Psychophysiological modeling - Current state and future directions

Dominik R. Bach[1,2,3], Giuseppe Castegnetti[1,2], Christoph W. Korn[1,2,4], Samuel Gerster[1,2], Filip Melinscak[1,2], Tobias Moser[1,2]

[1]Clinical Psychiatry Research, Department of Psychiatry, Psychotherapy, and Psychosomatics, Psychiatric Hospital, University of Zurich, Zurich, Switzerland
[2]Neuroscience Center Zurich, University of Zurich, Zurich, Switzerland
[3]Wellcome Trust Centre for Neuroimaging and Max Planck/UCL Centre for Computational Psychiatry and Ageing Research, University College London, United Kingdom
[4]Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Correspondence
Dominik R. Bach, Department of Psychiatry, Psychotherapy, and Psychosomatics, Psychiatric Hospital, University of Zurich, Lenggstrasse 31, CH-8032 Zurich; Switzerland. Email: dominik.bach@uzh.ch

Abstract

Psychologists often use peripheral physiological measures to infer a psychological variable. It is desirable to make this inverse inference in the most precise way, ideally standardized across research laboratories. In recent years, psychophysiological modeling has emerged as a method that rests on statistical techniques to invert mathematically formulated forward models (psychophysiological models, PsPMs). These PsPMs are based on psychophysiological knowledge and optimized with respect to the precision of the inference. Building on established experimental manipulations, known to create different values of a psychological variable, they can be benchmarked in terms of their sensitivity (e.g., effect size) to recover these values we have termed this predictive validity. In this review, we introduce the problem of inverse inference and psychophysiological modelling as a solution. We present background and application for all peripheral measures for which PsPMs have been developed: skin conductance, heart period, respiratory measures, pupil size, and startle eye blink. Many of these PsPMs are task invariant, implemented in open-source software, and can be used off the shelf for a

wide range of experiments. Psychophysiological modeling thus appears as a potentially powerful method to infer psychological variables.

Keywords

analysis/statistical methods, autonomic nervous system, computational modeling, electrodermal activity, heart rate, pupillometry

1 Introduction

Peripheral physiological measurements are often used to infer psychological variables. For example, to test a psychological intervention that reduces fear memory, a researcher may be interested in quantifying the strength of fear memory from skin conductance responses (SCR). This psychological perspective is the opposite of a basic psychophysiological perspective, in which researchers aim to describe how psychological variables impact on the peripheral measure. To enable inference on a psychological variable, the psychophysiological mapping from this variable to the measured signal (the "forward model" in statistical terminology) needs to be known with some certainty, and it needs to be exploited in the best possible way (the model needs to be "inverted"), to arrive at the most precise estimate of the psychological variable. This is the psychophysiological inverse problem (**Figure 1a**).
Psychophysiological modeling is a statistical framework to solve this problem in a principled manner (Bach & Friston, 2013). It can provide experiment-invariant, off-the-shelf applications that improve on current methods for inverse inference and thereby suggest meaningful methodological standards to enhance reproducibility.

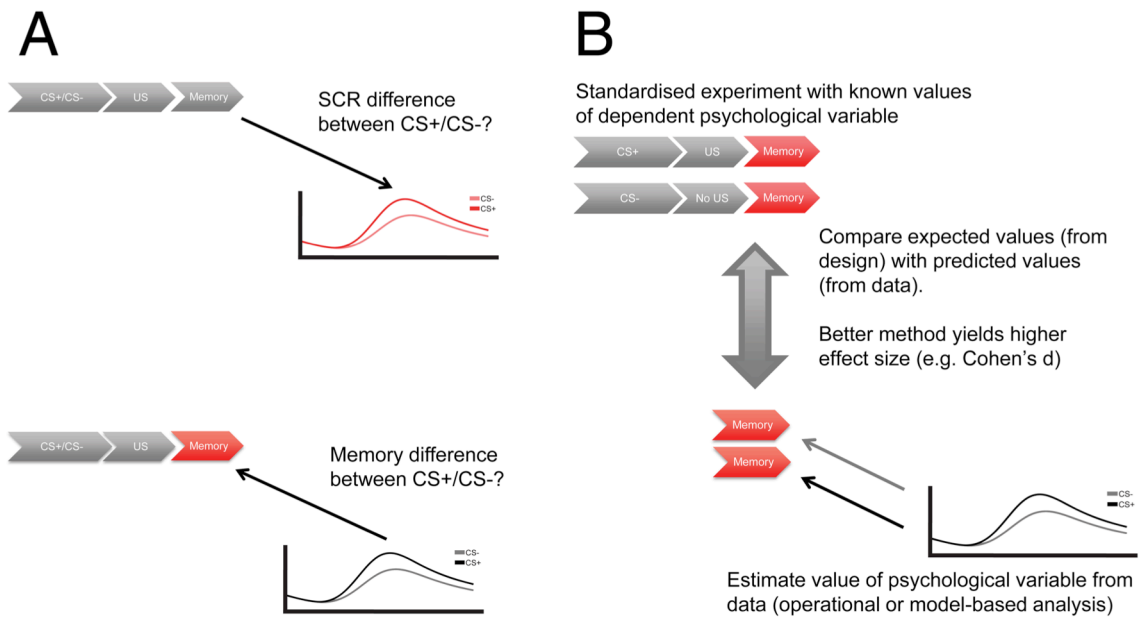

Figure 1. A: The psychophysiological inverse problem. Top: psychophysiological perspective (forward inference, e.g., Does aversive memory influence SCR?). Bottom: psychological perspective (inverse inference, e.g., Does my procedure establish aversive memory, as indexed by SCR?) B: Benchmarking an inverse inference method by assessing predictive validity: What is the sensitivity for inferring the effect of a known experimental manipulation.

Indeed, psychophysiological modeling approaches have been applied to analyze a wide variety of experiments: to infer attentional variables from pupil responses (e.g., de Gee et al., 2017; de Gee, Knapen, & Donner, 2014), to infer fear learning from SCR (e.g., Bach, Weiskopf, & Dolan, 2011; Bulganin, Bach, & Wittmann, 2014; Tzovara, Korn, & Bach, 2018) and startle eyeblink (Bach, Tzovara, & Vunder, 2017), and to quantify autonomic arousal during perception (e.g., Bach, Seifritz, & Dolan, 2015; Hayes et al., 2013; Koban, Kusko, & Wager, 2018; Koban & Wager, 2016; Sulzer et al., 2013), decision-making (e.g., Alvarez, et al., 2015; Bach, 2015a; de Berker et al., 2016; Nicolle, Fleming, Bach, Driver, & Dolan, 2011; Talmi, Dayan, Kiebel, Frith, & Dolan, 2009), and rest (Fan et al., 2012).

This review is structured in the following way. First, we discuss how to compare methods for inverse inference on psychological variables and introduce the concept of predictive validity. We then present psychophysiological modeling as a novel approach, including a specific implementation created by the authors together with related methods. In the major part of the review, we give a tutorial-style overview of the various forward models and inversion methods developed over the past decade, for different physiological measures. The field is moving rapidly. While nine methodological articles

on the topic were published between 1993 - 2013, 11 such papers came out in the 5 years since the last review on the topic (Bach & Friston, 2013). This last review contained a historical perspective on the development of models for SCR in the 1990s and 2000s (Alexander et al., 2005; Bach, Flandin, Friston, & Dolan, 2009; Barry, Feldmann, Gordon, Cocker, & Rennie, 1993; Lim et al., 1997) and on the emergent critique of operationalism (Green, 1992); here, we approach the problem in a systematic, nonhistorical manner.

2 Predictive validity

All analysis methods for psychophysiological signals are based on some knowledge about the forward mapping from psychology to physiology. A plethora of psychophysiological literature has addressed such forward mappings. However, even with a perfect forward model there are different ways of making inverse inference. For example, one can define different possible time windows to detect an SCR peak after an experimental event. Extending the peak window may increase the sensitivity of the method to detect a true event-related response, but decrease its specificity because experiment-unspecific peaks may be mistaken for event-related ones. Crucially, the optimal balance is difficult to intuit as is, for example, evident from the coexistence of different peak detection windows in analysis recommendations (Boucsein et al., 2012), and sometimes even within the same laboratories. Hence, it would be desirable to quantitatively evaluate an inverse inference method.

