Richter, M and Slade, K

Interpretation of Physiological Indicators of Motivation: Caveats and Recommendations

Richter, M and Slade, K (2017) Interpretation of Physiological Indicators of Motivation: Caveats and Recommendations. International Journal of Psychophysiology. ISSN 0167-8760

The International Journal of Psychophysiolgoy homepage: http://www.journals.elsevier.com/international-journal-of-psychophysiology

Richter, M., & Slade, K. (in press). Interpretation of physiological indicators of motivation: Caveats and recommendations. *International Journal of Psychophysiology.*

Interpretation of Physiological Indicators of Motivation: Caveats and Recommendations

Michael Richter[a] & Kate Slade[a]

[a] The Effort Lab, School of Natural Sciences and Psychology, Liverpool John Moores University, Tom Reilly Building, Byrom Street, Liverpool, L3 3AF, UK. E-mail: M. Richter@ljmu.ac.uk, K.Slade@2012.ljmu.ac.uk

Correspondence should be addressed to Michael Richter, School of Natural Sciences and Psychology, Liverpool John Moores University, Tom Reilly Building, Byrom Street, Liverpool L3 3AF, United Kingdom. E-mail: M.Richter@ljmu.ac.uk

Abstract

Motivation scientists employing physiological measures to gather information about motivation-related states are at risk of committing two fundamental errors: overstating the inferences that can be drawn from their physiological measures and circular reasoning. We critically discuss two complementary approaches, Cacioppo and colleagues' model of psychophysiological relations and construct validation theory, to highlight the conditions under which these errors are committed and provide guidance on how to avoid them. In particular, we demonstrate that the direct inference from changes in a physiological measure to changes in a motivation-related state requires the demonstration that the measure is not related to other relevant psychological states. We also point out that circular reasoning can be avoided by separating the definition of the motivation-related state from the hypotheses that are empirically tested.

Keywords: Motivation, psychophysiology, psychophysiological measures, indicator, marker

Interpretation of Physiological Indicators of Motivation: Caveats and Recommendations

1. Introduction

Most psychological professional bodies and associations, like the British Psychological Society (BPS), the German Psychological Society (DGPs), the American Psychological Association (APA), or the Association for Psychological Science (APS), emphasize the importance of sound method-related knowledge and skills for all psychologists independent of the specific field they are working in. It is therefore no wonder that classes on quantitative and qualitative methods, test and questionnaire construction, and philosophy of science are crucial features of many undergraduate and postgraduate psychology programs. Most psychologists underwent rigorous training in scientific methodology during their studies and this in-depth formation crystallizes in many research fields where psychologists conduct carefully designed, scientifically sound empirical research. Surprisingly, psychologists seem to struggle with applying their methodological expertise when it comes to employing physiological measures to study psychological phenomena like motivation.

Many psychologists using physiological measures to study motivation-related topics seem to be unaware of two major pitfalls and are consequently unable to avoid them. First, motivation scientists need to avoid interpreting changes in physiological measures as reflecting changes in motivation-related states if this conclusion is not warranted.[1] Second, they need to avoid circular reasoning when justifying the assessment of their physiological measures. Failing to avoid these pitfalls leads to inaccurate conclusions and decreases the scientific quality of the conducted research. However, many motivation scientists seem to be

1 Psychologists are of course not only interested in motivation-related states. They examine all kinds of psychological states and they are also interested in any type of psychological variable (like processes or properties). Given the focus of this special issue and for the sake of readability, we will use ‘motivation-related state’ throughout this article as placeholder for any psychological variable that a psychologist might be interested in.

unaware of these potential problems and lack the knowledge to avoid them. This article aims at providing them with the required knowledge by discussing two theoretical frameworks on the link between physiological measures and psychological states—Cacioppo, Tassinary, and Berntson's model of psychophysiological relations (Cacioppo & Tassinary, 1990; Cacioppo et al., 2000) and classical construct validity theory (e.g., Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Strauss & Smith, 2009; Trochim, 2016)—that illustrate the two pitfalls and their consequences. Referring to these frameworks, we will also provide guidance on how to avoid the pitfalls.

