

Serveur Académique Lausannois **SERVAL** serval.unil.ch

Author Manuscript

Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

Title: Experimental Designs and the 'Emotion Stimulus Critique': Hidden Problems and Potential Solutions in the Study of Emotion.

Authors: Pascual-Leone A, Herpertz SC, Kramer U

Journal: Psychopathology

Year: 2016

Issue: 49

Volume: 1

Pages: 60-8

DOI: 10.1159/000442294

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.

Running Head: EMOTIONAL STIMULI IN EXPERIMENTAL DESIGNS

IN PRESS PSYCHOPATHOLOGY

Experimental designs and the “emotion-stimulus critique”: Hidden problems and potential solutions in the study of emotion

Antonio Pascual-Leone¹, Sabine C. Herpertz², Ueli Kramer^{1,3}

¹ Department of Psychology, University of Windsor, Canada

² Department of Psychiatry, University of Heidelberg, Germany

³ Department of Psychiatry, Institute of Psychotherapy and General Psychiatry, University of Lausanne, Switzerland

The present research was funded by Swiss National Science Foundation International Cooperation grant IZK0Z3_150868/1 (PD Dr. Kramer and Prof. Herpertz).

All correspondence concerning this article should be addressed to PD Dr Ueli Kramer, IUP-Dpt Psychiatry-CHUV, University of Lausanne, Place Chauderon 18, CH-1003 Lausanne, Switzerland, ph. +41-21-314 00 50, fax +41-21-314 27 84; e-mail : Ueli.Kramer@chuv.ch

Running Head: EMOTIONAL STIMULI IN EXPERIMENTAL DESIGNS

Experimental designs and the "emotion-stimulus critique":

Hidden problems and potential solutions in the study of emotion

Experimental designs and the "emotion-stimulus critique": Hidden problems and potential solutions in the study of emotion

Abstract

Emotional experience is increasingly measured using experimental tasks, but the stimuli being used are often only proxies to the emotion being studied. Stimuli are intended to evoke a distinct emotional experience, but certain designs fail to adequately control for the actual experience in question. In this methodological paper, we review designs used in clinical psychology aiming at measuring emotion and develop the argument of the “emotion-stimulus critique”. Designs of neuroimaging studies on emotion in this context are given a preference. We argue that studies often concentrate on the standardization of the stimulus-material (i.e., words, images, movies) for eliciting an emotional experience, whereas the standardization of the actual participant's experience is seldom performed. Our proposal discusses the use of standardized stimuli in experimental designs and contrasts this with the necessity for controlling for a participant's unique emotional response. We highlight the importance of each participant's "inner metric", the individual's experiential anchor, which needs to be taken into account when examining the emotional correlates of psychiatric disorders or psychotherapeutic change. Implications of the emotion stimulus critique for research are discussed within the context of psychology, particularly clinical psychology and psychotherapy.

Key-Words:

Experimental; Methods; Emotion; Emotional induction; Subjective; Stimuli

Introduction

Emotional processing, broadly defined, encompasses several facets such as the perception of emotion, its regulation, and transformation and is a central variable for most mental health problems [1,2]. Emotional processing has been defined as (p. 51), "a process whereby emotional disturbances are absorbed, and decline to the extent that other experiences and behavior can proceed without disruption" [3]. Since then, the field has moved towards understanding change in emotional processing over the course of several forms of psychotherapy, for example dialectical-behavior therapy of borderline personality disorder [4,5] and emotion-focused therapy for depression [6]. Increasingly detailed models underlying the generation of emotion have been proposed (e.g., [7], p. 474) and complex research designs that examine distinct affective-meaning states are now used (for a review see [8]). From a methodological point of view, studying emotion in-vivo has always been challenging, because of the volatile nature of the phenomenon. It has become increasingly common, in particular among researchers in neuroscience, to use experimental tasks to measure emotional change, either as a way of measuring the effects of psychological treatments [5] or to assess prognostic information related to a disorder [9]. We believe that there are a certain number of problems that commonly appear in such experimental designs when conducted in a clinical context. This may explain some of the inconsistent findings in the literature on emotional experience and processing. These potential design problems will be spelled out in the present contribution.

Emotional experience can be generally understood as a two-phase phenomenon [10]. First, there is an emotion-eliciting, or –evoking phase, in which emotion and associated processing are often cued using words, or images, films, etc. [11]. A stimulus trigger, for example, could be an image from the International Affective Pictures System, or a word list from the Affect Norms for English Words [12], or validated databases of film excerpts [13].