To assess the quality of inverse inference, one would ideally compare the inferred with the actual value of the psychological variable (i.e., with "ground truth"). Of course, ground truth is never known for psychological variables.[1] To solve this conundrum, we have pragmatically proposed to use an experimental manipulation that can be assumed to influence the psychological variable in a certain way and is known to impact on the peripheral measure. One can then evaluate methods by their sensitivity to detect the impact of this experimental manipulation (**Figure 1b**). We have introduced the term predictive validity for this measure (Bach, Daunizeau, Friston, & Dolan, 2010), since it

[1] This is true for many areas of science and technology.

evaluates how well the psychological variable can be predicted.[2] Predictive validity analysis has often been performed on a categorical experimental manipulation (e.g., anticipating threat or safety in fear conditioning). Although it may thus appear on the surface that predictive validity boils down to classification performance, one usually aspires to infer psychological variables on a continuous scale, so that the method extends to situations in which the psychological variable varies parametrically across more than two levels. Crucially, since most psychophysiological measures are relatively unspecific, validation experiments require that the experimental conditions differ on only one dimension, the psychological variable of interest. This is true for any inverse inference method. Once a method with high predictive validity is identified, one can apply this method to other (methodologically similar) experiments to infer the same psychological variable.

Thus, in a validation experiment, a good inference method provides an estimator of the (known) psychological variable that has smaller variance than any other method. For a categorical validation experiment with two levels, this simply means - because the scale of the psychological variable is arbitrary - that the standardized difference in the estimated psychological variable between these two levels should be large. This can be evaluated by regarding the effect size, or test statistic, or the residual sums of squares in a predictive model, or the model evidence of that predictive model. All of these quantities are monotonically related. Using model evidence additionally allows us to make statements whether two methods are decisively different (Bach & Friston, 2013; see Appendix Equation 1). At the same time, the difference in the estimated psychological variable between two random partitions of the same experimental condition should be zero on average.

Predictive validity can be harnessed to validate any inverse inference method, including operational analysis. For the psychophysiological modeling approach, some additional considerations are warranted. Here, psychological variables are estimated by optimizing the goodness-of-fit of the forward model. Yet, for comparison of different methods, the goodness-of-fit of the forward model is not a suitable criterion. The goal of the forward

[2] Because the psychological variable is known a priori, one could also call it “retrodictive validity”.

model is to predict the signal, and the goal of inference is to find the most precise estimate of the psychological variable. These two goals can align with, be orthogonal to, or even oppose each other. Intuitively, one could assume that, if the forward model is known with certainty and formulated in mathematical terms, one should easily be able to invert the mapping. However, there are several statistical reasons why this intuition is incorrect in the general case (although it may be correct under specific circumstances). First, the forward model may use parameters that one is not interested in inferring. For example, the best known forward model for SCR assumes that the strength or amplitude of psychological input into the system is different on each trial (Gerster, Namer, Elam, & Bach, 2017). However, many researchers are not interested in psychological variables on a trial-by-trial basis but only in the average psychological variable within an experimental condition. The standard general linear model (GLM) inversion approach for SCR (Bach et al. 2009; Bach, Friston, & Dolan, 2013) will normally yield the same conditionwise estimates, regardless of whether estimation was done on a trial-by-trial basis followed by averaging, or on a condition-by-condition basis. In this case, the simpler model yields the same inference on the psychological state, although empirically it cannot predict SCR data so well, because it would assume the same (average) SCR amplitude for each trial (Bach et al., 2013). Hence, precision of the forward model and of the inference are unrelated. An example where they are opposed is given by individual response functions for SCR. All evidence suggests that the mapping from sudomotor nerve activity to skin conductance depends on subject-specific anatomical properties, and is variable between persons (Bach, Flandin, Friston, & Dolan, 2010; Gerster et al., , 2017). Hence, a forward model taking this heterogeneity into account will have a better goodness-of-fit than a model assuming a canonical response function across subjects, as we have also shown empirically (Bach et al., , 2013). At the same time, it can be difficult to estimate the shape of an individual's true response function from a limited number of trials with short inter trial intervals, and ensuing "overfitting" can make inference on the psychological variable worse, reducing predictive validity (Bach, Friston, & Dolan, 2013).

To summarize, predictive validity allows a statement on the quality of inverse inference, regardless of the method under study. It can be used to benchmark psychophysiological models, operational methods, or even machine-learning methods that try to find statistical regularities without any knowledge of the psychological or biophysical

relationships (Greco et al., 2017; Greco, Lanata, Valenza, Di Francesco, & Scilingo, 2016; Greco, Valenza, & Scilingo, 2016). As a statistical framework, it has a potential to improve inverse inference, to standardize methods across laboratories (by selecting the best one), and to provide an objective means for quality control within and between laboratories. We will return to these latter points in the discussion.

3 Psychophysiological Modeling

The goal of inverse inference is to find the best possible estimator of a psychological variable, from a measured data time series. This includes so-called operational methods, which "operationalize" (i.e., equate a noisy version of) the psychological variable with a single physiological data feature, for example, a peak-to-trough measure. Because operational methods use one or a very small number of data features, rather than the entire time series, they may suffer from information loss. Psychophysiological modeling is a way of using an entire data time series for inference (**Figure 2**). In a nutshell, a

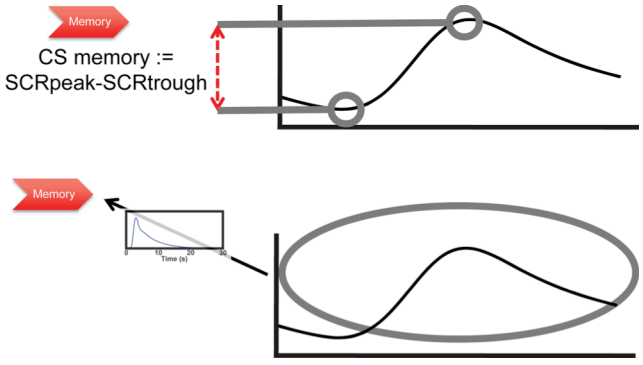

Figure 2. Operational analysis (top) assumes that selected data features are "equivalent" to a psychological variable, where the selection of data features is often based on informal models. Psychophysiological modelling (bottom) estimates the most likely psychological variable, given the entire data time series and a standard (experiment-invariant) response model.

psychophysiological model (PsPM) is a formal, quantitative model that maps a psychological variable onto an observed data time series. PsPMs are specified in mathematical form. The earliest PsPMs were developed for SCR and explicitly constituted a sequence of two models (**Figure 3a**): a neural model that specifies the mapping of the psychological variable onto sudomotor nerve activity (SNA), and a peripheral (effector organ) model that specifies how SNA maps onto measured SCR (Alexander et al., 2005; Bach et al., 2009; Lim et al., 1997). For SCR, this split is useful

because the peripheral model can be evaluated on its own by intraneural stimulation and recordings from well accessible peripheral nerves (Gerster et al., 2017). For some other measures, the peripheral model can be approximated by specific stimuli, for example, one can use luminance changes to elucidate pupil mechanics. However, for most PsPMs that have been created to date, neural and peripheral processes are more difficult to separate experimentally as the efferent nerves are less accessible (at least in humans), and in current models they are either collapsed, or the distinction is only used for mathematical convenience.

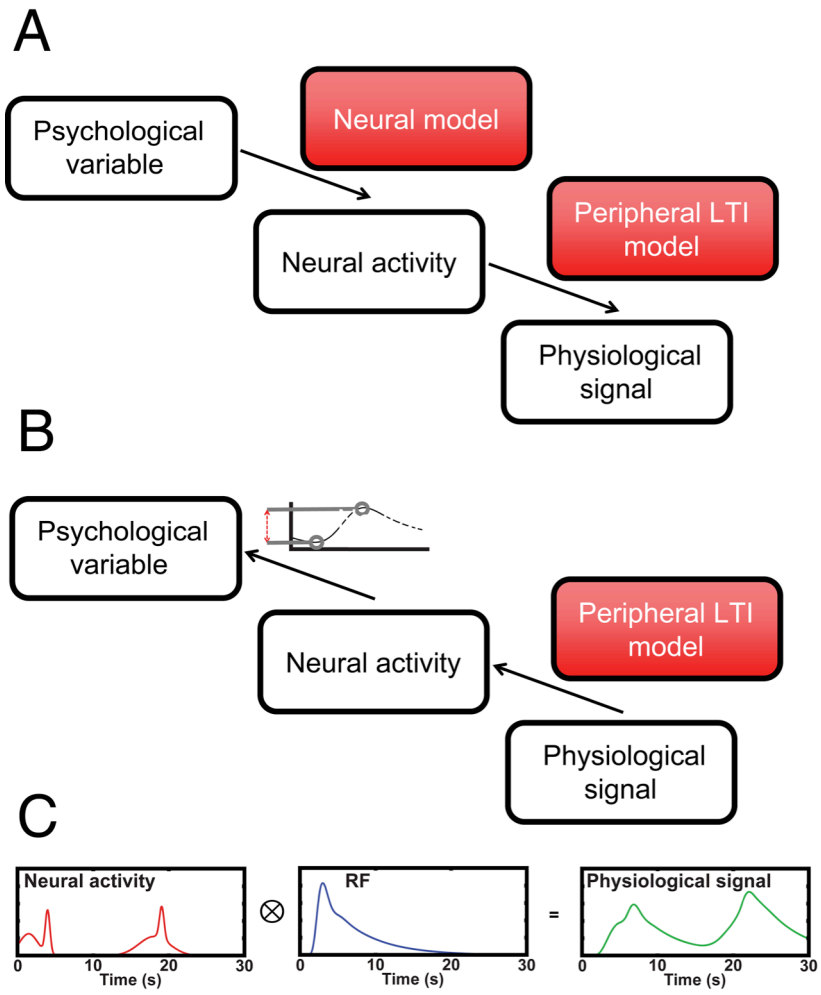

Figure 3. A: Basic formalism of most psychophysiological models. B: Related hybrid approaches (e.g. Ledalab, cvxEDA) use a standard response model to infer a (noisy) time series of neural inputs, and select data features of that time series as "equivalent" to the psychological variable. C: Linear time invariant (LTI) systems lie at the core of all existing psychophysiological modelling and hybrid approaches. A neural input is convolved with a canonical (experiment-invariant) response function, to yield a prediction for the measured signal.

