cognitive ability that interacts with contextual factors. Within contemporary psychology, such abilities are typically indexed by computer-based tasks, which enable numerous features that will strengthen measurement precision of hypnotic suggestibility assessment above and beyond traditional paper-and-pencil tests. For example, computer assessment enables greater standardization, shifting away from self-scoring of behavioural responses, which has multiple limitations (Varga et al, 2012), and randomization of suggestions, which mitigates the emergence of artefactual interdependencies between suggestions (see also (Benham et al., 2006; Woody et al., 2005). Here we highlight two especially valuable advantages of computerized assessment of hypnotic suggestibility.

Chronometry refers to the use of response times (RTs) to gauge performance in a cognitive task. In the wake of behaviourism, RT-based measures have formed the basis for the empirical assessment of cognition as they provide valuable information regarding the speed at which a stimulus is processed, a decision is made, and a response is implemented. RT variability across individuals is readily apparent on the standard scales but goes unmeasured. For example, in suggestions for arm heaviness, it is common to observe participants whose arm falls down very rapidly and others whose arm falls down very slowly. Hull (Hull, 1933) primarily used RTs as the dependent variable when gauging responses to different

suggestions. A wide variety of studies have demonstrated that specific hypnotic suggestions impact RTs on a range of behavioural tasks (for a review, see Oakley & Halligan, 2009). Despite these findings, responses on the standard scales are scored in a simple binary manner without regard to the speed of one's response. Such an approach masks important variability across individuals. Indexing RTs during response to suggestions offers unique opportunities to substantially augment measurement precision.

As noted above, single-trial sampling is inadequate for robustly measuring response magnitude to a suggestion. By contrast, the use of multi-trial sampling enables a psychophysical approach involving the presentation of stimuli or tests of responsiveness at different magnitudes or levels of difficulty, respectively (Kingdom & Prins, 2010). This method would allow for a more sensitive estimate of an individual's level of responsiveness to suggestion, such as a quantification of the magnitude of the attenuation of an auditory stimulus experienced following a negative auditory hallucination suggestion, which is inaccessible using binary scoring on a single item. A simple version of this approach is applied on multiple items of the RSPSs (Weitzenhoffer & Hilgard, 1967) and psychophysics has been occasionally used to study hypnotic suggestion (Tataryn & Kihlstrom, 2017). A psychometric variant of this method is applied in the standard scales through the inclusion of items of increasing difficulty levels. Nevertheless, it is not employed at the individual item level, resulting in reduced precision of the measurement of responsiveness to individual suggestions.

Summary and recommendations for future scale development

Nearly all clinical and experimental hypnosis research uses hypnotic suggestibility scales that were developed more than half a century ago, or similar derivatives. These scales constitute excellent measures for the time period in which they were developed but the myriad limitations we and others have highlighted negatively impact the measurement of hypnotic suggestibility as well as their utility. All current knowledge of hypnosis is based on the use of these scales and their limitations now impede progress in our understanding of suggestion-based phenomena. These limitations warrant the

development of new scales. On the basis of the foregoing critique, we conclude by briefly outlining some prescriptions for the next generation of hypnosis scales.

The development of the next generation of hypnotic suggestibility scales should aim to increase measurement precision and more robustly index hypnotic responding by recognizing responsiveness to individual suggestions as a distribution and depart from the assumption that hypnotic suggestibility is unidimensional. Measurement precision can be augmented by assessing individual suggestions with multiple trials with continuous or pseudo-continuous scoring methods that incorporate behavioural and experiential assessments (Bowers, 1981; Polito et al., 2014). These approaches should be further strengthened through the inclusion of control trials, randomization of suggestions, and chronometric and psychometric methods. Suggestions should be diverse (see Table 1), flexible and relevant to widely-studied clinical and experimental hypnotic phenomena. They should also be representative of the broad domain of hypnosis and designed so as to measure a putative general ability and ancillary componential abilities (Woody et al., 2005), and index individual differences in strategy utilization. Future scales should further aim to optimize hypnotic inductions and suggestion content in order to achieve suggestion coherence and greater standardization.

Many of our proposed changes require a shift from traditional pencil-and-paper measures to computer-based assessment of hypnotic suggestibility. Such a shift opens up novel opportunities for online administration of scales (Palfi et al., 2019; Wieder & Terhune, 2019), harnessing smartphone technology, and complementing behavioural assessment with physiological measures (e.g., pupillometry). Some issues we have identified are unresolved (e.g., induction optimization) and will be difficult to address in the next round of measures. Nevertheless, our critique and corresponding prescriptions offer numerous paths for future research and scale development. We are confident that the implementation of these prescribed changes will significantly augment the utility and precision of hypnotic suggestibility screening with substantive consequences for our understanding of this ability. Just as the development of modern scales helped to usher in a mature scientific discipline devoted to the study of hypnosis in the 20th century (Hilgard, 1965), the development of the next generation of measures of hypnotic suggestibility

promises to radically improve this research domain, provide a stronger integration with germane research fields, and significantly advance our understanding of the impact of verbal suggestions on cognition and perception.