The second part of this phenomenon is an emotion-response - or –experience - phase, in which emotions and associated processing are experienced and lived by the participant. However, while great attention is often given to controlling and designing the first phase of this phenomenon, the target of study is often the second phase. In this sense, the stimulus “trigger” being presented in a study (i.e., cues, words, images, etc.) is often not really the intended “stimulus” (i.e., an activated emotional experience) but rather a trigger, or proxy, to elicit said stimulus. Moreover, beyond the general consensus on conceptualizing the phenomenon of emotional experience in this way, there is a great heterogeneity in the degree to which studies optimally take into account the underlying methodological constraints related to internal and external validity of the emotional experience being studied.

Methodologists agree on the fundamental idea that there is a trade-off in most experimental designs between internal and external validity - the more one facilitates one side of this equation, the more the other side is compromised. This overall conflict translates into two opposing methodological constraints and has important implications for any controlled task that uses stimuli-material as a means to indirectly measure a participant’s emotion. On one hand, to ensure internal validity, (a) emotion or emotional processing conditions must be congruent across participants (i.e., between-participant comparability). On the other hand, to ensure external validity: (b) the proposed task that aims to study the subjective experience of emotion should allow one to measure some individually and idiosyncratically relevant process or content (i.e., *subjective relevance*). Research designs studying emotional experiences can therefore be placed on a continuum between controlling for between-participant comparability and the personal relevance to a given participant. This control can be exercised either in terms of the presented stimulus or in terms of a participant’s actual subjective response.

Experimental research on emotional experience is usually designed in a fashion that appeals to face validity and the conventional concerns of experimental design to ensure that

all stimuli in the emotion-eliciting phase are comparable, which is presumed to mean there is acceptable inter-participant comparability. In the meantime, the issue of subjective relevance is not as transparent. This, *inner metric*, the individual's experiential anchor, of each participant is usually overlooked and consequentially several conceptual and or methodological short-cuts are sometimes inadvertently used. For instance, participant's experiences in the emotion-response phase are often diverse and remain uncontrolled for, usually with the tacit assumption that they are comparable, given that presenting trigger stimuli were standardized. This assumption seems reasonable until the experimenter stops to imagine the task within the unique perspective of one or another participant: a methodological consideration first proposed by Tolman and Brunswik [14] and then more fully described by J. Pascual-Leone [15] as taking a *meta-subjective perspective*. This epistemological perspective is characterized by a constructivist investigation of emotion "from within the organism" ([15], p. 739), which contrasts with the description from an external observer's perspective of the phenomenon.

In sum, the present contribution aims at outlining the *emotion-stimulus critique* as a caution to researchers and highlights a number of solutions used to solve the problem. For example, some of the possible solutions discussed include: having participants explicitly label the emotional response, as well using their subjective ratings of emotional intensity from an idiosyncratic perspective to inform the selection of stimuli triggers. Finally, we will discuss the complexity of studying multi-step emotional sequences (as opposed to discrete emotion categories) as well as the problems entailed in studying emotion that is outside an individual's awareness.

The "emotion-stimulus critique": Hidden problems in the study of emotion

Stimuli in experimental designs are often taken to be universal cues to which participants respond. The Stimulus-Response (S-R) model used since early behavioral

experiments, such as in the affective conditioning of “little Albert” [16], are a classic example of this design and basic assumption. This basic S-R model continues to be used today in experimental research, in particularly neuroscience, because it offers a straightforward and convenient paradigm. However, as highlighted by humanistic theory [17,2], the agentic and dynamic role of the participant plays an increasingly important role as one moves from research questions on pure “behavioral responses” to more subjectively interpreted “meaning and feelings.” Indeed, even contemporary behavioral theory has acknowledged that the S-R relationship is mediated on several levels, the most basic of which is that a presenting Stimulus is subjectively interpreted by a participant (as organism; *O*) and then Responded to (S-O-R; [18]). Similarly, contemporary neuroscience has suggested the central role of specific neuronal correlates (i.e., the amygdala) in the detection of *personal salience* of a stimulus [19]. Based on the individual salience of a stimulus, the individual may construct meaning in an idiosyncratic fashion (i.e., when tacitly posed the question: “what does this stimulus mean to me?”). Such individual meaning construction, therefore, becomes an important piece to consider when studying emotion. As suggested, conceptual short-hand in research design has sometimes clouded this issue by not sufficiently discriminating between the objectively presented (trigger) *stimuli* (i.e., a picture, a loud noise, a word, or a phrase) versus the participants’ subjective reaction to the stimuli (i.e., feeling calm, feeling frightened, or feeling sad). It follows that the unfortunate assumption made by some research designs is that the presented stimuli is de facto of, “a calming picture” or “a frightening noise,” where object of emotion and experience of emotion come to inadvertently overlap. This is important because the subjective relevance is confounded with the actual stimulus.