3.1 Hybrid approaches

The empirical distinction between neural and peripheral models for SCR has early on motivated a hybrid approach (Alexander et al., 2005), engendered, for example, in the

softwares Ledalab (Benedek & Kaernbach, 2010a, 2010b) or cvxEDA (Greco, Valenza, Lanata, Scilingo, & Citi, 2015). In this approach, a deterministic peripheral model is inverted to compute a noisy SNA time series from the time series of measured SCR data. To make inference on psychological variables, some data features (peak-to-trough measures) of the SNA time series are extracted, in line with more traditional operational analysis (**Figure 3b**). This second mapping is heuristically motivated, not quantitatively specified or evaluated. Two different study groups have compared Ledalab with peak-to-trough methods on the one hand and a full PsPM-based approach on the other, in paradigms with relatively short intertrial intervals. On average, the hybrid Ledalab approach was found to yield similar predictive validity as directly using peak-to-trough measures of the SCR data, and its predictive validity was surpassed by inversion of full PsPMs (Bach, 2014; Green, Kragel, Fecteau, & LaBar, 2014). Systematic evaluation of cvxEDA has not yet been conducted. Notably, these approaches differ from the PsPM approach also in their peripheral forward models and statistical methods for model inversion. They could possibly be extended to directly estimate psychological variables, but retaining their specific forward models and inversion methods.

3.2 Linear time invariant systems

All PsPMs and hybrid models that have been proposed up to today contain at their heart a linear time invariant (LTI) system (**Figure 3c**). A LTI system is a system the output of which does not explicitly depend on time (time invariance), and the output to the sum of two inputs is just the sum of the outputs of the individual inputs (linearity). The first principle implies that different response shapes are explained by different inputs. According to the linearity principle, the magnitude of a response does not depend on the baseline. LTI systems are unambiguously described by the mathematical operation of convolution (overlap integral) of an input time series with a response function (RF), which corresponds in signal processing terms to a linear filter (see Appendix Equation 2). LTI systems constitute a mathematical simplification of real biophysical systems, which contain many parameters that cannot be usefully constrained from measured data. We will report for each measure how can it be described by a LTI system.

3.3 Model inversion

Many PsPMs assume that experimental events rapidly engage a psychological process, which then feeds into the physiological system with constant latency and shape. Under

these assumptions, the amplitude of the psychological input can be estimated in a general linear (convolution) model (Bach et al., 2009), similar to standard approaches for analysis of fMRIs (Friston, Jezzard, & Turner, 1994). In a nutshell, the RF is convolved with a time series of impulses (delta functions) centered on experimental events for each condition, and the ensuing time series form columns in the design matrix of a multiple regression model. The estimated coefficient of an individual event column constitutes the amplitude estimate for that condition (see Appendix Equation 3-5). Loosely speaking, the RF is regressed onto the observed response, and the regression coefficient is the estimate of the input amplitude. If latency and/or shape of the psychological input cannot be assumed to be constant, then they need to be estimated as well, and the inversion model becomes nonlinear, for example, in SCR models for fear conditioning (Bach, Daunizeau et al., 2010) or startle eye blink response (SEBR) models (Khemka, Tzovara, Gerster, Quednow, & Bach, 2017).

4 Psychophysiological models for different measurements

Table 1 gives an overview of the different psychophysiological models proposed until today.

Table 1: Psychophysiological models developed until today. PSR: pupil size responses. SCR: skin conductance responses. HPR: Heart period responses. RPR: respiration period responses. RAR: respiration amplitude responses. RFRR: respiratory flow rate responses. SEBR: startle eye blink responses.

Measurement	Psychological variable	Neural model	Peripheral model	Model inversion	Software implementation	Published in
SCR	Generic phasic arousal	Instantaneous impulse	LTI system with parameters from empirical data	GLM	PsPM	Bach, 2014; Bach, Flandin, Friston, & Dolan, 2009; Bach, Flandin, Friston, & Dolan, 2010; Bach, Friston, & Dolan, 2013
	Generic phasic arousal (validated for fear conditioning)	Constrained Gaussian impulse	LTI system with parameters from empirical data	Variational Bayes	PsPM	Bach, Daunizeau, Friston, & Dolan, 2010; Staib, Castegnetti, & Bach, 2015
	Generic tonic arousal (validated for anxiety and cognitive load)	Gaussian impulses with constant shape and unconstrained onset	LTI system with parameters from empirical data	Variational Bayes	PsPM	Bach, Daunizeau, Kuelzow, Friston, & Dolan, 2011; Bach, Friston, & Dolan, 2010
	Generic	Not specified	LTI system with parameters from theoretical considerations	Deterministic inverse filter	Ledalab	Benedek & Kaernbach, 2010a; Benedek & Kaernbach, 2010b
	Generic	Discrete impulses with unconstrained onset	LTI system with parameters from theoretical considerations	Convex optimisation	cvxEDA	Greco, Valenza, Lanata, Scilingo, & Citi, 2015
PSR	Luminance adaptation	Instantaneous impulse	Combination of 2 LTIs with parameters from empirical data	GLM	PsPM	Korn & Bach, 2016
	Attention	Instantaneous impulse	LTI with parameters from empirical data	OLS estimation in frequency domain	Pupil	Hoeks & Levelt, 1993
	Fear conditioning	Instantaneous impulse	LTI with parameters from	GLM	PsPM	Korn, Staib, Tzovara, Castegnetti, & Bach,

			empirical data			
RPR	Not yet specified	Instantaneous impulse	LTI with parameters from empirical data	GLM	PsPM	Bach, Gerster, Tzovara, & Castegnetti, 2016
RFRR	Not yet specified	Instantaneous impulse	LTI with parameters from empirical data	GLM	PsPM	Bach, Gerster, Tzovara, & Castegnetti, 2016
RAR	Not yet specified	Instantaneous impulse	LTI with parameters from empirical data	GLM	PsPM	Bach, Gerster, Tzovara, & Castegnetti, 2016
	Fear conditioning	Instantaneous impulse	LTI with parameters from empirical data	GLM	PsPM	Castegnetti, Tzovara, Staib, Gerster, & Bach, 2017
SEBR	Generic startle reflex	Instantaneous impulse with variable latency	LTI with parameters from empirical data	Template matching/GLM	PsPM	Khemka, Tzovara, Gerster, Quednow, & Bach, 2017

4.1 Skin conductance

4.1.1 Forward model

Skin conductance is often used to infer phasic or tonic sympathetic arousal generated by a wide range of psychological stimuli and tasks (Boucsein, 2012). Opening of sweat glands, elicited via the sympathetic nervous system with negligible parasympathetic transmission, causes phasic increases of skin conductance that are termed SCR (see Boucsein, 2012, for the physiology of SCR). Slow C fibers carrying impulses to the sweat glands are termed sudomotor (SN), and their activity can be measured by intraneural recordings. From their end terminal, the neurotransmitter acetylcholine diffuses through the skin to reach sweat glands, a process on the time scale of up to a second. In the history of psychophysiological modeling, it was recognized early on that nonoverlapping SCR can be well described by a simple response function, and overlapping SCR can be seen as being generated by a LTI system (Alexander et al., 2005; Bach et al, 2009). All published SCR models have therefore assumed that the mapping from SNA to SCR (but not necessarily from psychological variable to SNA) is well described by a LTI system.

Two types of response functions have been proposed: one based on a biophysical model of the sweat gland, with parameters set from theoretical considerations (Alexander et al., 2005; Benedek & Kaernbach, 2010a, 2010b; Greco et al., 2015), and a purely phenomenological function with parameters fitted to a database of 1,278 SCRs from 64 individuals in six different experimental conditions (Bach, Daunizeau et al., 2010; Bach et al., 2009; Bach, Flandin et al.,, 2010). While the former approach appears theoretically more rigorous, many of the biophysical parameters were not known from physiological research and had to be guessed. The ensuing forward model has not been systematically evaluated, and it is unclear how well this response function fits actual SCR. In contrast, the latter model is defined by its fit to empirical data. This phenomenological response function is mathematically described by a Gaussian-smoothed exponential (Bach, Flandin et al., 2010) or a third-order (linear, constant-coefficient) ordinary differential equation ; see Appendix- Equation 6, 7).