Acknowledgments
This work was supported by Bial Foundation bursaries 198/12 and 144/14 and a Marie Skłodowska-Curie Intra-European Fellowship within the 7th European Community Framework Programme to DBT.

References

Anllo, H., Becchio, J., & Sackur, J. (2017). French Norms for the Harvard Group Scale of Hypnotic Susceptibility, Form A. *Int J Clin Exp Hypn, 65*(2), 241-255. https://doi.org/10.1080/00207144.2017.1276369

Balthazard, C. G. (1993). The hypnosis scales at their centenary: some fundamental issues still unresolved. *Int J Clin Exp Hypn, 41*(1), 47-73. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8425775

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. http://www.ncbi.nlm.nih.gov/pubmed/767390

Barber, T. X. (1965). Measuring "hypnotic-like" suggestibility with and without "hypnotic induction": Psychometric properties, norms, and variables influencing response to the Barber Suggestibility Scale (BSS). *Psychological Reports, 16*, 809-844. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14303013

Barber, T. X., & Wilson, S. C. (1978). The Barber Suggestibility Scale and the Creative Imagination Scale: Experimental and clinical applications [Research Support, U.S. Gov't, P.H.S.]. *American Journal of Clinical Hypnosis, 21*(2-3), 84-108. http://www.ncbi.nlm.nih.gov/pubmed/747170

Barnes, S. M., Lynn, S. J., & Pekala, R. J. (2009). Not all group hypnotic suggestibility scales are created equal: Individual differences in behavioral and subjective responses. *Conscious Cogn, 18*(1), 255-265. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18778955

Barnier, A. J., Dienes, Z., & Mitchell, C. J. (2008). How hypnosis happens: New cognitive theories of hypnotic responding. In M. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 141-178). Oxford University Press.

Barnier, A. J., & McConkey, K. M. (2004). Defining and identifying the highly hypnotizable person. In M. Heap, R. J. Brown, & D. A. Oakley (Eds.), *The highly hypnotizable person: Theoretical, experimental and clinical issues* (pp. 30-60). Brunner-Routledge.

Barnier, A. J., Terhune, D. B., Polito, V., & Woody, E. Z. (in revision). A componential approach to individual differences in hypnotizability.

Benham, G., Woody, E. Z., Wilson, K. S., & Nash, M. R. (2006). Expect the unexpected: Ability, attitude, and responsiveness to hypnosis. *Journal of Personality and Social Psychology, 91*(2), 342-350. https://doi.org/10.1037/0022-3514.91.2.342

Bowers, K. S. (1981). Do the Stanford Scales tap the "classic suggestion effect"? *International Journal of Clinical and Experimental Hypnosis, 29*(1), 42-53. https://doi.org/10.1080/00207148108409142

Bowers, K. S. (1993). The Waterloo-Stanford Group C (WSGC) scale of hypnotic susceptibility: normative and comparative data. *Int J Clin Exp Hypn, 41*(1), 35-46. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8425774

Bowers, P., Laurence, J. R., & Hart, D. (1988). The experience of hypnotic suggestions. *International Journal of Clinical and Experimental Hypnosis, 36*(4), 336-349. https://doi.org/10.1080/00207148808410523

Braffman, W., & Kirsch, I. (1999). Imaginative suggestibility and hypnotizability: An empirical analysis. *Journal of Personality and Social Psychology, 77*(3), 578-587. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10510510

Brown, R. J., Antonova, E., Langley, A., & Oakley, D. A. (2001). The effects of absorption and reduced critical though on suggestibility in an hypnotic context. *Contemporary hypnosis, 18*, 62-72.

Brown, R. J., & Oakley, D. A. (2004). An integrative cognitive theory of hypnosis and high hypnotizability. In M. Heap, R. J. Brown, & D. A. Oakley (Eds.), *The highly hypnotizable person: Theoretical, experimental and clinical issues* (pp. 152-186). Brunner-Routledge.

Bryant, R. A., Guthrie, R. M., & Moulds, M. L. (2001). Hypnotizability in acute stress disorder. *American Journal of Psychiatry, 158*(4), 600-604. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11282695

Cardeña, E. (2005). The phenomenology of deep hypnosis: Quiescent and physically active. *International Journal of Clinical and Experimental Hypnosis, 53*(1), 37-59. https://doi.org/10.1080/00207140490914234

Cardeña, E., & Terhune, D. B. (2014). Hypnotizability, personality traits and the propensity to experience alterations of consciousness. *Psychology of Consciousness: Theory, Research, and Practice, 1*, 292-307.