Thus, while experimental research, for example on perception or basic components of emotion, is often studying participants’ reactions to a given stimuli (e.g., what happens when the participant suddenly hears a loud crash?), research on emotional experience is (or should

be) studying the *response to an emotion that is being evoked*. However, even as it is the target of inquiry, this latter issue (the second step in the phenomenon; [10]) is all too often insufficiently distinguished from the actual stimulus used to evoke or elicit feeling. Still, it is this covert internal and subjective emotional reaction which the researcher intends to manipulate, by proxy, through the external manipulation of stimuli. Nevertheless, the researcher cannot directly manipulate emotional experience as one might otherwise manipulate variables in conventional experimental designs on more explicit phenomena.

As a case in point, in a study on the impact of “fearful stimuli” on cognitive processes a researcher might present alarming or dangerous stimuli in an effort to evoke fear in participants. In a conventional design that follows good research practice, the presented stimulus will be the same across participants in an effort to standardize research procedures, and this is both expected and valued by reviewers of research. However, despite the best of intentions, two assumptions are often inadvertently made in these research designs. We will examine them below.

The first assumption is that the experimental stimulus being used will be evocative, in particular it will be evocative of a *specific* target emotion (i.e., fear), and that the emotional experience being studied will be comparable across participants. Of course, this initial issue is often (but not always) controlled for to some degree in designs that use a manipulation check. Still, the tacit assumption has already been made, such that a manipulation check often only assesses the degree to which some target emotion was evoked, and (as if often the case in hypothesis testing) whether there were different or more complex reactions is a question that remains unasked. Thus, a considerable margin of variability is usually accepted because the notion of a manipulation check, which is secondary and post hoc to the actual stimulus, is usually satisfied with minimal data. For example, if the stimulus is presumed to evoke fear – so according to earlier research has demonstrated relative fear specificity –, then emotional

arousal following presentation of that stimulus is taken as evidence to indicate that the induction procedure was effective. This evidence of a successful manipulation is often a rating of arousal, negative affect, or subjective units of distress - perhaps reported on a scale of 1 to 10. However, while a manipulation check confirms that induction of emotional arousal was successful, it often takes for granted that the *quality* of emotion, was the same as prescribed by the experimental design. Moreover, much research, particularly studies conducted in a clinical context, presumes that this quality of emotion (e.g., “fear”) is uniform across participants when in fact it may vary: e.g., fear vs. vague distress vs. shame-anxiety vs. helplessness. In short, the common logical fallacy here is as follows: a researcher uses stimuli to induce fear, then participant reports are used to confirm some change in arousal, and therefore “the manipulation of participants to experience fear” is taken to have been successful – and the researcher eagerly moves on to examine findings. These short-cuts are particularly tenuous when examining emotions in patients with specific disorders of the emotion recognition system, as found, for example, in borderline personality disorder [20]. Evidence suggests that even healthy individuals tend to misattribute certain emotions, such as disgust [21]. Demonstrated relative fear specificity and manipulation checks of arousal do not guarantee that fear was really activated.

The second assumption being made in this common type of design, however, is more serious to its internal validity. The manipulation of a presented stimulus (the first phase described by Sander [10]), is functionally equated with the manipulation of a target emotion (the second phase). Despite the fact that they are often carefully standardized, the stimuli being presented, is *only a proxy* to emotion (i.e., the target variable under inquiry). This means that standardizing stimuli, which may give the outward appearance of standardization and internal validity in a research design, may actually ignore at best, or undermine at worst, the standardization of the actual target variable one presumes to be manipulating (i.e., a given

emotional experience). For patients with borderline personality disorder, for example, it has been shown that idiosyncratic stimuli elicit greater levels of specific emotion as compared to standardized stimuli [22]. The process of stimulus standardization may therefore create additional problems in the measurement of the targeted emotion.

This observation presents a fundamental problem, which we refer to as the "emotion-stimulus critique." Some of these problems, and how they may be solved, will be discussed below by using exemplary studies from the literature.