There are good empirical arguments to motivate the use of LTI systems to model the SNA/SCR relationship, provided that SCR data are high-pass filtered. Three kinds of tests have been exploited to evaluate the forward model. Indirect tests made the auxiliary

assumption that SN bursts follow external stimulation with constant shape and latency (this assumption is not part of the LTI system). These tests showed, that for short events (< 1 s duration) that are separated by at least 30 s, more than 60% of the variance in (high - pass filtered) SCR can be explained under a LTI model (Bach, Flandin et al., 2010), supporting the plausibility of the time invariance approximation. In the absence of stimulation, baseline variance exceeded the residual variance under stimulation, implying that the residual variance is due to noise rather than LTI violations (Bach, Flandin et al., 2010). SCR to pairs of stimuli separated by different intervals (2 - 9 s) do not depend on the interval, in line with the linearity principle (Bach, Flandin et al., 2010). A more direct test of the time invariance principle is furnished by intraneural recordings, which show that 60% - 75% of SCR variance is explained by a LTI model that takes SN activity as input, although this is still suffering from interfering non-SN (e.g., vasomotor nerve) activity (Gerster et al., 2017). A third approach relies on intraneural stimulation while blocking interfering nerve traffic by regional anesthesia. Here, SN is stimulated at different repetition frequencies, thus simultaneously addressing the linearity and time invariance principle. In this case, 93% - 99% of SCR variance is explained under a LTI model when stimulation frequency is below 0.6 Hz. Above this stimulation frequency, the LTI model cannot be usefully applied due to strong nonlinearities (Gerster et al., 2017); however, this limitation should be largely irrelevant for most psychological experiments with slower stimulation rates. To summarize, it appears that under suitable conditions, a LTI model is not just an approximation but rather an accurate description of biophysical reality in the SNA/SCR system. Notably, tonic skin conductance components, which are filtered out in all modeling approaches, do not appear to be modeled by a finite LTI system.

The LTI model does not make the assumption that SCR shape is constant between individuals. On the contrary, all evidence suggests that this is not the case. Nevertheless, inference based on a canonical response function already provides better inference than operational methods. We have shown that this inference can be further improved by allowing some variability between individuals, but only if the possible individual response functions are very strongly constrained to avoid overfitting (Bach et al., 2013; Staib, Castegnetti, & Bach, 2015).

For the mapping from psychological variable to SNA, three different forward models have been developed and evaluated. First, short stimuli elicit rapid SNA with constant Q6 latency (Bach et al., 2009). It appears that the mapping from psychological variable to SCR is largely invariant to the type of experiment stimulus (aversive white noise bursts, aversive electric stimulation, aversive pictures, auditory oddballs, and a visual detection task; Bach, Flandin et al.,, 2010), which motivates the use of this model for phasic arousal independent of the eliciting stimulus or experiment. For longer stimuli or anticipation of stimuli, a model with variable SNA latency and shape can be used, and these two parameters are estimated from the experimental data (Bach, Daunizeau et al., 2010). This model was motivated specifically to analyze fear condition experiments in which participants are exposed to conditioned stimuli (CS), one of which (CS+) predicts an aversive event (US). Typically, there is a time delay between CS+ and US, and so participants will anticipate the US during CS+ and (to a diminishing extent) during CS- presentation and express SCR at some (unknown and possibly variable) time point during this interval. Notably (and different from models discussed later), responses to CS+ and CS- are thought to be governed by the same psychological process, but to be quantitatively dissimilar. Hence, in this application of the model, a conditioned sudomotor nerve response to each CS is estimated, and the difference in their amplitude between CS+ and CS- constitutes the inference on fear memory. Finally, to account for spontaneous SCR fluctuations, a model is proposed in which constant-shape SN bursts occur with variable onset and amplitude, which are estimated from the data (Bach, Daunizeau et al.,, 2011).

4.1.2 Inference on psychological variables

The constant-latency forward model is inverted in a GLM approach (Bach et al.,, 2009) and has been evaluated on independent data sets (Bach, 2014; Bach et al., 2013) in comparison to different peak-scoring measures (Boucsein et al., 2012) and to measures from the hybrid Ledalab approach (Benedek & Kaernbach, 2010a,, 2010b). Predictive validity was assessed for the comparison of negative arousing versus neutral picture presentation, positive arousing versus neutral pictures, picture presentation versus no stimulus, and fearful versus angry face presentation. Predictive validity was decisively higher for the GLM-based approach on the majority of comparisons, and it was never decisively surpassed by another method in the remaining comparisons (Bach, 2014). In an evaluation study from an independent laboratory, a GLM approach had higher

predictive validity than peak-scoring or Ledalab for distinguishing CS+ and CS- in aversive learning (Green et al., , 2014). For distinguishing five different phases of a fear generalization experiment, peak scoring appeared to have higher predictive validity (Green et al., 2014). However, note that the predictive validity evaluation assumes that distinguishable different psychological states are created by the experiment, which is less well established for the latter manipulation. This study did not allow for assessing whether any method was decisively better or worse than another.[3]

The flexible-latency forward model is inverted in a variational Bayes approach (Bach, Daunizeau et al., 2010) and was evaluated on independent fear conditioning data sets (Staib et al, 2015) in comparison to different peak-scoring measures (Boucsein et al., 2012) and to measures from the hybrid Ledalab approach (Benedek & Kaernbach, 2010a, 2010b). Predictive validity was assessed for the comparison of CS+ versus CS-, and was found to be decisively higher for the model-based approach than for peak-scoring or Ledalab measures (Staib et al., 2015).

Finally, the flexible-onset model is inverted with the same variational Bayes approach (Bach, Daunizeau et al., 2011) and can be used to infer tonic arousal, which is often operationalized as the number of spontaneous skin conductance fluctuations (Boucsein et al., 2012). This method was evaluated with respect to distinguishing public speaking anxiety from rest or nonpublic speaking anxiety, and mental load from rest. Predictive validity of the model-based approach and of an automated peak-count measure was comparable (Bach & Staib, 2015). Furthermore, under the peripheral LTI model, the area under the curve of a time series equals the number of spontaneous fluctuations times their amplitude. While this relation has been empirically validated, the same study also showed that the number of spontaneous fluctuations alone allows better inference on tonic arousal than the area under the curve (Bach, Friston, & Dolan, 2010).

4.1.3 Future directions

[3] Bayes factors are reported in this study as well, but they are not comparable between the models evaluated. Comparison of model evidence is only possible if the dependent variable in the model is the same (Burnham & Anderson, 2004) but Green et al. (2014) use estimates of psychological state as dependent variable, which are obviously different between methods.

The constant-latency GLM approach appears fairly mature. It has been optimized with respect to model complexity and data preprocessing (Bach et al.,, 2013) and evaluated by two different laboratories (Bach, 2014; Green et al., 2014). Current research focuses on incremental improvements, such as data preprocessing and modeling of nonlinearities under specific conditions such as rapid event succession, or when the sweat ducts qualitatively change their response behavior (Tronstad, Kalvoy, Grimnes, & Martinsen, 2013).

The flexible-latency model is also in a mature stage and has been optimized (Staib et al., 2015), but a formal evaluation by different research laboratories is lacking. In terms of the forward model, a question remains as to under what conditions additional complexity afforded by estimating response latency improves inference on the psychological variable (i.e., What is the level of variability in response latency that should motivate preferring this method to the constant-latency GLM approach?). This question is awaiting empirical investigation. A weakness of the flexible-latency model is its complexity, such that it is impossible to estimate all parameters at the same time on standard PCs and cluster cores. Parameters are therefore estimated in a trial-by-trial fashion such that estimation errors that occur in one trial will propagate into the next one. While some technical tricks reduce a detrimental impact of this sequential estimation, it may be possible to improve on parameter estimation with a global optimization method that evaluates the entire parameter space at the same time.

Finally, the flexible-onset model is the least mature and has been evaluated on only a small number of questions and data sets; there is room for optimization of this method.

Beyond SCR, other measures also allow inference on SNA and, thus, on the same psychological variables. It remains to be determined whether measures such as skin potential and skin susceptance (Tronstad et al., 2013) or the memristor properties of the skin (Pabst, Tronstad, & Martinsen, 2017) yield additional information or can help reduce noise in the inverse inference.