Carlson, E. B., & Putnam, F. W. (1989). Integrating research on dissociation and hypnotizability: Are there two pathways to hypnotizability? *Dissociation, 2*, 32-38.

Coltheart, M., Cox, R., Sowman, P., Morgan, H., Barnier, A., Langdon, R., Connaughton, E., Teichmann, L., Williams, N., & Polito, V. (2018). Belief, delusion, hypnosis, and the right dorsolateral prefrontal cortex: A transcranial magnetic stimulation study. *Cortex, 101*, 234-248. https://doi.org/10.1016/j.cortex.2018.01.001

Connors, M. H. (2015). Hypnosis and belief: A review of hypnotic delusions. *Conscious Cogn, 36*, 27-43. https://doi.org/10.1016/j.concog.2015.05.015

Council, J. R. (2002). A historical overview of hypnotizability assessment. *Am J Clin Hypn, 44*(3-4), 199-208. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11799534

Cunningham, P. F., & Ramos, P. (2012). Involuntary experiencing and the performance of hypnotic test suggestions. *Int J Clin Exp Hypn, 60*(4), 416-431. https://doi.org/10.1080/00207144.2012.701090

Dienes, Z., & Hutton, S. (2013). Understanding hypnosis metacognitively: rTMS applied to left DLPFC increases hypnotic suggestibility. *Cortex, 49*(2), 386-392. https://doi.org/10.1016/j.cortex.2012.07.009

Elkins, G. (2014). *Hypnotic relaxation therapy: Principles and applications*. Springer Publishing Co.

Elkins, G. R., Barabasz, A. F., Council, J. R., & Spiegel, D. (2015). Advancing Research and Practice: The Revised APA Division 30 Definition of Hypnosis. *American Journal of Clinical Hypnosis, 57*(4), 378-385. https://doi.org/10.1080/00029157.2015.1011465

Field, P. B. (1965a). An inventory scale of hypnotic depth. *The International journal of clinical and experimental hypnosis, 13*(4), 238-249. https://doi.org/10.1080/00207146508412946

Field, P. B. (1965b). An inventory scale of hypnotic depth. *Int J Clin Exp Hypn, 13*(4), 238-249. https://doi.org/10.1080/00207146508412946

Freedman, S. (2012). *The measurement of posthypnotic amnesia with the Harvard Group Scale of Hypnotic Susceptibility: Form A* Concordia University].

Galea, V., Woody, E. Z., Szechtman, H., & Pierrynowski, M. R. (2010). Motion in response to the hypnotic suggestion of arm rigidity: A window on underlying mechanisms. *International Journal of Clinical and Experimental Hypnosis, 58*(3), 251-268. https://doi.org/10.1080/00207141003760561

Haggard, P. (2017). Sense of agency in the human brain. *Nature reviews. Neuroscience, 18*(4), 196-207. https://doi.org/10.1038/nrn.2017.14

Haggard, P., Cartledge, P., Dafydd, M., & Oakley, D. A. (2004). Anomalous control: When 'free-will' is not conscious. *Conscious Cogn, 13*(3), 646-654. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15336254

Hilgard, E. R. (1965). *Hypnotic susceptibility*. Harcourt, Brace & World, Inc.

Hilgard, E. R., Crawford, H. J., Bowers, P., & Kihlstrom, J. F. (1979). A tailored SHSS:C permitting user modification for special purposes. *International Journal of Clinical and Experimental Hypnosis, 27*(2), 125-133. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=541130

Hilgard, E. R., Crawford, H. J., & Wert, A. (1979). The Stanford Hypnotic Arm Levitation Induction and Test (SHALIT): A six-minute hypnotic induction and measurement scale. *International Journal of Clinical and Experimental Hypnosis, 27*(2), 111-124.

Hilgard, E. R., Weitzenhoffer, A. M., & Gough, P. (1958). Individual Differences in Susceptibility to Hypnosis. *Proceedings of the National Academy of Sciences of the United States of America, 44*(12), 1255-1259. https://doi.org/10.1073/pnas.44.12.1255

Hull, C. L. (1933). *Hypnosis and suggestibility: An experimental approach*. D. Appleton-Century Co.

Kihlstrom, J. F. (2008). The domain of hypnosis, revisited. In M. R. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis* (pp. 21–52). Oxford University Press.

King, B. J., & Council, J. R. (1998). Intentionality during hypnosis: An ironic process analysis. *International Journal of Clinical and Experimental Hypnosis, 46*(3), 295-313. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9650441

Kingdom, F. A. A., & Prins, N. (2010). Psychophysics: A practical introduction. In. Elsevier.