2. Pitfall 1: Inferring psychological states from physiological measures

As noted in the preceding section, motivation scientists often interpret changes in physiological measures as reflecting changes in motivation-related states. Examples include suggestions that the error-related negativity of event-related brain potentials indicates defensive reactivity (Weinberg et al., 2012), that heart rate measures affective arousal (Sideridis et al., 2014), or that the combination of EEG activity, pupillometric response, and skin conductance change provides information about the level of task engagement (Gergelyfi et al., 2015). It is understandable that motivation scientists are interested in using physiological measures to gather information about motivation-related states. It is, however, important to acknowledge that the existing empirical research on most physiological measures does not warrant such a strong inference (e.g., Cacioppo et al., 2000).

2.1. Psychophysiological relations

Cacioppo and colleagues provided a classification of the relationships between physiological measures and psychological states (Cacioppo & Tassinary, 1990; Cacioppo et al., 2000; see also Allanson & Fairclough 2004; Fairclough, 2009) which highlights the characteristics that a physiological measure needs to possess to enable inferences about a psychological (motivation-related) state. They distinguished four classes of relations between physiological measures and psychological states according to the level of specificity and generality. Table 1 provides an overview of these relations, the inferences that they enable, and the required validation.

A physiological measure that changes as a function of the manipulation of a psychological state is considered an *outcome*. An outcome is characterized by a *one-to-many relation* between the physiological measure and psychological states. It has been demonstrated that a specific psychological state affects the physiological measure in a certain context but there might be other psychological states that also influence the measure or the relation might not hold in other contexts.[2] This applies probably to most physiological measures used in the motivation-related literature and these measures thus constitute physiological outcomes of motivation-related states. For instance, de Morree and Marcora's (2010) observation that corrugator supercilii amplitude increased as a function of increasing difficulty of a leg extension task demonstrated that corrugator supercilii amplitude is an outcome of effort (assuming that leg extension difficulty manipulates effort).

If additional research demonstrates that the physiological measure responds similarly in many different contexts to variations in the psychological state, the physiological measure is called a *concomitant*. Physiological concomitants have a general one-to-many relation with psychological states. They are affected by many psychological states but the relations are independent of the specific context. In the case of the corrugator supercilii example, research

2 Cacioppo and colleagues' framework uses the term context in a broad sense. It refers to any aspect that can differ between two situations and does include the specific stimulus configuration present in a certain context. Psychophysiological studies on stimulus specificity (e.g., Brenner et al., 2005; Edelberg & Wright, 1964) thus constitute specific demonstrations of context-dependent relationships.

would need to demonstrate that corrugator supercilii amplitude responds in general—in many different contexts—to variations in effort.

If one can demonstrate that a physiological measure is only affected by a single psychological state, one has evidence for a *one-to-one relationship*. If this relationship only holds in a certain context, the physiological measure is called a *marker* of the psychological state. If the relationship is general, the physiological measure is called an *invariant*.[3] To warrant the conclusion that corrugator supercilii amplitude is a marker of effort, one would need to show that it is—in a certain context—only affected by changes in effort and not by changes in any other psychological state. To conclude that it is an invariant, one would have to demonstrate the one-to-one relationship in every context. Empirical evidence revealing that corrugator supercilii amplitude is also affected by other psychological states—like Cacioppo et al.’s (2000) demonstration that it changes as a function of mood—disqualifies the measure as an invariant of effort and would also disqualify it as a marker of effort if the empirical evidence had been gathered in a context similar to the context of de Morree and Marcora’s study.

Outcomes, concomitants, markers, and invariants differ considerably regarding the inferences that they enable. Invariants enable the type of conclusion that most motivation scientists are probably looking for. Given that the physiological measure and the motivation-related state have a general one-to-one relationship, one can directly infer the motivation-related state from the physiological measure. Any change in the measure reflects a change in the state. If corrugator supercilii amplitude were an invariant of effort, any change in its

3 Cacioppo and colleagues’ distinction between context-independent invariants and concomitants and context-dependent outcomes and markers resembles the distinction between endophenotypes, intermediate phenotypes, and biomarkers in psychopathology (Beauchaine, 2009; Lenzenweger, 2013a, 2013b; Puntmann, 2009). Endophenotypes and intermediate phenotypes are supposed to be context-independent because of their genetic underpinnings, whereas biomarkers are measures that correlate with some aspects of a disease but not necessarily in all contexts.

amplitude would announce a change in effort. When psychologists use physiological measures to find out whether drivers are in an optimal state for driving (Brookhuis & de Waard, 2010), to design ambulatory devices that monitor mental stress (Choi, Ahmed, & Gutierrez-Osuna, 2012), or to predict whether athletes feel challenged or threatened in a competition (Jones et al., 2009), they are keen on having a physiological invariant of the psychological state that they are interested in. In cognitive neuroscience this desire to infer psychological states from physiological (neurological) activity has been labeled reverse inference (e.g., Poldrack, 2011). Examples include the inference of reward processing from ventral striatum activity (Takahashi et al., 2009) and valuation from orbitofrontal cortex activation (Padoa-Schioppia & Assad, 2006).