4.2 Pupil size

4.2.1 Forward model

A wide range of psychological variables impacts on pupil size (e.g., de Gee et al., 2014; Joshi, Li, Kalwani, & Gold, 2016). Pupil size is controlled by two antagonist muscles: the radial M. dilatator pupillae, which receives sympathetic innervation via preganglionic neurons from the spinal cord and postganglionic neurons from the superior cervical ganglion, and the circular M. sphincter pupillae, which receive parasympathetic input from preganglionic neurons in
the Edinger–Westphal nucleus within the midbrain and postganglionic neurons in the ciliary ganglion (McDougal & Gamlin, 2008). The Edinger–Westphal nucleus receives both luminance-mediated inputs from the retina and appears to relay inputs that are not related to luminance (e.g., from the locus coeruleus; Joshi et al., 2016; Liu, Rodenkirch, Moskowitz, Schriver, & Wang, 2017). This innervation motivates using luminance changes to probe the biophysics of phasic pupil responses. The fact that the pupil is controlled separately by both branches of the autonomic nervous system, and by antagonistic muscles with different mechanical properties, already suggests that pupil responses may best be modeled by two parallel LTI systems. As a parsimonious description, a model is proposed that does not split the system into contributions of the two muscles but rather separates a slower dilation/constriction response from a faster component that only occurs for constrictions (see Appendix Equation 9, 10 and parameters for PSR_dil and PSR_con; Korn & Bach, 2016). Interestingly, the former component directly (exponentially) relates to luminance, while the contribution of the latter component appears independent from the amount of luminance change (Korn & Bach, 2016). Because of the different time constants of the two systems, the model makes an interesting counterintuitive prediction: light flashes should lead to pupil constriction, but brief darkness periods should also lead to constriction, when the faster response induced by the return to light precedes the slower response induced by the darkness period. Indeed, this prediction was confirmed by experimental observation in our own laboratory (Korn & Bach, 2016) and by others (Barbur, Harlow, & Sahraie, 1992). This forward model explained around 60% of signal variance, speaking to the validity of LTI approximation (Korn & Bach, 2016). Furthermore, it was used to infer the neural input into the pupil system for different psychological tasks (visual detection, auditory oddball detection, listening to emotional words). The inferred input latency meaningfully related to known underlying psychological processes (Korn & Bach, 2016). This also suggests that the fast time course of pupil responses implies a necessity to build neural forward models that are to some extent specific to the psychological

process studied - different, for example, from the unspecific SCR models - because the time constants of the psychological processes differ. One such model was developed for fear conditioning (Korn, Staib, Tzovara, Castegnetti, & Bach, 2017). Here, we found that CS- responses are different between experiments, depending on the perceptual modality and physical properties of the CS. However, the added impact of the CS+ was rather constant across experiments and could be modeled with a constant-latency CS-evoked neural input that peaks between CS and US (Korn et al., 2017). This neural input appeared to be time-locked to the CS, for different intervals between CS and US. Different from SCR models, this means that there is no common CS response that differs only in amplitude between conditions. Instead, the model seeks to estimate to what extent a specific CS+ component is expressed on each (CS+ and CS-) trial. Since this CS+ component may not be orthogonal to the experiment-specific CS response, this means that all estimates (CS+ and CS-) are only interpretable up to a constant: it is possible to interpret differences between trial sets, or temporal changes in the estimated CS response, but not the magnitude of response estimates for individual trials or conditions. For mathematical convenience, neural and peripheral model are collapsed into one response function, which makes GLM inversion possible (Korn et al., 2017; see Appendix Equation 9 and parameters for PSR_FC). Another psychological model with similar structure but a different response function was proposed to capture the impact of attention on pupil size (Hoeks & Levelt, 1993).

4.2.2 Inference on psychological variables

Pupil PsPMs have been used for two purposes. The first is direct inference on a psychological variable. Inference on CS+ memory in fear conditioning is implemented in a GLM inversion approach, and yields predictive validity superior to peak scoring or area-under-the-curve measures (Korn et al., 2017). Given the short latency and signal-to-noise levels similar or better than SCR, inference can be performed on a single-trial level. Similarly, inference on attention has been used in numerous studies (e.g., de Gee et al., 2014,2017; Knapen et al., 2016), although to our knowledge this method has not been systematically investigated and validated beyond the initial development data set (Hoeks & Ellenbroek, 1993; Hoeks & Levelt, 1993). Another application, distinct from all other models presented in this article, is to infer the time course of a psychological process. This is possible because luminance-related responses are mediated by near-instantaneous neural activity and thereby allow characterizing pupil biomechanics.

Comparing a measured pupil time course with the time course of a luminance-related response therefore affords estimating the neural input into the pupillary system, and thus makes inference on the dynamics of the underlying psychological process (Korn & Bach, 2016).

4.2.3 Future directions.

Pupil-size modeling is a relatively new approach and requires further study. Importantly, an independent validation of psychological inference is yet lacking. PsPMs exist for luminance-conditioned, fear-conditioned, and attention-related responses, and appear to crucially depend on the timing of the psychological process under study. Potentially, pupil-size modeling thus offers a more precise window into the temporal dynamics of cognitive processes than many other psychophysiological variables. Finally, a lot of research is currently being done on pupil size and its relation to cognitive processes, in humans and other species (e.g., Eldar, Cohen, & Niv, 2013; Joshi et al., 2016). It is likely that new models and methods will emanate from this basic research.

4.3 Heart period

4.3.1 Forward model

Heart rate or heart period are often used to infer emotional arousal, for example, while watching pictures or during fear conditioning (Berntson, Quigley, & Lozano, 2007; Bradley, Codispoti, Cuthbert, & Lang, 2001) and can be measured with electrocardiogram or pulse oxymeters. Cardiac rhythm is generated locally in the heart, but modulated under slower sympathetic and faster parasympathetic influence (Akselrod et al., 1981). Sympathetic stimulation frequency appears to linearly scale with heart period changes, not heart rate (Berntson, Cacioppo, & Quigley, 1995). Therefore current PsPMs for phasic cardiac responses model heart period, which is mapped onto the following R spike and linearly interpolated. It appears that various short stimuli induce phasic heart period responses (HPR), and six response components could be identified in a systematic investigation see Appendix Equation 8 and parameters for HPR_E1-E6). However, neither this study nor previous research based on operational methods allow a definite conclusion as to which psychological variables influence each of the components and relatedly, whether these components are independently controlled. In contrast, a well-replicated phenomenon is fear-conditioned bradycardia

(Castegnetti et al., 2016). This bradycardia response appears to be added to a stimulus-specific HPR that occurs for both CS+ and CS- in fear conditioning, similar to what is observed for pupil size and with the same limitations for interpretation of response estimates. However, different from pupil responses, it appears to be time-locked to the US when the CS-US interval is varied (Castegnetti, Tzovara, Staib, Gerster, & Bach, 2017; Castegnetti et al., 2016). Furthermore, relating the bradycardia response to the first response component elicited by short stimuli revealed a putative neural input that peaks at the US (Castegnetti et al., 2016). Pragmatically, the PsPM for fear-conditioned HPR collapses a neural and peripheral system into one response function, allowing GLM inversion (see Appendix Equation 9 and parameters for HPR_FC).

4.3.2 Inference on psychological variables

The range of psychological variables that can be inferred from phasic HPR to short stimuli appears unclear at present. Fear-conditioned bradycardia is a notable, well-studied exception. Using the GLM approach, fear memory could be inferred with higher predictive validity than with peak-scoring methods (Castegnetti et al., 2016). Unlike for SCR and PSR, attempts to perform single-trial analyses in our own laboratory have not succeeded, probably because heart-period time series are dominated by respiratory arrhythmia, and many trials are required to reduce the impact of this noise component.

4.3.3 Future directions

Elucidating the forward model for HPR appears an important task both in the context of PsPMs and operational approaches. The fidelity of inference on fear memory has been demonstrated in several data sets but requires independent confirmation from different laboratories. Castegnetti et al. (2017) have suggested that fear-conditioned bradycardia could potentially be - at least partly - induced by increased thorax pressure induced via respiration amplitude responses; this may be another interesting avenue of research.

4.4 Respiration measures

4.4.1 Forward model

Most respiratory psychophysiology research has focused on how psychological states on a time scale of 10- 20 s up to minutes influence respiration parameters (Boiten, Frijda, & Wientjes, 1994; Grassmann, Vlemincx, von Leupoldt, & Van den Bergh, 2015; Ritz et al., 2010; Vlemincx, Van Diest, & Van den Bergh, 2015; Wuyts, Vlemincx, Bogaerts,

Van Diest, & Van den Bergh, 2011), but relatively little is known about phasic respiratory responses. PsPMs have been developed for respiratory period, respiratory amplitude, and respiratory flow rate responses to brief external events, all measured with a simple single-chest belt system as standardly employed in fMRI laboratories (see Appendix Equation 8 and parameters for RPR, RAR_E, and RFRR). External events cause responses in these three measures that are captured with LTI systems, but as yet it is not clear which psychological states could be inferred from these responses (Bach, Gerster, Tzovara, & Castegnetti, 2016). In contrast, we have shown in several experiments that a CS+ in fear conditioning elicits a biphasic respiratory amplitude response that can be modeled in a LTI system, thus allowing GLM inversion (Castegnetti et al., 2017) (see Appendix Equation 9 and parameters for RAR_FC). The approach and its interpretation are similar to that for pupil size and heart period.