Kirsch, I. (1997). Suggestibility or hypnosis: what do our scales really measure? *Int J Clin Exp Hypn, 45*(3), 212-225. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9204635

Kirsch, I. (1999). Hypnosis and placebos: Response expectancy as a mediator of suggestion effects. *Anales de Psicologia, 15*, 99-110.

Kirsch, I., & Braffman, W. (2001). Imaginative suggestibility and hypnotizability. *Current Directions in Psychological Science, 10*(2), 57-61. https://doi.org/10.1111/1467-8721.00115

Kirsch, I., Council, J. R., & Wickless, C. (1990). Subjective scoring for the Harvard Group Scale of Hypnotic Susceptibility, Form A. *International Journal of Clinical and Experimental Hypnosis, 38*(2), 112-124. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2347668

Kirsch, I., Milling, L. S., & Burgess, C. (1998). Experiential scoring for the Waterloo-Stanford Group C scale. *International Journal of Clinical and Experimental Hypnosis, 46*(3), 269-279. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9650439

Lauer, L. W. (1966). *Factorial components of hypnotic susceptibility* [Ph.D. dissertation, Stanford University].

Laurence, J.-R., Beaulieu-Prévost, D., & du Chéné, T. (2008). Measuring and understanding individual differences in hypnotizability. In M. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research and practice* (pp. 225-253). Oxford University Press.

London, P. (1963). *Children's Hypnotic Susceptibility Scale*. Consulting Psychologists Press.

London, P. (1988). *The Children's Hypnotic Susceptibility Scale*. Consulting Psychologists Press.

Ludwig, V. U., Stelzel, C., Krutiak, H., Magrabi, A., Steimke, R., Paschke, L. M., & Walter, H. (2013). The suggestible brain: Posthypnotic effects on value-based decision-making. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nst110

Lush, P., Moga, G., McLatchie, N., & Dienes, Z. (2018). The Sussex-Waterloo Scale of hypnotisability (SWASH): Measuring capacity for altering conscious experience. *Neuroscience of Consciousness, 4*, niy006.

Lynn, S. J., Laurence, J.-R., & Kirsch, I. (2015). Hypnosis, suggestion, and suggestibility: An integrative model. *American Journal of Clinical Hypnosis, 57*(3), 314-329. https://doi.org/10.1080/00029157.2014.976783

McConkey, K. M., & Barnier, A. J. (2004). High hypnotizability: Unity and diversity in behavior and experience. In M. Heap, R. J. Brown, & D. A. Oakley (Eds.), *The highly hypnotizable person: Theoretical, experimental and clinical issues* (pp. 61-84). Routledge.

McConkey, K. M., Wende, V., & Barnier, A. J. (1999). Measuring change in the subjective experience of hypnosis. *Int J Clin Exp Hypn, 47*(1), 23-39. https://doi.org/10.1080/00207149908410020

Montgomery, G. H., Schnur, J. B., & David, D. (2011). The impact of hypnotic suggestibility in clinical care settings. *Int J Clin Exp Hypn, 59*(3), 294-309. https://doi.org/10.1080/00207144.2011.570656

Moore, J. W. (2016). What Is the sense of agency and why does it matter? *Frontiers in psychology, 7*, 1272. https://doi.org/10.3389/fpsyg.2016.01272

Moran, T. E., Kurtz, R. M., & Strube, M. J. (2002). The efficacy of the Waterloo-Stanford Group Scale of hypnotic susceptibility: Form C. *Am J Clin Hypn, 44*(3-4), 221-230. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11799536

Morgan, A. H. (1973). The heritability of hypnotic susceptibility in twins. *J Abnorm Psychol, 82*(1), 55-61. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4738327

Morgan, A. H., & Hilgard, J. R. (1978/1979). The Stanford Hypnotic Clinical Scale for adults [Research Support, U.S. Gov't, P.H.S.]. *American Journal of Clinical Hypnosis, 21*(2-3), 134-147. http://www.ncbi.nlm.nih.gov/pubmed/747162

Oakley, D. A., & Halligan, P. W. (2009). Hypnotic suggestion and cognitive neuroscience. *Trends Cogn Sci, 13*(6), 264-270. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19428287

Oakley, D. A., & Halligan, P. W. (2013). Hypnotic suggestion: Opportunities for cognitive neuroscience. *Nature Reviews Neuroscience, 14*(8), 565-576. https://doi.org/10.1038/nrn3538

Palfi, B., Moga, G., Lush, P., Scott, R. B., & Dienes, Z. (2019). Can hypnotic suggestibility be measured online? *Psychological Research*. https://doi.org/10.1007/s00426-019-01162-w

Parris, B. A. (2017). The Role of Frontal Executive Functions in Hypnosis and Hypnotic Suggestibility. *Psychology of Consciousness: Theory, Research, and Practice, 42*, 211-229. https://doi.org/10.1037/cns0000106

Pekala, R. J. (1991). *The phenomenology of consciousness inventory*. West Chester, PA: Mid-Atlantic Educational Institute, Inc.