As pointed out in the preceding paragraphs, an invariant is a physiological measure that is exclusively related to a single psychological variable and that shows this relationship independent of the specific context. To establish a physiological invariant, one would thus need to demonstrate that the physiological measure 1) is sensitive to variations in the motivation-related state of interest, 2) does not respond to changes in any other psychological variable, and 3) that this holds in any context. It is obvious that it is very difficult to fulfill the last two criteria. Given that the number of different psychological variables and contexts that exist (or that one can imagine) is probably unlimited or at least very high, one might never be able to demonstrate a general one-to-one relationship. One might lessen the requirements by only asking for an examination of the psychological variables and contexts that have frequently been the topic of psychophysiological research. However, even under these lowered requirements, there is probably no evidence for a physiological invariant of a psychological state, so far. At least, we do not know of any systematic, successful endeavor to demonstrate that a certain physiological measure is only related to a single motivation-related state and that this relationship is context-independent.

Physiological markers provide the same information as invariants except that their application is context-limited. Given that they also have a one-to-one relationship with the associated psychological state, they allow researchers to deduce a change in the psychological state from a change in the physiological marker. This inference would, however, only be valid in the context for which the one-to-one relation has been demonstrated. The requirements for validating a marker are nearly as high as the requirements for the validation of an invariant. One would need to demonstrate that—in a certain context—the physiological measure is 1) sensitive to variations in the motivation-related state of interest and 2) does not respond to changes in any other psychological variable.

To validate the pattern of increased cardiac output and increased total peripheral resistance that has been used by Jones and colleagues (2009) to identify athletes who perceive a task as threat, one would need to demonstrate in a first step that this pattern is actually linked to perceived threat. In a second step, one would need to demonstrate that the pattern does not appear in other psychological states, like stress, effort, anger, or competitiveness. To our knowledge there is some positive evidence showing that perceived threat results in the predicted pattern of cardiac output and total peripheral resistance (e.g., Tomaka et al., 1997). There is, however, no comprehensive empirical research demonstrating that other psychological states do not lead to the same pattern. Our view on the available empirical evidence for other physiological measures that have been suggested as markers of motivation-related states is as pessimistic. There seems to be no physiological measures for which empirical research has demonstrated that it is exclusively related to a single motivation-related state in a specific context.

The lack of systematic research demonstrating that a certain physiological measure is exclusively related to a single motivation-related state is understandable. It is obvious that motivation scientists are mainly interested in examining how the motivation-related variables that they are interested in affect physiological measures. It is much less exciting to demonstrate that psychological variables that one is not interested in have no impact on the measures. If you are interested in approach motivation like Harmon-Jones (2003), it is certainly more interesting for you to examine the relationship between approach motivation and asymmetrical frontal cortical activity than to show that frontal cortical activity differences are not related to mindfulness, stress, or listening effort. Moreover, it might be difficult to get funding for research that mainly aims at providing evidence for no effect of psychological states on a physiological measure.

However, the lack of this type of research resulted in a situation where we do not have any established physiological markers or invariants of motivation-related states—at least not in the sense of Cacioppo and colleagues' framework. Existing psychophysiological research does thus not offer physiological measures that can be used to draw direct inferences about the existence, absence, or magnitude of motivation-related states. This has important implications for the application of psychophysiological research. Consider the use of physiological measures to infer whether drivers are in an appropriate mental state for driving (e.g., Brookhuis & de Waard, 2010). A physiological marker or invariant enables the development of a system that effectively prevents individuals, who are at risk of causing an accident, from driving. The system could, for instance, automatically slow down the car when the physiological measures indicate that the driver is in a state that constitutes a risk for safety (for instance, when she or he is sleepy or distracted). Given the one-to-one relation associated with markers and invariants, the system would be specific. It would only respond to changes in mental states that are safety-relevant. It would not respond to changes in any other psychological state.