4.4.2 Inference on psychological variables

It appears that fear memory can be inferred from respiration amplitude responses, and predictive validity of this inference is higher for a GLM inversion than peak scoring (Castegnetti et al., 2017); however, it is lower than for many other psychophysiological measures. A distinct advantage of the respiratory PsPM could be that it only requires single-chest belt data, which is standardly available in many MRI scanners.

4.4.3 Future directions

More research is required to elucidate the range of psychological variables that can be inferred from respiratory measures. Modeling more sophisticated respiratory measures could be an interesting possibility.

4.5 Startle eyeblink electromyogram

4.5.1 Forward model

Different from the previously discussed measures, which are under the direct influence of a psychological variable, the impact of psychological variables on startle eyeblink is only modulatory and requires elicitation of a startle response to reveal it. This startle eyeblink response (SEBR) itself has rather stereotypical dynamics, while its amplitude is modulated by different psychological variables (Yeomans, Li, Scott, & Frankland, 2002). This modulation has been suggested to balance the protective utility of the startle response with its metabolic and opportunity cost (Bach, 2015b). Importantly, SEBR

dissociates CS+ and CS- in fear conditioning, a phenomenon termed fear-potentiated startle (Brown, Kalish, & Faber, 1951). A PsPM was developed to model the immediate, brief SEBR to startle probes in the absence of any psychological manipulation (Khemka et al., 2017). This model explained about 60% of signal variance under LTI assumptions. For inference, the neural input was allowed a flexible latency, to better capture slight latency variation between individuals and trials (Khemka et al., 2017). As common in the literature, the model assumes that the shape of SEBR is independent of the psychological or cognitive state, which only impacts on its amplitude.

4.5.2 Inference on psychological variables

This PsPM was employed to infer fear memory, both during acquisition and memory recall under extinction (Khemka et al., 2017). Predictive validity of the inference was compared to four peak-scoring methods with different preprocessing steps. For each of three experiments, a different peak-scoring method performed best, but across all experiments, the PsPM approach yielded highest predictive validity (Khemka et al., 2017).

4.5.3 Future directions

The impact of different preprocessing methods on SEBR analysis appears not well understood. This leads to a heterogeneous picture when comparing PsPM with different peak-scoring methods, and should be a focus of future research. Startle-independent eyeblinks are a typical source of noise in SEBR research, and modeling these eyeblinks could be an important topic for further investigation.

4.6 Combining psychophysiological models

In psychophysiological research, different measurement methods are often used simultaneously (in the spirit of convergent operationalization), but not commonly combined for statistical inference on a psychological variable. In a PsPM approach, this may be a possibility and could improve inference under circumstances where several measures are indicative of the same psychological variable. In order to enable such combination, it would be desirable to clarify qualitatively that two measures are impacted by the same psychological variable, and to investigate quantitatively the dimensional structure of these measures across different individuals.

5 Discussion

Psychological investigation relies on solving the inverse problem: making inference on essentially unobservable psychological variables. Psychophysiology benefits from many decades of research on the forward mapping from psychological variables onto physiological measures. This has allowed the building of precise and explicit forward models, which can be specified in mathematical form and inverted to yield inference on the psychological variable: psychophysiological modeling. Building on simple experimental manipulations that yield a known psychological state, methods can be evaluated in terms of their predictive validity (i.e., the fidelity with which they recover the known state). This allows comparison of operational as well as model-based methods and has revealed that in many cases PsPMs allow more precise inference than traditional operational methods, for example peak scoring.

As a limitation, this conclusion is based on a limited number of studies and data sets, most of which - with one exception (Green et al., 2014) - come from our laboratory and are thus based on the same recording equipment and rather comparable experimental procedures. It is possible that the PsPMs and inversion methods developed in this context overfit these experimental circumstances and do not generalize well to data acquired in different contexts, for example, using different experimental timings or involving other types of artifacts. Notably, the same limitation applies to the variety of operational methods that have often been developed in and for specific laboratories such that a multitude of operational analysis methods coexists in the literature. We have provided examples for application of the PsPMs in different laboratories, which provides circumstantial evidence that overfitting is not a major issue for the approach presented here. However, to entirely rule out such possibility, it would be desirable to compare the PsPMs with different operational analyses methods in many experimental situations.

Research on psychophysiological modeling underlines the necessity for precise specification of forward model, and thus for the detailed and meticulous work of basic psychophysiology. The application of PsPMs remains restricted to situations in which this mapping is well known. Indeed, PsPMs exist only for a small number of experimental scenarios. However, these include standard experiments such as fear conditioning and picture viewing, and may well comprise a large proportion of applied

psychophysiology research. For these, PsPMs offer improved inference compared to currently used methods. For exploratory research, operational methods with their flexibility may be more appropriate. However, as an advantage of model-based methods theirprecise specification implementation reduces researcher degrees of freedom. Analysis flexibility is a problem recognized across the entire field of psychology (Simmons, Nelson, & Simonsohn, 2011) and possibly more prevalent with flexible operational methods. While standardization of methods is an ongoing effort (Boucsein et al., 2012; Lonsdorf et al., 2017), PsPM offers rational criteria beyond community consensus for choosing the standards, namely, the quality of the inference. Notably, all methods discussed in this review are available in open-source toolboxes, most of them in the Matlab-based PsPM toolbox (pspm.sourceforge.net).

As a framework for evaluating methods, we have proposed to benchmark their predictive validity, that is, their ability to recover known psychological states (Bach & Friston, 2013). We note that this framework has potentially many more applications than evaluating PsPMs or comparing them to operational methods. For example, it allows power analyses. If the fidelity of a method is known in a standard experiment, it is often possible to derive best-case additional assumptions that define minimum sample sizes required to achieve a desired power level. We have provided an experimental example for this in the context of an intervention to reduce synaptic plasticity during fear conditioning, as measured with SEBR (Bach et al., 2017). As an important insight, power analysis based on predictive validity research has revealed that the required sample sizes - especially when using traditional operational methods - can be much higher than what is the standard in the field. Consequently, studies not basing their sample size on this or other formal analyses may be underpowered. Another potential application is quality control. Labs can compare their measurement methods between each other, assure measurement fidelity over time, or benchmark research trainees, using predictive validity of standard measures. Finally, a potentially powerful use of this framework is the optimization of experimental design. If the effect of a standard experiment on a psychological variable is known a priori, then one can choose an experimental design that best allows detecting this effect. As an example, this approach can help to find the optimal balance of retention trials to measure fear memory recall, for which we have provided an empirical example (Khemka et al., 2017).

To summarize, the field of psychophysiological modeling is moving rapidly and has in parts already matured. With these developments, we hope to have provided task-unspecific tools that free researchers' resources from having to develop data analysis procedures for every study, and instead focus on the psychological or cognitive questions they want to answer.

Acknowledgments

The authors thank their colleagues who over the years helped establishing the software PsPM and the methods described in this review: Jean Daunizeau, Raymond J. Dolan, Guillaume Flandin, Karl J. Friston, Gabriel Gräni, Saurabh Khemka, Philipp C. Paulus, Linus Rüttimann, Matthias Staib, Athina Tzovara.

References

Akselrod, S., Gordon, D., Ubel, F.A., Shannon, D.C., Berger, A.C., & Cohen, R.J. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science*, **213**, 220-222.

Alexander, D.M., Trengove, C., Johnston, P., Cooper, T., August, J.P., & Gordon, E. (2005). Separating individual skin conductance responses in a short interstimulus-interval paradigm. *Journal of Neuroscience Methods*, **146**, 116-123.

Alvarez, R.P., Kirlic, N., Misaki, M., Bodurka, J., Rhudy, J.L., Paulus, M.P., & Drevets, W.C. (2015). Increased anterior insula activity in anxious individuals is linked to diminished perceived control. *Translational psychiatry*, **5**, e591.

Bach, D.R. (2014). A head-to-head comparison of SCRalyze and Ledalab, two model-based methods for skin conductance analysis. *Biol Psychol*, **103C**, 63-68.

Bach, D.R. (2015a). Anxiety-Like Behavioural Inhibition Is Normative under Environmental Threat-Reward Correlations. *PLoS Comput Biol*, **11**, e1004646.

Bach, D.R. (2015b). A cost minimisation and Bayesian inference model predicts startle reflex modulation across species. *J Theor Biol*, **370**, 53-60.

Bach, D.R., Daunizeau, J., Friston, K.J., & Dolan, R.J. (2010). Dynamic causal modelling of anticipatory skin conductance responses. *Biol Psychol*, **85**, 163-170.

Bach, D.R., Daunizeau, J., Kuelzow, N., Friston, K.J., & Dolan, R.J. (2011). Dynamic causal modeling of spontaneous fluctuations in skin conductance. *Psychophysiology*, **48**, 252-257.