Arousal, pain, and more complex affective-meaning states

The emotion-stimulus critique has largely gone unnoticed partly because it has already been adequately addressed when it comes to the single component of emotional arousal, which is linear, meaning that arousal is usually measured as higher or lower. A simple and reliable way of managing this issue is to ask participants to essentially rate their in-the-moment emotional arousal on some Likert scale. In neuroscientific research (e.g., [5]), as well as psychological research on affective arousal in general, a widely used example of this is the subjective rating of general distress made often from 0 to 10, or similar ratings on the Self Assessment Manikin, a self-report measure in which participants are presented with pictographs arranged along a nine-point continuum [23]. In this approach, a given rating from one participant does not necessarily correspond to the same "objective" arousal rating given by some other participant. Indeed, the measure relies on a person's *inner metric*, which is to say that arousal is measured from an individual's subjective perspective [23]. Even when the stimuli used to elicit arousal are quite different, the actual emotional response (on a single dimension) may be comparable across participants.

The emotion-stimulus critique is also sufficiently addressed in domains adjacent to emotion, such as in pain research. Pain, like emotion, produces a physiological response with a subjective evaluation of its intensity; therefore, this example seems relevant. In a recently

developed paradigm measuring pain perception in patients presenting with borderline personality disorder (BPD), Niedtfeld et al. [24] were mindful of the impact of individual differences in the tolerance of pain experience and administered individualized intensities of pain stimuli, according to a subjective pre-assessment by the participants of a given temperature as feeling “moderately painful.” In a pre-test, the participants had to respond to the question of what stimulus intensity was experienced as e.g., a “6 on a scale ranging from 0 to 10.” Again, whereas the actual temperature (presenting stimulus) could objectively vary, the subjectively felt experience of pain (subjective experience) was reportedly the same across patients (“moderately painful” is a 6 out of 10). As it happens, this objective variation could even be measured using standard deviation: for patients with BPD, moderately painful temperature was $M = 47.4$ degrees Celsius, $SD = 0.79$. Nonetheless, quantitatively different stimuli beget qualitatively similar responses. Hence, even though objective intensity of the stimulus varied, the painful experience was the same.

Beyond clinical psychology, a sophisticated approach on measuring the intensity of sensations (i.e., related to the taste of food) is represented by the Labeled Magnitude Scaling (LMS[25]). It aims at the representation of the “inner metric” by transforming rated intensity using log-linear computations of the raw scores. It was shown that these LMS ratings correlate highly, but not perfectly, with the traditional measurements (i.e., Likert-type scales) of the intensity of the sensations [26]. This results demonstrates that Likert-type scales might be sufficiently adequate for measuring linear concepts, such as intensity of arousal.

The field of research on “feelings” has progressively moved from designing studies that examine questions related to arousal, or the experience of pain (both usually treated as linear concepts), towards the examination of qualitatively distinct emotional states (e.g., anger, shame, see [8]). However, the study of elaborated emotional states demands increasingly complex designs. In a recent study on emotional responses that used

physiological and self-report measures, Elices et al. [27] used a validated set of film-stimuli [13] and presented them to 30 patients with borderline personality disorder (BPD) and to 30 controls. Based on such preliminary work and the well established stimuli, the researchers' core assumption was that certain films excerpts produced discrete emotional reactions. For example, a two minute excerpt from the movie "Man Bites Dog," where the main character threatens an old lady with a gun and she dies, is intended to evoke anger. It is hypothesized that participants with BPD would experience a wider range of emotions and with greater intensity than controls. Such general links between stimuli and emotional responses were shown in previous research using this particular set of stimuli [13], but we would argue that such a design seems to overlook the experiential anchoring ("inner metric") within the individual in regards to their subjective emotional response. A number of participants may indeed experience anger when viewing the threatening scene described, perhaps even a majority as shown by Schaefer et al. [13]. Even so, it is possible to have other spontaneous emotional reactions to the same complex stimulus, such as fear, sadness, or much more complex and mixed emotions. Also, for certain individuals, anger might indeed be the dominant emotion (i.e., rated by the participant as being their most intense reaction), but other emotions might also be present. Conversely, another participant might experience anger as a less intense emotion, in relative terms, and fear as the predominant emotion. In short, the question of how a group of people respond to a stimulus on average is different from the question of how each member in one's sample has responded. Overlooking the experiential anchoring in the participants in this design means that there may be unacknowledged limitations to the conclusions drawn. Authors of the study [27] did not find any significant differences in the self-reports of categorical emotions between BPD and controls, nor did they find a correlation between physiological measure of arousal and its self-reported counterpart. In considering these findings, one might speculate that the absence of significant findings may

be attributed to the method and explained by the emotion-stimulus critique. By assuming that so-called “anger-eliciting” excerpts really evokes anger in participants, potential between-group differences may have been washed out by the grouping together of the real variation in the subjective experiences among participants.