A physiological outcome or concomitant would also enable a system that prevents individuals from driving when they are not in an appropriate mental state. However, an outcome- or concomitant-based system would not respond exclusively to changes in safety-relevant states. It would also respond to changes in psychological states that are irrelevant to driving safety. For instance, the system might slow down the car because it erroneously interpreted the driver's physiological response to a joke made by a fellow passenger as indicating a high level of distraction. It is obvious that such a system would be of limited use. High-quality applications that use physiological measures to infer the user's state consequently require physiological markers or invariants.

Even if the application-related value of outcomes and concomitants is low, it is noteworthy that they enable sound scientific research. Under certain conditions, they allow researchers to test hypotheses and to compare psychological theories. If two or more theories differ regarding the predictions for one and the same physiological parameter, crucial experiments can be conducted to contrast the explanatory power of the theories (Cacioppo et al., 2000, Platt, 1964). Empirical observations that can be explained by only one of the competing theories provide evidence in favor of this theory. If one integrates assumptions about a psychophysiological relation with the predictions of a psychological theory so that the predicted variable becomes a physiological one, one can also use physiological measures to test the hypotheses of a single theory. The cardiovascular research on motivational intensity theory (Richter et al., 2016, for a review) constitutes an example for this approach. By suggesting that effort mobilization is associated with increased impact of the sympathetic nervous system on the heart, Wright (1996) transformed Brehm's motivational intensity theory (Brehm & Self, 1989) from a theory that predicted a psychological variable, effort, to one that predicted a physiological one, sympathetic activity on the heart. This enabled researchers to use cardiovascular measures influenced by sympathetic nervous system activity to test hypotheses about effort-related cardiovascular activity.

In sum, Cacioppo and colleagues' framework of psychophysiological relations highlight the requirements for being able to infer motivation-related states from physiological measures. Only after one has shown that a certain physiological measure changes as a function of the motivation-related state of interest and that it does not change as a function of other relevant psychological states, one can draw valid inferences from the measure to the state. Given that for most if not all psychophysiological measures evidence demonstrating that the physiological measure is related exclusively to a single motivation-related state is lacking, motivation scientists need to be careful when interpreting their physiological data. They have at best physiological outcomes or concomitants of the motivation-related states that they are interested in. This allows meaningful tests of psychological hypotheses and the comparison of competing theories but it does not warrant the conclusion that the physiological measure indicates the absence, presence, or magnitude of a motivation-related state. For instance, it is not (yet) warranted to infer from Corrugator supercilii amplitude to the level of effort investment, from skin conductance responses to emotional arousal, or from the amount of left frontal cortical activity to approach motivation.

2.2. Construct validation

Classical construct validation theory (e.g., Cronbach & Meehl, 1955; Kelley, 1927) provides a second approach to understanding the characteristics that physiological measures need to possess to enable inferences to motivation-related states. Most motivation-related states (like achievement motivation or effort) are theoretical constructs that cannot be observed directly, and physiological measures are used to collect information about these unobservable states. For instance, researchers working on the biopsychosocial model of challenge and threat (Blascovich & Tomaka, 1996) are interested in whether participants are in a challenge or threat state when performing tasks. Given that challenge and threat states cannot be observed directly, researchers have relied on measures of cardiac output and total peripheral resistance to acquire information about participants' state (e.g., Rith-Najarian et al., 2014). Other examples are the use of systolic blood pressure to assess effort in the frame of the research on motivational intensity theory (Richter et al., 2016, for an overview), heart rate variability to measure changes in mental engagement (e.g., Pendleton et al., 2016), skin conductance level to assess the impact of unfairness on emotional arousal (van 't Wout et al., 2006), and pre-ejection period as an indicator of the activity of the behavioral activation system (e.g., Derefinko et al., 2016).

The quality of the inferences that can be drawn from physiological measures depends on their validity, the question whether the measures actually assess the psychological states that they are supposed to measure (Cronbach & Meehl, 1955; Kelley, 1927). Most psychologists learned during their studies about measurement validity, threats to it, and how to demonstrate validity. They know about face validity, content validity, concurrent validity, predictive validity, convergent validity, and discriminant validity, and they are aware that each one of these validity subtypes provides a potential means to show the validity of a measure.