Bach, D.R., Flandin, G., Friston, K.J., & Dolan, R.J. (2009). Time-series analysis for rapid event-related skin conductance responses. *J Neurosci Methods*, **184**, 224-234.

Bach, D.R., Flandin, G., Friston, K.J., & Dolan, R.J. (2010). Modelling event-related skin conductance responses. *International Journal of Psychophysiology*, **75**, 349-356.

Bach, D.R., & Friston, K.J. (2013). Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, **50**, 15-22.

Bach, D.R., Friston, K.J., & Dolan, R.J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International Journal of Psychophysiology*, **76**, 52-55.

Bach, D.R., Friston, K.J., & Dolan, R.J. (2013). An improved algorithm for model-based analysis of evoked skin conductance responses. *Biol Psychol*, **94**, 490-497.

Bach, D.R., Gerster, S., Tzovara, A., & Castegnetti, G. (2016). A linear model for event-related respiration responses. *J Neurosci Methods*, **270**, 147-155.
Bach, D.R., Seifritz, E., & Dolan, R.J. (2015). Temporally Unpredictable Sounds Exert a Context-Dependent Influence on Evaluation of Unrelated Images. *PLoS One*, **10**, e0131065.
Bach, D.R., & Staib, M. (2015). A matching pursuit algorithm for inferring tonic sympathetic arousal from spontaneous skin conductance fluctuations. *Psychophysiology*, **52**, 1106-1112.
Bach, D.R., Tzovara, A., & Vunder, J. (2017). Blocking human fear memory with the matrix metalloproteinase inhibitor doxycycline. *Mol Psychiatry*.
Bach, D.R., Weiskopf, N., & Dolan, R.J. (2011). A stable sparse fear memory trace in human amygdala. *Journal of Neuroscience*, **31**, 9383-9389.
Barbur, J.L., Harlow, A.J., & Sahraie, A. (1992). Pupillary responses to stimulus structure, colour and movement. *Ophthalmic Physiol Opt*, **12**, 137-141.
Barry, R.J., Feldmann, S., Gordon, E., Cocker, K.I., & Rennie, C. (1993). Elicitation and habituation of the electrodermal orienting response in a short interstimulus interval paradigm. *International Journal of Psychophysiology*, **15**, 247-253.
Benedek, M., & Kaernbach, C. (2010a). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, **190**, 80-91.
Benedek, M., & Kaernbach, C. (2010b). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, **47**, 647-658.
Berntson, G.G., Cacioppo, J.T., & Quigley, K.S. (1995). The metrics of cardiac chronotropism: biometric perspectives. *Psychophysiology*, **32**, 162-171.
Berntson, G.G., Quigley, K.S., & Lozano, D. (2007). Cardiovascular Psychophysiology. In J.T.T. Cacioppo, L.G.; Berntson,G.G. (Ed.), *Handbook of Psychophysiology*. New York City: Cambridge University Press.
Boiten, F.A., Frijda, N.H., & Wientjes, C.J. (1994). Emotions and respiratory patterns: review and critical analysis. *Int J Psychophysiol*, **17**, 103-128.
Boucsein, W. (2012). *Electrodermal activity*. New York Springer.
Boucsein, W., Fowles, D.C., Grimnes, S., Ben-Shakhar, G., roth, W.T., Dawson, M.E., & Filion, D.L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, **49**, 1017-1034.
Bradley, M.M., Codispoti, M., Cuthbert, B.N., & Lang, P.J. (2001). Emotion and Motivation I: Defensive and Appetitive Reactions in Picture Processing. *Emotion*, **1**, 276-298.
Brown, J.S., Kalish, H.I., & Faber, I.E. (1951). Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *Journal of Experimental Psychology*, **41**, 317-328.
Bulganin, L., Bach, D.R., & Wittmann, B.C. (2014). Prior fear conditioning and reward learning interact in fear and reward networks. *Frontiers in Behavioral Neuroscience*, **8**, 67.
Castegnetti, G., Tzovara, A., Staib, M., Gerster, S., & Bach, D.R. (2017). Assessing fear learning via conditioned respiratory amplitude responses. *Psychophysiology*, **54**, 215-223.
Castegnetti, G., Tzovara, A., Staib, M., Paulus, P.C., Hofer, N., & Bach, D.R. (2016). Modeling fear-conditioned bradycardia in humans. *Psychophysiology*, **53**, 930-939.
de Berker, A.O., Rutledge, R.B., Mathys, C., Marshall, L., Cross, G.F., Dolan, R.J., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nat Commun*, **7**, 10996.

de Gee, J.W., Colizoli, O., Kloosterman, N.A., Knapen, T., Nieuwenhuis, S., & Donner, T.H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *Elife*, **6**.

de Gee, J.W., Knapen, T., & Donner, T.H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences of the USA*, **111**, E618-625.

Eldar, E., Cohen, J.D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, **16**, 1146-1153.

Fan, J., Xu, P., Van Dam, N.T., Eilam-Stock, T., Gu, X., Luo, Y.-j., & Hof, P.R. (2012). Spontaneous brain activity relates to autonomic arousal. *The Journal of Neuroscience*, **32**, 11176-11186.

Friston, K.J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Hum Brain Mapp*, **1** 153-171.

Gerster, S., Namer, B., Elam, M., & Bach, D.R. (2017). Testing a linear time invariant model for skin conductance responses by intraneural recording and stimulation. *Psychophysiology*.

Grassmann, M., Vlemincx, E., von Leupoldt, A., & Van den Bergh, O. (2015). The role of respiratory measures to assess mental load in pilot selection. *Ergonomics*, 1-9.

Greco, A., Guidi, A., Felici, F., Leo, A., Ricciardi, E., Bianchi, M., Bicchi, A., Citi, L., Valenza, G., & Scilingo, E.P. (2017). Muscle fatigue assessment through electrodermal activity analysis during isometric contraction. *Conf Proc IEEE Eng Med Biol Soc*, **2017**, 398-401.

Greco, A., Lanata, A., Valenza, G., Di Francesco, F., & Scilingo, E.P. (2016). Gender-specific automatic valence recognition of affective olfactory stimulation through the analysis of the electrodermal activity. *Conf Proc IEEE Eng Med Biol Soc*, **2016**, 399-402.

Greco, A., Valenza, G., Lanata, A., Scilingo, E., & Citi, L. (2015). cvxEDA: a Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Trans Biomed Eng*.

Greco, A., Valenza, G., & Scilingo, E.P. (2016). Investigating mechanical properties of a fabric-based affective haptic display through electrodermal activity analysis. *Conf Proc IEEE Eng Med Biol Soc*, **2016**, 407-410.

Green, C.D. (1992). Of Immortal Mythological Beasts: Operationism in Psychology. *Theory & Psychology*, **2**, 291-320.

Green, S.R., Kragel, P.A., Fecteau, M.E., & LaBar, K.S. (2014). Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis. *Int J Psychophysiol*, **91**, 186-193.

Hayes, D.J., Duncan, N.W., Wiebking, C., Pietruska, K., Qin, P., Lang, S., Gagnon, J., Bing, P.G., Verhaeghe, J., Kostikov, A.P., Schirrmacher, R., Reader, A.J., Doyon, J., Rainville, P., & Northoff, G. (2013). GABAA receptors predict aversion-related brain responses: an fMRI-PET investigation in healthy humans. *Neuropsychopharmacology*, **38**, 1438-1450.

Hoeks, B., & Ellenbroek, B.A. (1993). A neural basis for a quantitative pupillary model. *Journal of Psychophysiology*, **7**, 315-315.

Hoeks, B., & Levelt, W.J.M. (1993). Pupillary Dilation as a Measure of Attention - a Quantitative System-Analysis. *Behavior Research Methods Instruments & Computers*, **25**, 16-26.

Joshi, S., Li, Y., Kalwani, R.M., & Gold, J.I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, **89**, 221-234.