Pekala, R. J., & Kumar, V. K. (2007). An empirical-phenomenological approach to quantifying consciousness and states of consciousness: With particular reference to understanding the nature of hypnosis. In G. A. Jamieson (Ed.), *Hypnosis and conscious states: The cognitive neuroscience perspective* (pp. 167-194). Oxford University Press.

Pekala, R. J., Kumar, V. K., Maurer, R., Elliott-Carter, N., Moon, E., & Mullen, K. (2010). Suggestibility, expectancy, trance state effects, and hypnotic depth: I. Implications for understanding hypnotism. *Am J Clin Hypn, 52*(4), 275-290. http://www.ncbi.nlm.nih.gov/pubmed/20499542

Piccione, C., Hilgard, E. R., & Zimbardo, P. G. (1989). On the degree of stability of measured hypnotizability over a 25-year period. *Journal of Personality and Social Psychology, 56*(2), 289-295. http://www.ncbi.nlm.nih.gov/pubmed/2926631

Polczyk, R., & Pasek, T. (2006). Types of suggestibility: Relationships among compliance, indirect, and direct suggestibility. *Int J Clin Exp Hypn, 54*(4), 392-415. https://doi.org/10.1080/00207140600856764

Polito, V., Barnier, A. J., & Woody, E. Z. (2013). Developing the Sense of Agency Rating Scale (SOARS): An empirical measure of agency disruption in hypnosis. *Consciousness and Cognition, 22*(3), 684-696. https://doi.org/10.1016/j.concog.2013.04.003

Polito, V., Barnier, A. J., Woody, E. Z., & Connors, M. H. (2014). Measuring agency change across the domain of hypnosis. *Psychology of Consciousness: Theory, Research, and Practice, 1*(1), 3-19.

Raz, A., Kirsch, I., Pollard, J., & Nitkin-Kaner, Y. (2006). Suggestion reduces the stroop effect. *Psychological Science, 17*(2), 91-95. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16466414

Rominger, C., Weiss, E. M., Nagl, S., Niederstatter, H., Parson, W., & Papousek, I. (2014). Carriers of the COMT Met/Met allele have higher degrees of hypnotizability, provided that they have good attentional control: A case of gene-trait interaction. *Int J Clin Exp Hypn, 62*(4), 455-482. https://doi.org/10.1080/00207144.2014.931177

Sadler, P., & Woody, E. (in revision). Multicomponent theories of hypnotizability: History and prospects.

Sadler, P., & Woody, E. Z. (2004). Four decades of group hypnosis scales: What does item-response theory tell us about what we've been measuring? *Int J Clin Exp Hypn, 52*(2), 132-158. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15115059

Scacchia, P., & De Pascalis, V. (2020). Effects of Prehypnotic Instructions on Hypnotizability and Relationships Between Hypnotizability, Absorption, and Empathy. *Am J Clin Hypn, 62*(3), 231-266. https://doi.org/10.1080/00029157.2019.1586639

Sheehan, P. W., & McConkey, K. M. (1982). *Hypnosis and experience: The exploration of phenomena and process*. Lawrence Erlbaum.

Shor, R. E., & Orne, E. C. (1962). *Harvard Group Scale of Hypnotic Susceptibility, Form A*. Consulting Psychologists Press.

Spanos, N. P., & Gorassini, D. R. (1984). Structure of hypnotic test suggestions and attributions of responding involuntarily. *Journal of Personality and Social Psychology, 46*(3), 688-696. https://doi.org/Doi 10.1037//0022-3514.46.3.688

Spanos, N. P., Radtke, H. L., Hodgins, D. C., Stam, H. J., & Bertrand, L. D. (1983). The Carleton University responsiveness to suggestion scale: Normative data and psychometric properties [Research Support, Non-U.S. Gov't]. *Psychological Reports, 53*(2), 523-535. https://doi.org/10.2466/pr0.1983.53.2.523

Spiegel, H., & Spiegel, D. (1978). *Trance and treatment: Clinical uses of hypnosis*. Basic Books.