In addition, as we have discussed, once stimuli are held constant, inter-participant comparability is usually not performed on the level of the participant's internal emotional response. This problem seems all too often to be overlooked because a series of tacit assumptions are made with regard to the “universal” links between an evocative stimulus and the emotional response of participants. We think it is useful to untangle these and argue that an optimal research design should purposefully include idiosyncratic stimuli. In short, studies that include the participant's "own words," or other personalized and experience-evoking stimuli, or by using individualized script-driven imagery are highly desirable [22]. The aim of this would be to maximize personal and *subjective relevance* while at the same time standardize as best as possible the ensuing emotional responses (Sander's second phase of emotion) to presenting stimuli against an inner metric. This would maximize between-participant comparability on emotion (the target variable), rather than maximizing comparability on the intended evocative stimuli (the proxy variable). Experimental tasks that do this would shore up against the emotion-stimulus critique.

Guidelines for resolving the emotion-stimulus critique: Potential solutions

The objective of experimental studies on emotion is generally to gain insight into a participant's reaction to an evoked emotion and in order to do this, the presenting stimuli are often held constant, assuming they evoke comparable emotional reactions across individuals. However, the "true" variable of interest which is the one ultimately targeted for manipulation is a participant's actual emotional response, evoked by some stimulus. Consequentially, a number of research designs, produce ambiguous data unless a participant's actual emotional

experience is adequately appraised and controlled for. This can be corrected if research designs meet as many of the following guidelines as possible:

(1) *A broad category of idiosyncratic stimuli are selected*, where all items target one specific emotional process in a comparable way. However, while stimuli may not be uniform or held constant across participants (as when presenting participants with a standard face of e.g., an angry man), they are within some range or within a given category, where the category of stimuli will have as many degrees of freedom as reasonably possible (e.g., presenting each participants with the angry face of his or her care givers). In this case and in particular for group comparisons, problems with stimuli familiarity need to be attended to. A compelling example of this is the study by Hooley, Siegle and Gruber [28] on perceived self criticism, in which the researchers presented participants with audio recorded critical statements from their own caregivers. It seems important to consider that the selection of idiosyncratic stimuli (i.e., the participant's own words, or the picture of one's own caregiver) should be done based on some preliminary discovery-oriented testing or psychological assessment. For Hooley and colleagues, this required telephone interviews with participants' mothers. In other types of examples, the preliminary strategy for gathering source material might be to ask participants for narratives on auto-biographical memories [29], or using an imaginary self-critical dialogue from emotion focused psychotherapy [30].

Using such material where a target emotional experience, for example shame, is directly evoked or referred to by each participant, the researcher must then develop a word extraction procedure (if words/phrases are the intended stimuli). In this procedure, the most poignant words related to these, say, shameful experiences would be identified for each participant. (If images or other stimuli are being used, an analogous extraction, selection, or cropping procedure should be used.) These materials will subsequently serve as the idiosyncratic and personally relevant stimuli for when one manipulates participant's emotional

experience (i.e., during the actual experimental task in an fMRI environment). This preliminary work can be done directly by the researcher or alternatively in collaboration with the participants themselves through a structured task. For example, participants could be invited to choose a stimulus from a large set of possibilities (i.e., the internet, a test bank) an image, film clip, or text that evokes the target emotion (e.g., shame) in him or her. Again, this will become the idiosyncratic stimulus for a later experiment on emotional experience or emotional processing.

(2) *Control the context of the stimulus* as much as possible, even though individual stimuli will be different. This context will help to increase the comparability and increase the likelihood that different stimuli are interpreted similarly across participants. For example, the context of the stimulus is controlled for when choosing a moderately evocative context, as arousal should ideally be constant across specific stimuli. Examples of this are “emotional experiences related with traumatic memories”, “unresolved interpersonal problems”, “current self-criticisms,” or by framing other target processes identified with specific disorders (i.e. example anxious rumination in generalized anxiety disorder). These general contexts for study should have some face validity as being evocative in nature. For instance, in collecting critical stimuli pre-testing, Hooley et al. [28] asked participants’ care givers to provide critical remarks in a standard way (e.g., “Stacey, one thing that really bothers me about you is...”; p. 3). Here, the caregiver’s moderately arousing criticism is the stable context of the stimulus, whereas the formulations of these stimuli are idiosyncratic to each participant in the study. We should note that the cultural context, among other variables, may impact his variability. Also, it needs to be noted that the actual presentation of the care giver might confound with effects of familiarity facing such a picture, which should be controlled for or circumvented (i.e., by preferring verbal stimuli for example).