Khemka, S., Tzovara, A., Gerster, S., Quednow, B.B., & Bach, D.R. (2017). Modeling startle eyeblink electromyogram to assess fear learning. *Psychophysiology*, **54**, 204-214.
Knapen, T., de Gee, J.W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and Ocular Factors Jointly Determine Pupil Responses under Equiluminance. *PLoS One*, **11**, e0155574.
Koban, L., Kusko, D., & Wager, T.D. (2018). Generalization of learned pain modulation depends on explicit learning. *Acta Psychol (Amst)*, **184**, 75-84.
Koban, L., & Wager, T.D. (2016). Beyond conformity: Social influences on pain reports and physiology. *Emotion*, **16**, 24.
Korn, C.W., & Bach, D.R. (2016). A solid frame for the window on cognition: Modeling event-related pupil responses. *J Vis*, **16**, 28.
Korn, C.W., Staib, M., Tzovara, A., Castegnetti, G., & Bach, D.R. (2017). A pupil size response model to assess fear learning. *Psychophysiology*, **54**, 330-343.
Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6*. Gainesville, FL: University of Florida.
Lim, C.L., Rennie, C., Barry, R.J., Bahramali, H., Lazzaro, I., Manor, B., & Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. *International Journal of Psychophysiology*, **25**, 97-109.
Liu, Y., Rodenkirch, C., Moskowitz, N., Schriver, B., & Wang, Q. (2017). Dynamic Lateralization of Pupil Dilation Evoked by Locus Coeruleus Activation Results from Sympathetic, Not Parasympathetic, Contributions. *Cell Rep*, **20**, 3099-3112.
Lonsdorf, T.B., Menz, M.M., Andreatta, M., Fullana, M.A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Romer, S., Shiban, Y., Schmitz, A., Straube, B., Vervliet, B., Wendt, J., Baas, J.M.P., & Merz, C.J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neurosci Biobehav Rev*, **77**, 247-285.
McDougal, D.H., & Gamlin, P.D.R. (2008). Pupillary control pathways.
Nicolle, A., Fleming, S.M., Bach, D.R., Driver, J., & Dolan, R.J. (2011). A regret-induced status quo bias. *Journal of Neuroscience*, **31**, 3320-3327.
Pabst, O., Tronstad, C., & Martinsen, O.G. (2017). Instrumentation, electrode choice and challenges in human skin memristor measurement. *Conf Proc IEEE Eng Med Biol Soc*, **2017**, 1844-1848.
Paulus, P.C., Castegnetti, G., & Bach, D.R. (2016). Modeling event-related heart period responses. *Psychophysiology*, **53**, 837-846.
Ritz, T., Kullowatz, A., Goldman, M.D., Smith, H.J., Kanniess, F., Dahme, B., & Magnussen, H. (2010). Airway response to emotional stimuli in asthma: the role of the cholinergic pathway. *J Appl Physiol (1985)*, **108**, 1542-1549.
Simmons, J.P., Nelson, L.D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, **22**, 1359-1366.
Staib, M., Castegnetti, G., & Bach, D.R. (2015). Optimising a model-based approach to inferring fear learning from skin conductance responses. *J Neurosci Methods*, **255**, 131-138.
Sulzer, J., Sitaram, R., Blefari, M.L., Kollias, S., Birbaumer, N., Stephan, K.E., Luft, A., & Gassert, R. (2013). Neurofeedback-mediated self-regulation of the dopaminergic midbrain. *Neuroimage*, **83**, 817-825.

Talmi, D., Dayan, P., Kiebel, S.J., Frith, C.D., & Dolan, R.J. (2009). How humans integrate the prospects of pain and reward during choice. *Journal of Neuroscience*, **29**, 14617-14626.
Tronstad, C., Kalvoy, H., Grimnes, S., & Martinsen, O.G. (2013). Waveform difference between skin conductance and skin potential responses in relation to electrical and evaporative properties of skin. *Psychophysiology*, **50**, 1070-1078.
Tzovara, A., Korn, C. W., & Bach, D. R. (2018). Human Pavlovian fear conditioning conforms to probabilistic learning. PLOS Computational Biology, 14, e1006243.
Vlemincx, E., Van Diest, I., & Van den Bergh, O. (2015). Emotion, sighing, and respiratory variability. *Psychophysiology*, **52**, 657-666.
Wuyts, R., Vlemincx, E., Bogaerts, K., Van Diest, I., & Van den Bergh, O. (2011). Sigh rate and respiratory variability during normal breathing and the role of negative affectivity. *Int J Psychophysiol*, **82**, 175-179.
Yeomans, J.S., Li, L., Scott, B.W., & Frankland, P.W. (2002). Tactile, acoustic and vestibular systems sum to elicit the startle reflex. *Neurosci Biobehav Rev*, **26**, 1-11.

Appendix

Model evidence

To quantify model evidence, PsPM uses the following identity to compute Akaike Information Criterion (AIC):

$$AIC = -2\ln(L) + 2k = n\ln\left(\frac{RSS}{n}\right) + 2k \quad (1)$$

where n is the number of data points in the predictive model, L is the maximum of the likelihood function, and k the number of parameters in the predictive model. k is constant for all methods that are evaluated in a methods comparison. The predictive model uses the a priori defined psychological variable as dependent variable (this ensures it is the same across all methods), and the design matrix contains the estimated psychological variable and (possibly subject-specific) intercept terms.

LTI systems

Linear time invariant systems are defined by the following convolution operation:

$$y(t) = u \times h = \int_0^\infty u(t-\tau)h(\tau)\mathrm{d}\tau \quad (2)$$

where $u(t)$ is the input into the system at time t, h is the impulse response function (RF), and τ is a dummy variable over which integration is performed. Note that since we are dealing with a causal system (i.e,. in time), we have explicitly set the lower limit of the

integral to zero, such that an input that occurs after *t* can have no impact upon the output at time t.

GLM

A GLM can be written as

$$Y = X\beta + \epsilon \tag{3}$$

where *Y* is the vector of observations, β is a vector of input amplitude parameters, and ϵ is the error. *X* is the design matrix in which each column is obtained by convolving impulse functions at known time points with each component of the RF. Each column takes the form:

$$X \quad = u_i(t) \times h_j(t) \tag{4}$$

where $u_i(t)$ is the neural input with unit amplitude for condition *i*, and *j* is the index of the RF component. Finally, X also contains a column for the intercept. The maximum-likelihood amplitude estimates $\hat{\beta}$ can be computed using the Moore-Penrose pseudoinverse X^+, for example implemented in the Matlab function pinv.m:

$$\hat{\beta} = X^+ Y \tag{5}$$

Skin conductance forward model

(a) Phenomenological RF described as a Gaussian-smoothed exponential:

$$\begin{aligned} &h(t) \propto \int_0^t N(t-\tau)\left(E_1(\tau) + E_2(\tau)\right)d\tau, \\ &t \geq 0, \\ &N(t) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-t_0)^2}{2\sigma^2}}, \\ &E_i(t) = e^{\lambda_i t} \end{aligned} \tag{6}$$

with estimated parameters: $\hat{t}_0 = 3.0745\ s$ for peak latency; $\hat{\sigma} = 0.7013$ for definition of the rise time; $\hat{\lambda}_1 = 0.3176$ and $\hat{\lambda}_2 = 0.0708$ to define the two decay components.

(b) Phenomenological RF described as third-order (linear, constant-coefficient) ordinary differential equation:

$$\dddot{y} = \vartheta_1 \ddot{y} + \vartheta_2 \dot{y} + \vartheta_3 y - u(t - \vartheta_4) = 0 \tag{7}$$

with $\hat{\vartheta}_1 = 1.342052, \hat{\vartheta}_2 = 1.411425, \hat{\vartheta}_3 = 0.122505, \hat{\vartheta}_4 = 1.533879$

Models with Gaussian response functions

Gaussian function:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad (8)$$

Parameters see table (HPR_E1-6: constant-latency heart period responses; RPR: respiration period responses; RAR_E: constant-latency respiration amplitude responses; RFRR: respiratory flow rate responses).

Response function (RF)	Parameters of Gaussian Function	
	μ (mean)	*σ (standard deviation)*
HPR_E 1	1.0	1.9
HPR_E 2	5.2	1.9
HPR_E 3	7.2	1.5
HPR_E 4	7.2	4.0
HPR_E 5	12.6	2.0
HPR_E 6	18.85	1.8
RPR	4.20	1.65
RAR_E	8.07	3.74
RFRR	6.00	3.23

Models with gamma response functions

Gamma function:

$$y = \frac{(x - x_0)^{k-1} e^{-\frac{x-x_0}{\theta}}}{\theta^k \Gamma(k)} \quad (9)$$

Parameters see table (HPR_RC: fear-conditioned heart period responses; PSR_dil: luminance-evoked pupil dilation; PSR_con: luminance-evoked pupil constriction; PSR_FC: fear-conditioned pupil size responses; RAR_FC :fear-conditioned respiration amplitude responses; SEBR: startle eyeblink responses). Note that the amplitude parameter is left free for all models other than the luminance models in which it has a physical interpretation.

Response function (RF)	Parameters of gamma Function			
	k (shape)	θ (scale)	x_0 (onset)	A (amplitude)

HPR_FC	48.5	0.182	-3.86	-
PSR_dil	2.40	2.40	0.2	0.77
PSR_con	3.24	0.18	0.2	0.43
PSR_FC	5.94	0.75	0.002	-
RAR_FC (early)	2.570×10^{7}	3.124×10^{-4}	-8.024×10^{3}	-
RAR_FC (late)	3.413	1.107	7.583	-
SEBR	3.5114	0.0108		-

Model for steady-state pupil size

$$d(E_v) = C + Ae^{BE_v} \tag{10}$$

where d is the z scored steady-state pupil diameter and E_v is the respective illuminance level in (in $[lx]$). The parameter values are:
$A = 49.79; B = -0.50\left[\frac{1}{lx}\right]; C = -1.05.$