Szechtman, H., Woody, E., Bowers, K. S., & Nahmias, C. (1998). Where the imaginal appears real: A positron emission tomography study of auditory hallucinations. *Proceedings of the National Academy of Sciences of the United States of America, 95*(4), 1956-1960. http://www.ncbi.nlm.nih.gov/pubmed/9465124

Tart, C. T., & Hilgard, E. R. (1966). Responsiveness to suggestions under "hypnosis" and "waking-imagination" conditions: A methodological observation. *International Journal of Clinical and Experimental Hypnosis, 14*(3), 247-256. https://doi.org/10.1080/00207146608412967

Tataryn, D. J., & Kihlstrom, J. F. (2017). Hypnotic tactile anesthesia: Psychophysical and signal-detection analyses. *Int J Clin Exp Hypn, 65*(2), 133-161. https://doi.org/10.1080/00207144.2017.1276358

Terhune, D. B. (2014). Defining hypnosis: The pitfalls of prioritizing spontaneous experience over response to suggestion. *Journal of Mind-Body Regulation, 2*, 116-117.

Terhune, D. B. (2015). Discrete response patterns in the upper range of hypnotic suggestibility: A latent profile analysis. *Consciousness and Cognition, 33*, 334-341. https://doi.org/10.1016/j.concog.2015.01.018

Terhune, D. B., & Cardeña, E. (2010a). Differential patterns of spontaneous experiential response to a hypnotic induction: A latent profile analysis [Research Support, Non-U.S. Gov't]. *Consciousness and Cognition, 19*(4), 1140-1150. https://doi.org/10.1016/j.concog.2010.03.006

Terhune, D. B., & Cardeña, E. (2010b). Methodological and interpretative issues regarding the Phenomenology of Consciousness Inventory – Hypnotic Assessment Procedure: A comment on Pekala et al. *American Journal of Clinical Hypnosis, 53*, 105-113.

Terhune, D. B., & Cardeña, E. (2015). Dissociative subtypes in posttraumatic stress disorders and hypnosis: Neurocognitive parallels and clinical implications. *Current Directions in Psychological Science, 24*, 452-457. https://doi.org/10.1177/0963721415604611

Terhune, D. B., & Cardeña, E. (2016). Nuances and uncertainties regarding hypnotic inductions: Toward a theoretically informed praxis. *Am J Clin Hypn, 59*(2), 155-174. https://doi.org/10.1080/00029157.2016.1201454

Terhune, D. B., Cardeña, E., & Lindgren, M. (2011). Dissociative tendencies and individual differences in high hypnotic suggestibility [Research Support, Non-U.S. Gov't]. *Cogn Neuropsychiatry, 16*(2), 113-135. https://doi.org/10.1080/13546805.2010.503048

Terhune, D. B., Cleeremans, A., Raz, A., & Lynn, S. J. (2017). Hypnosis and top-down regulation of consciousness. *Neuroscience and Biobehavioral Reviews, 81*(Pt A), 59-74. https://doi.org/10.1016/j.neubiorev.2017.02.002

Terhune, D. B., & Cohen Kadosh, R. (2012). The emerging neuroscience of hypnosis. *Cortex, 48*, 382-386.

Terhune, D. B., & Oakley, D. A. (2020). Hypnosis and imagination. In A. Abraham (Ed.), *Cambridge handbook of imagination*. Cambridge University Press.

Varga, K., Farkas, L., & Mero, L. (2012). On the objectivity of the scoring of Harvard Group Scale of Hypnotic Susceptibility. *Int J Clin Exp Hypn, 60*(4), 458-479. https://doi.org/10.1080/00207144.2012.675298

Wagstaff, G. F. (2014). On the centrallity of the concept of an altered state to definitions of hypnosis. *Journal of Mind-Body Regulation, 2*(2), 90-108.

Weitzenhoffer, A. M. (1974). When is an "instruction" an "instruction"? *Int J Clin Exp Hypn, 22*(3), 258-269. https://doi.org/10.1080/00207147408413005

Weitzenhoffer, A. M. (1980). Hypnotic susceptibility revisited. *American Journal of Clinical Hypnosis, 22*(3), 130-146. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7386402

Weitzenhoffer, A. M., & Hilgard, E. R. (1959). *Stanford Hypnotic Susceptibility Scales, Forms A and B*. Consulting Psychologists Press.

Weitzenhoffer, A. M., & Hilgard, E. R. (1962). *Stanford Hypnotic Susceptibility Scale: Form C*. Consulting Psychologists Press.

Weitzenhoffer, A. M., & Hilgard, E. R. (1963). *Stanford Profile Scales of Hypnotic Susceptibility, Forms I and II*. Consulting Psychologists Press.

Weitzenhoffer, A. M., & Hilgard, E. R. (1967). *Revised Stanford Profile Scales of Hypnotic Susceptibility: Forms I and II*. Consulting Psychologists Press.

Whalley, M. G., & Brooks, G. B. (2009). Enhancement of suggestibility and imaginative ability with nitrous oxide. *Psychopharmacology, 203*(4), 745-752.