(3) *Stimuli should have high face validity from each participant's personal perspective.* Researchers and lay persons alike may not find immediate face validity in the selected stimuli without knowing the personal associations and meaning the participant attributes to the chosen stimuli. In some sense, there may be “a story” related to why a participant’s stimulus will be evocative for him or her and this may not be self-evident. In a case example from unpublished research, one of our participant's most poignant and contemptuous self-criticism was the accusation: “You don’t even wear a tie!” The painfully contemptuous meaning behind this sentence (e.g., personal shame about his lower social class, real concerns about being employable, not living up to family standards) was immediately recognizable to the participant himself (i.e., high face validity), even if the comment about ones dress seems relatively benign to an outsider. Note that for this specific example, it was sufficient to rely on the participant’s self-report, but there may be cases where researchers need to use other methods (see #5, below). In another study by Doerig et al. [31] on self criticism, healthy participants were instructed in a pre-fMRI test to pick six “prototypes” from a list of self-critical adjectives that most personally related to them. In order to assess the subjective relevance of each of these, participants were then asked to rate them on a Likert-type scale, answering the question: “How negatively would you rate this characteristic of yourself?”

(4) Given that idiosyncratic and personalized stimuli vary, but are selected based on the criterion that they evoke a given emotional experience, there needs to be an internal check that these *stimuli are functionally equivalent between participants* within the design, despite their being uniquely customized. This standard is to establish that the stimuli really evoke what they should be eliciting for each individual participant, enabling the design to control for their subjective relevance. This assessment can often be easily taken from participant's subjective reports. To do this a researcher must consider: Does the presenting stimulus (the

proxy variable) elicit in the client a state that is consistent with the emotional experience being studied (target variable)? Thus, although stimuli across participants may differ, the pattern of participant responding is considered comparable. For example, there needs to be a check if all extracted words (or chosen images, films, etc.) provoke the actual target emotion with sufficient specificity and to a sufficient degree. Self-reported measures assessing distinct emotion categories are possible tools for this, as they often give a range of emotions on which participants rate the intensity and qualities of their immediate experience (allowing researchers to consider the profile of emotional experience). Examples of this include the self-reported Differential Emotions Scale assessing distinct emotion categories [32], or the Positive and Negative Affect Rating Scale [33].

Finally, (5) *observation-based classifications of participant's feeling and meaning* are also sometimes necessary. Indeed, in some types of research, data collection using participant's self-rating, as suggested in guideline #4, is not available or may not be feasible. Alternatively, in specific cases, self-report of emotions may not be considered as sufficiently reliable information. This reliability problem is sometimes more acute when studying clinical population, for example, as already cited in borderline personality disorder [20].

Emotion is a rapid and ephemeral phenomenon of meaning construction and it progressively unfolds in a multi-step sequence. When researchers wish to study these patterns of emotion, interrupting the process or flow of a participant's emotional experience would directly interfere with the phenomenon of interest. For example, in a quasi-experimental design, Whelton and Greenberg [30] asked participants *first* to imagine speaking critically to themselves and *then*, to immediately respond to their own criticisms. This procedure allowed the researchers to examine participants' emotional resilience in vivo, moment-by-moment. Similarly, Rochman and Diamond [34] used an experiment to test whether specific emotional processing sequences were related to overcoming anger about interpersonal grievances. By

randomly assigning participants to conditions where they expressed either sadness before anger, or anger before sadness, researchers showed the latter sequence was accompanied by greater indices of productive change and resilience. Studies like these, on moment-by-moment resilience and change sequences, have strong clinical implications but they also require that emotion unfolds in a continuous two (or more) step process. Yet, for these designs to be internally valid, researchers must also confirm that each of targeted emotional states (e.g., contempt followed by assertion; or anger followed by sadness) were indeed successful as experimental manipulations. Moreover, this assessment must be done for each individual participant in vivo, which precludes self reports, as described in guideline #4. For this reason, observational criteria must be used, often by blind raters who examine archival data, i.e., video, transcripts of the process.

Still other research questions exploring emotional experiences may examine emotion that is somewhat outside of the participant's own awareness, either non-conscious, or sub-conscious, or where participants are immersed in the moment of affective experience (i.e., incidental emotion without deliberate self-awareness). In studies of this kind, even if participants could feasibly be asked directly about their emotional states, they may not be able to respond accurately or, in other cases, drawing attention to the measurement of emotion per se risks precipitating some artefactual change in the target state. For example, in a study on trauma and expressive writing, Pascual-Leone et al. [35] used priming tasks to target specific emotions in experimental conditions and then asked participants to write about their "deepest thoughts and feelings" related to a personal trauma. Because emotional engagement and writing styles are known to vary dramatically in this paradigm, and participants may or may not have responded to the emotional-priming, the content of participants' narratives needed to be directly examined to confirm that different conditions were indeed successful.