Face validity and content validity refer to the translation of a theoretical construct into a measure. For many physiological measures, it is difficult to demonstrate validity by using one of these two validity subtypes because of the complex nature of the measures or the lack of a precise definition of the motivation-related construct. For instance, it is difficult to show the face validity of a measure like heart rate variability for the assessment of mental engagement (e.g., Pendleton et al., 2016). It is not self-evident why the variation in the length of the interval between two consecutive heart beats should be indicative of mental engagement. It is also not obvious that high mental engagement translates into a low amount of variation in interbeat intervals. It is, of course, possible to construct a line of argument linking mental engagement, central parasympathetic activity, the interaction of the parasympathetic nervous system and the respiratory system, peripheral parasympathetic activity, and changes in interbeat intervals (e.g., Berntson et al., 1997) but this requires additional assumptions, explanations, and theorizing. It is not understandable at first glance whether heart rate variability is a valid measure of mental engagement.

It is also often difficult to refer to content validity to demonstrate that a physiological measure is valid. Physiological measures are often used to assess motivation-related constructs that are not well-defined. Given that content validity requires that one can evaluate whether the physiological measure captures all facets of the psychological construct, an ill-defined motivation-related concept prevents the demonstration of content validity. The psychophysiological literature on arousal illustrates this problem (see National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria, 2016, for an another discussion of the problematic status of the arousal concept). Due to a lack of agreement on the definition of arousal (e.g., Andrew, 1974; Blascovich, 1992; Neiss 1988), researchers used either very broad or very narrow definitions of arousal. Both types of definitions make the demonstration of content validity either very difficult or unnecessary. For example, Groeppel-Klein defined arousal as the “neurophysiological basis underlying all processes in the human organism” (Groeppel-Klein, 2005, p. 428). It is difficult to check the content validity of a physiological measure against such a broad definition. Bonnet and colleagues (1992) defined arousal as a specific abrupt shift in EEG activity. Such a very specific definition renders the demonstration of content validity unnecessary given that the physiological measure defines the construct.

Given that face validity and content validity often do not provide adequate means to assess the validity of physiological measures as indicators of motivation-related states, researchers might resort to the criterion-related subtypes of validity (Trochim, 2016). The demonstration of predictive, concurrent, convergent or discriminant validity always requires a criterion with which the physiological measure can be compared. Predictive validity requires the demonstration that the physiological measure predicts the effects that it should theoretically predict. For instance, drawing on the idea that arousal and performance should show an inverted-U shape-like relationship (e.g., Anderson, Revelle, & Lynch, 1989), one could demonstrate predictive validity by showing that a physiological measure of arousal successfully predicts the performance effect. If the measure indicated low or high arousal if performance is low and if the measure indicated moderate arousal if performance is high, one would have demonstrated the predictive validity of this measure.

To demonstrate concurrent validity of a physiological measure, one would have to demonstrate that the measure successfully differentiates between groups that are, for theoretical reasons, supposed to differ regarding the theoretical construct. For instance, the concurrent validity of pre-ejection period as an indicator of reward sensitivity (Brenner et al., 2005) could be demonstrated by showing that children with conduct problems and children without conduct problems differ in their pre-ejection period responses to reward (Beauchaine et al., 2007). For theoretical reasons, children with conduct problems are supposed to have decreased reward sensitivity and for this reason the demonstration that pre-ejection period responses differentiate between children with and without conduct problems suggests concurrent validity of pre-ejection period as a measure of reward sensitivity.

Convergent and discriminant validity can be established by demonstrating that a physiological measure is related (for convergent validity) or not related (for discriminant validity) to other measures. Convergent validity can be demonstrated by showing that the physiological measure is related to other measures of the same construct. Discriminant validity requires the demonstration that the physiological measure is not related to measures of conceptually different constructs. A researcher aiming at demonstrating that pre-ejection period is a valid measure of effort could aim at demonstrating that pre-ejection period is related to other measures of effort and unrelated to measures that measure other psychological constructs. However, the practical problem that the researcher would be facing is that there is neither an agreement on appropriate measures of effort, nor an agreement on a definition of effort. Some authors would consider self-reported effort a valid measure (e.g., Alhanbali et al., 2016; Meyer & Hallermann, 1977), whereas others would not (e.g., Gendolla & Richter, 2010). Depending on their subjective preferences, these authors would consider the demonstration that pre-ejection period and self-reported effort are correlated as a successful demonstration of convergent validity or as a failure to demonstrate discriminant validity.