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19057896

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychol Sci, 16*(10), 780-784. https://doi.org/10.1111/j.1467-9280.2005.01614.x

Wieder, L., & Terhune, D. B. (2019). Trauma and anxious attachment influence the relationship between suggestibility and dissociation: A moderated-moderation analysis. *Cogn Neuropsychiatry*, 1-17. https://doi.org/10.1080/13546805.2019.1606705

Winkel, J. D., Younger, J. W., Tomcik, N., Borckardt, J. J., & Nash, M. R. (2006). Anatomy of a hypnotic response: Self-report estimates, actual behavior, and physiological response to the hypnotic suggestion of arm rigidity. *International Journal of Clinical and Experimental Hypnosis, 54*, 186-205.

Woody, E. Z. (1997). Have the hypnotic susceptibility scales outlived their usefulness? *Int J Clin Exp Hypn, 45*(3), 226-238. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9204636

Woody, E. Z., & Barnier, A. J. (2008). Hypnosis scales for the twenty-first century: What do we know and how should we use them? In M. Nash & A. J. Barnier (Eds.), *The Oxford handbook of hypnosis: Theory, research and practice* (pp. 255-281). Oxford University Press.

Woody, E. Z., Barnier, A. J., & McConkey, K. M. (2005). Multiple hypnotizabilities: Differentiating the building blocks of hypnotic response. *Psychological Assessment, 17*(2), 200-211. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16029107

Woody, E. Z., & McConkey, K. M. (2003). What we don't know about the brain and hypnosis, but need to: A view from the Buckhorn Inn. *Int J Clin Exp Hypn, 51*(3), 309-338. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12934845

Woody, E. Z., & Szechtman, H. (2007). To see feelingly: Emotion, motivation, and hypnosis. In G. A. Jamieson (Ed.), *Hypnosis and conscious states: The cognitive neuroscience perspective* (pp. 241-256). Oxford University Press.

Younger, J., Kemmerer, D. D., Winkel, J. D., & Nash, M. R. (2005). The Harvard Group Scale of Hypnotic Susceptibility: accuracy of self-report and the memory for items. *Int J Clin Exp Hypn, 53*(3), 306-320. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16076667

Table 1. *Categories of hypnotic suggestion and representative examples.*

Suggestion Type	Psychological function			
	Motor	Cognitive	Perceptual	Affective
Facilitative	Involuntary arm movement	Delusion	Auditory hallucination	Disgust
Inhibitory	Motor paralysis	Amnesia	Analgesia	Emotional numbing