In short, research that examines complex dynamic processes, or emotional processes that may be outside of awareness will often require procedures in the design to ensure that target emotion(s) and engagement are comparable across participants within a given condition. Researchers can verify the target emotion(s) using blind, third-party observer's ratings based on objective criteria to capture participants "emotional performance" (i.e., does the elicited emotional reaction correspond to criteria of a particular emotion category?). When data is based on video recordings, one example of this is to rate participants' responses on a behaviorally anchored observational criteria (e.g., as done in [30] or [34]) or a more comprehensive measure such as the Classification of Affective Meaning States (e.g., [6,35]). For very specific queries about emotion, such as subtle expressions of contempt, one might assess facial muscular activity using electromyographic measures (e.g., [36]). Still other forms of "observational" measures might include systematic linguistic analyses. Tausczik and Pennebaker [37] have shown that computer software examining both style words (e.g. pronouns, articles, prepositions) and content words (e.g., nouns, adjectives, adverbs) can be used to capture the "footprint" of a participant's affective and cognitive experience.

Conclusions

Experimental research on emotion, in particular the dynamic study of complex emotional experience and responding, has moved from the study of intensity of emotions towards questions about distinct categories of emotional experience. What is too often overlooked in such studies is the hidden problem of a stimulus' idiosyncratic emotional relevance for an individual and the diversity of the actual participants' experience as they are presented with "standardized" research stimuli. Indeed, standardizing the presented stimuli that evoke emotion is not the same as standardizing the emotion they in turn evoke, which more often than not is the actual aim of inquiry. This conceptual short hand has led some research astray from optimal designs and vulnerable to what we call the *emotion stimulus*

critique. It is not sufficient to show that a certain percentage of participants respond with a certain category of emotion when faced with a standardized stimulus. Research designs should adopt a *meta*-subjective perspective, considering the significance of a given experimental tasks from within participant's subjective perspectives [15]. In this paper we have offered some initial guidelines to help address these design problems, although we expect recommendations will continue to evolve through an ongoing discussion on the methods used to experimentally study emotion, in particular in the context of clinical populations. It is time for experimental research on emotion to reintroduce the participant's experience back into the laboratory and to attend to the *participant's* actual processes.

References

- [1] Lane RD, Ryan L, Nadel L, Greenberg L: Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from Brain Science. *Beh Brain Sci*, in press.
- [2] Pascual-Leone A, Paivio S, Harrington S: Emotion in psychotherapy: An experiential-humanistic perspective. In D Cain, S Rubin, K Keenan (eds.) *Humanistic Psychotherapies: Handbook of Research and Practice* (Second edition), in press.
- [3] Rachman R: Emotional processing. *Beh Res Ther* 1980;18:51-60.
- [4] Neacsiu AD, Rizvi SL, Linehan MM: Dialectical behavior therapy skills use as a mediator and outcome of treatment for borderline personality disorder. *Beh Res Ther* 2010;48:832-839.
- [5] Schnell K, Herpertz SC: Effects of dialectical-behavioral-therapy on the neural correlates of affective hyperarousal in borderline personality disorder. *J Psychiatr Res* 2007;41:837-847.
- [6] Pascual-Leone A: Dynamic emotional processing in experiential therapy: Two steps forward, one step back. *J Consult Clin Psychol* 2009;77(1):113-126.

- [7] Feist G, Rosenberg E, Stamp J, Poole J: Psychology: Making connections (Canadian Edition). Whitby, McGraw-Hill Ryerson Higher Education, 2012.
- [8] Lewis M, Haviland-Jones JM, Feldman Barrett L: Handbook of Emotions. New York, NY, Guilford Press, 2010.
- [9] Siegle GJ, Thompson WK, Collier A, Berman SR, Feldmiller J, Thase ME, Friedman ES: Toward clinically meaningful neuroimaging in depression treatment. Arch Gen Psychiatr 2012;69(9):913-924.
- [10] Sander D: Models of emotion: The affective neuroscience approach. In J Armony P Vuilleumier (eds.): The Cambridge Handbook of Human Affective Neuroscience. Cambridge, Cambridge University Press, 2013, pp 5-56.
- [11] Banich MT, Compton RJ: Cognitive Neuroscience. Third Edition. Wadsworth, CA, Cengage Learning, 2011.
- [12] Bradley MM, Lang PJ: Affective Norms for English Words ANEW Technical Manual and Affective Ratings. Gainesville FL, The Center for Research in Psychophysiology, University of Florida, 1997
- [13] Schaefer A, Nils F, Sanchez X, Philippot P: Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cogn Emot 2010;24(7):1153-1172.
- [14] Tolman EC, Brunswik E: The organism and the causal texture of the environment. Psychol Rev 1935;42:43-77.
- [15] Pascual-Leone J: Can we model organismic causes of working memory, efficiency and fluid intelligence? A meta-subjective perspective. Intell 2013;41:738-743.
- [16] Watson JB, Rayner R; Conditioned emotional reactions. J Exper Psychol 1920;3:1-4.
- [17] Gendlin ET: A theory of personality change. In P Worchel D Byrne (eds.): Personality Change. New York, Wiley & Sons, 1964.