The lack of agreement on a single definition of effort makes it also difficult to know how to deal with measures of related constructs. Given that it is unclear whether effort, cognitive control (e.g., Kuipers et al., in press) and mental workload (e.g., Brookhuis & de Waard, 2010), refer to the same or different motivation-related constructs, it is unclear whether one should aim at finding correlations between their measures or not. If they referred to the same motivation-related construct, one could use a high correlation between measures of effort, measures of cognitive control, and measures of mental workload to demonstrate convergent validity. However, if they referred to different constructs, one would be keen on finding no associations between these measures to demonstrate discriminant validity.

The lack of precise definitions of motivation-related constructs is mainly problematic for the demonstration of face validity, content validity, convergent validity, or discriminant validity. Without a specific definition, it is impossible to know which theoretical content needs to be covered by the physiological measure and which criterion has to be used to demonstrate that a physiological measure is a valid. The lack of a precise definition is less problematic for predictive and concurrent validity given that these types of validity rely on notions about effects that the psychological state should have according to a theoretical model or hypothesis.

2.3. Summary and conclusion

In the preceding sections, we discussed two approaches, Cacioppo and colleagues' framework and construct validation theory, that highlight the conditions for valid inference to motivation-related states from physiological measures. Both approaches are complementary and mainly differ in their focus. Classical construct validation focuses on the first step in Cacioppo and colleagues' framework, the question whether the physiological measure is associated with the motivation-related state of interest. Demonstrating that a physiological measure changes as a function of the motivation-related state corresponds to the demonstration of construct validity—the demonstration that the measure assesses what it is supposed to assess. Cacioppo and colleagues' framework focuses on the specificity of physiological measures, that is, on the number of psychological states that are associated with a certain physiological measure. Measures that are related to a single motivation-related state—markers and invariants—enable direct inferences from the physiological measure to the motivation-related state. Consequently, the crucial step in the validation of a physiological marker or invariant is the demonstration that the measure is not related to any other psychological state than the motivation-related state of interest.

Given the overlap between Cacioppo and colleagues' framework and construct validation, specific definitions of motivation-related constructs have the same importance in both frameworks. Without a specific definition, it is impossible to demonstrate content validity, convergent, or discriminant validity, and it is thus difficult to show that a certain physiological measure is a valid measure of a motivation-related state. Without a specific definition, it is also impossible to know if a change in the motivation-related state has occurred and it is consequently impossible to show that changes in the motivation-related state are associated with changes in a physiological measure.

Comparing the advantages and disadvantages of both approaches, researchers might be inclined to prefer the construct validation approach. Construct validation avoids the time consuming process of demonstrating that a physiological measure is only related to a single motivation-related state. Moreover, with the exception of predictive and concurrent validity, all subtypes of construct validity require only a definition of the motivation-related state. No theoretical model is needed. It would, of course, be desirable to have validation studies that draw on definitions that are shared by many researchers. However, this not required for construct validity. Researchers only need to formulate their own, specific definition of the motivation-related construct that they are interested in to be able to demonstrate face validity, content validity, convergent validity, or discriminant validity.

Construct validation also offers a solution to researchers who would like to avoid defining their motivation-related state of interest. Those researchers might aim at demonstrating the predictive and concurrent validity of their physiological measures. They would only need a theoretical model that offers predictions about determinants, outcomes, associated states, or group differences. Unfortunately, the demonstration of predictive and concurrent validity leads often to circular reasoning—the second pitfall that researchers interested in using physiological measures to obtain information about motivation-related states need to avoid.

3. Pitfall 2: Circular reasoning

Motivation scientists sometimes engage in circular reasoning when trying to justify the use of a certain physiological measure in their empirical work. Circular reasoning is often visible in formal validation studies but it also happens frequently when researchers present the general rationale of their work. Consider the example of a researcher interested in demonstrating that systolic blood pressure is a valid measure of effort. The researcher might decide to use motivational intensity theory (Brehm & Self, 1989) as a theoretical framework to provide evidence for the predictive validity of systolic blood pressure. According to motivational intensity theory, effort should increase with increasing task demand if the required effort is justified by the importance of task success. If the required effort exceeds the justified effort, individuals should disengage and effort should be low. Showing that systolic blood pressure changes as a function of task demand as predicted would establish the predictive validity of systolic blood pressure as a measure of effort.