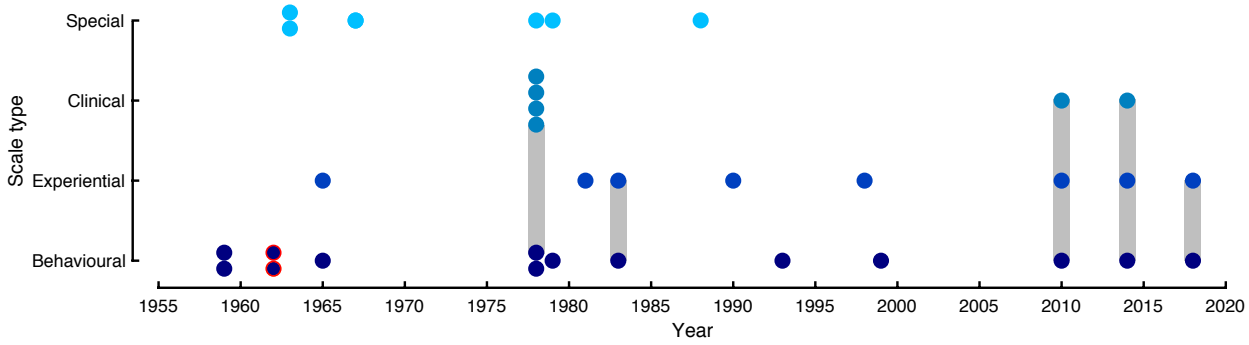

Figure 1. A timeline of standardized measures of hypnotic suggestibility. Scales are distinguished on the basis of whether they were introduced primarily as behavioural, experiential, clinical, or special-population scales with corresponding different colours. Connecting grey bars denote scales that fall into multiple categories whereas red markers denote the two most widely used scales. *Stanford Hypnotic Susceptibility Scale: Form A* (SHSS:A; Weitzenhoffer & Hilgard (1959)); *Stanford Hypnotic Susceptibility Scale: Form B* (SHSS:B; Weitzenhoffer & Hilgard (1959)); *Harvard Group Scale of Hypnotic Susceptibility: Form A* (HGSHS:A; Shor & Orne (1962)); *Stanford Hypnotic Susceptibility Scale: Form C* (SHSS:C; Weitzenhoffer & Hilgard (1962)); *Children's Hypnotic Susceptibility Scale* (CHSS; (London, 1963)); *Stanford Profile Scales of Hypnotic Susceptibility* (SPS; Weitzenhoffer & Hilgard (1963)); *Barber Suggestibility Scale* (BSS; (Barber, 1965)); *Inventory Scale of Hypnotic Depth* (ISHD; (Field, 1965a)); *Revised Stanford Profile Scales of Hypnotic Susceptibility* (RSPS; Weitzenhoffer & Hilgard (1967)); *Diagnostic Rating Scale* (DRS; Weitzenhoffer & Hilgard (1967)); *Hypnotic Induction Profile* (HIP; (Spiegel & Spiegel, 1978)); *Stanford Hypnotic Clinical Scale for Adults* (SHCS:A; (Morgan & Hilgard, 1978/1979)); *Stanford Hypnotic Clinical Scale for Children* (SHCS:C; Morgan & Hilgard, 1978/1979b)); *Creative Imagination Scale* (CIS; (Barber & Wilson, 1978)); *Stanford Hypnotic Arm Levitation* (SHALIT; (Hilgard, Crawford, & Wert, 1979)); *Tailored Stanford Hypnotic Susceptibility Scale: Form C* (TSHSS:C; Hilgard et al. (1979)); *Bowers Involuntariness Scale* (BIS; (Bowers, 1981)); *Carleton University Responsiveness to Suggestion Scale* (CURSS; (Spanos et al., 1983)); *Children's Hypnotic Susceptibility Scale* (CHSS; (London, 1988)); *HGSHS:A Subjective Scale* (HGSHS:A-S; Kirsch et al. (1990)); *Waterloo Stanford Group Scale of Hypnotic Susceptibility: Form C* (WSGC; (Bowers, 1993)); *WSGC Subjective Scale* (WSGC-S; Kirsch et al. (1998)); *Group Scale of Hypnotic Ability* (GSHA; Hawkins & Wenzel, 1999); *PCI Hypnotic Assessment Procedure* (PCI-HAP; Pekala et al., 2010a); *Elkins Hypnotizability Scale* (EHS; (G. Elkins, 2014)); *Sussex-Waterloo Scale of Hypnotizability* (SWASH; (Lush et al., 2018).

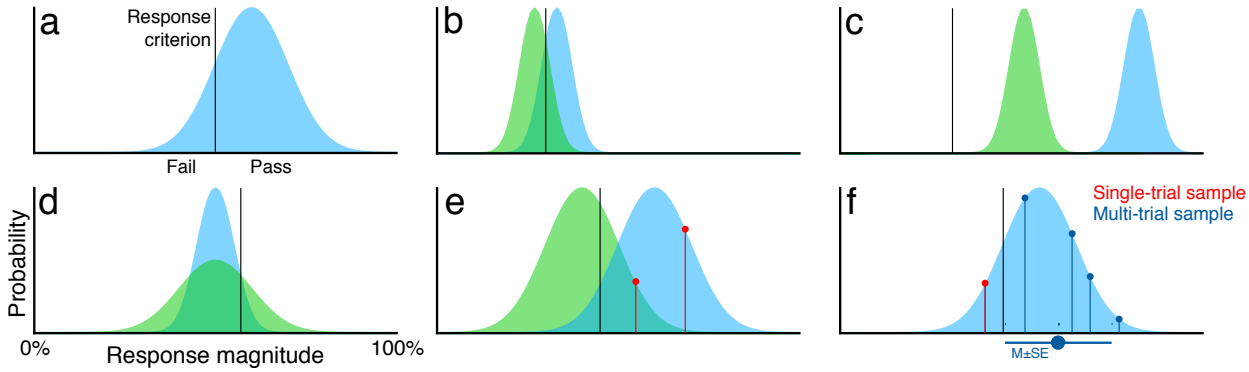

Figure 2. Response to suggestion as a random variable and the consequences of binary scoring and single-trial sampling. (a) In the standard scales, response to a suggestion is coded as fail or pass, according to a response criterion relative to the probability distribution of the response magnitude; (c) Two participants with similar levels of responsiveness who will tend to receive different scores; (d) Two participants with different levels of responsiveness who will receive the same score; (d) Two participants with the same mean level of response to a suggestion but different distribution variances. The participant with higher variance (lower reliability) will tend to pass the item more often; (e) Two participants with different levels of responsiveness to a suggestion will occasionally be coded as being equally responsive; (b) Scoring based on single-trial sampling of a participant's response distribution is more likely to yield a misleading outcome than multi-trial sampling, which can be further strengthened through the use of continuous scoring, enabling better estimation of the mean (*M*) and variability (standard error [*SE*]) of the response distribution.