- [18] Goldfried MR, Davidson GC: Clinical behavior therapy (expanded ed.). New York, Wiley-Interscience, 1994.
- [19] Cunningham WA, Brosch T: Motivational salience: Amygdala tuning from traits, needs, values and goals. *Curr Dir Psychol Sci* 2012;21(1):54-59.
- [20] Herpertz SC: Affektregulation und ihre neurobiologischen Grundlagen [Affect regulation and its neurobiological underpinnings]. In B Dulz SC Herpertz U Sachsse (eds.): *Handbuch der Borderline-Störungen (2. Auflage)*. Stuttgart: Schattauer, 2011, pp 75-85.
- [21] Nabi RL: The theoretical vs the lay meaning of disgust: Implications for emotion research. *Cogn Emot* 2002;16(5):695-703.
- [22] Kuo JR, Neacsiu AD, Fitzpatrick S, MacDonald DE: A methodological examination of emotion inductions in borderline personality disorder: A comparison of standardized versus idiographic stimuli. *J Psychopath Behav Assess* 2014;36:155-164.
- [23] Bradley MM, Lang PJ: Measuring emotion: the self-assessment Manikin and the semantic differential. *J Beh Ther Exper Psychiatr* 1994;25:49-59.
- [24] Niedtfeld I, Kirsch P, Schulze L, Herpertz SC, Bohus M, Schmahl C: Functional connectivity of pain-mediated affect regulation in borderline personality disorder. *PLoS One* 2012;7(3):e33293. doi: 10.1371/journal.pone.0033293
- [25] Green BG, Shaffer, GS, Gilmore, MM: Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chem Sens* 1993; 18:683-702.
- [26] Lawless, HT, Popper, R, Kroll, BJ: Comparison of the labeled affective magnitude (LAM) scale, an 11-point category scale and the traditional nine-point hedonic scale. *Food Qual Pref* 2010; 21:2-12.

- [27] Elices M, Soler J, Fernandez C, Martin-Blanco A, Portella MJ, Perez V, et al.: Physiological and self-assessed emotional responses to emotional-eliciting films in borderline personality disorder. *Psychiatr Res* 2012;200:437-443.
- [28] Hooley JM, Siegle G, Gruber SA: Affective and neural reactivity to criticism in individuals high and low on perceived criticism. *PloS One* 2012;7(9):e44412.
- [29] Beblo T, Driessen M, Mertens M, Wingenfeld K, Piefke M, Rullkoetter N, et al.: Functional MRI correlates of the recall of unresolved life events in borderline personality disorder. *Psychol Med* 2006;36(6):845-856.
- [30] Whelton WJ, Greenberg LS: Emotion in self-criticism. *Pers Individ Diff* 2005;38:1583-1595.
- [31] Doerig N, Schlumpf Y, Spinell S, Späti J, Brakowski J, Quednow BB, et al.: Neural representation and clinically relevant moderators of individualised self-criticism in healthy subjects. *Soc Cog Affect Neurosci* 2013;1-8;doi: 10.1093/scan/nst123.
- [32] Youngstrom EA, Green KW: Reliability generalization of self-report of emotions when using the Differential Emotions Scale. *Educ Psychol Measur* 2003;63:279-295.
- [33] Watson D, Clark LA, Tellegen A: Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 1988;54:1063-1070.
- [34] Rochman D, Diamond GM: From unresolved anger to sadness: Identifying physiological correlates. *J Counsel Psychol* 2008;55(1):96-105.
- [35] Pascual-Leone A, Sirois F, Metler S, Singh T, Crozier M, Porter L: Productive emotion in expressive writing: An experimental manipulation of EFT change. Presentation at the 42nd International Meeting of the Society for Psychotherapy Research. Bern, June, 2011.
- [36] Kamen G, Gabriel DA: *Essentials of Electromyography*. Champaign, IL, Human Kinetics, 2010.
- [37] Tausczik YR, Pennebaker JW: The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2010;29(1):24-54.