If the researcher would now use systolic blood pressure response in a second study to test motivational intensity theory's hypothesis that task demand is a determinant of effort, she or he would likely engage in circular reasoning. To justify the use of systolic blood pressure as a measure of effort in the second study, the researcher would probably rely on the predictive validity of systolic blood pressure demonstrated in the first study. However, the demonstration of predictive validity in the first study relied on the assumption that effort changes as a function of task demand, which is questioned in the second study that aims at examining whether effort changes as a function of task demand. It is obvious that one can either question whether task demand is a determinant of effort using systolic blood pressure as an indicator of effort or use the hypothesized relationship between task demand and effort to demonstrate the predictive validity of systolic blood pressure as a measure of effort. Doing both at the same time leads to circular reasoning and makes the research meaningless.

The solution that has been suggested for this problem is a specific definition of the construct of interest that is independent of the hypotheses that one tests (e.g., Bechtoldt, 1959). The empirical research on motivational intensity theory that has employed cardiovascular measures constitutes an example for this approach. Drawing on physiological research on active coping (Obrist, 1981), Wright (1996) suggested that effort investment is associated with increased sympathetic influence on the heart. This definition of effort as myocardial sympathetic activity resulted in numerous empirical studies that examined motivational intensity theory's effort-related predictions using various measures of sympathetic influence on the heart (e.g., Richter & Gendolla, 2009; Richter & Knappe, 2014; Richter et al., 2012; see Richter et al., 2016, for a recent review). The use of Wright's notion as a definition of effort—and not as a hypothesis—allowed researchers to justify their physiological measure of effort independently from the predictions of the theory enabling them to avoid the problem of circular reasoning.

However, this solution to the problem of circular reasoning comes with a disadvantage. Linking a physiological measure (or parameter) by definition to a motivation-related state makes the motivation-related construct redundant (Clark, 1983). If effort is defined as myocardial sympathetic activity, both concepts can be used interchangeably. Given that there is no scientific reason for employing two concepts if one is sufficient, the label 'effort' could be dropped and researchers could directly refer to myocardial sympathetic activity. The only reason for using the label effort in this case might be that it attracts the interest of a broader audience if effort is used instead of myocardial sympathetic activity.

It is also noteworthy that this solution explicitly separates the definition of the motivation-related state from the empirical research on the phenomenon. Empirical research on a motivation-related state can only be evaluated in the light of the definition used by the researchers who conducted the research. Using a different definition of the same motivation-related state will automatically render the empirical research meaningless. If one disagrees that effort is reflected in myocardial sympathetic activity, the cardiovascular research on motivational intensity theory that has relied on this definition becomes meaningless. The observation that task demand has an impact on pre-ejection period (e.g., Richter, 2016)—an indicator of myocardial sympathetic activity—does not provide any information about effort if one disagrees that effort is associated with myocardial sympathetic activity.

It is obvious that an explicit presentation of the specific definition of the motivation-related state that underlies a specific piece of empirical research limits the potential reach of the work. For instance, without a presentation of the underlying definition of effort, a piece of effort-related research can be easily related to any other publications on effort. An explicit presentation of the definition, makes this much harder. Once effort has been defined as myocardial sympathetic activity, it is difficult to create links to publications that have drawn on other definitions of effort (for instance, effort as self-reported perception of effort investment, Marcora, 2009). However, given that science differs from pre-science by the fact that the phenomena of interest are clearly defined before empirical research on the phenomena is conducted (Kuhn, 1970), we think that it is desirable that researchers explicitly define their motivation-related constructs of interest before collecting data.

Clark (1983) suggested an alternative solution to the problem of circularity. He suggested that construct validation and testing theoretical predictions cannot be separated. According to his analysis, the validity of a measure depends on the support for the theory as a whole. The relation between the physiological measure and the motivation-related state is considered one of many hypotheses that make up the theory. Moreover, Clark suggested that the validity of the hypotheses included in a theoretical model cannot be evaluated independently. There is either support for the whole network of hypotheses that makes up the theory—including the relation between the motivation-related state and the physiological measure—or not.

There are two important implications of this approach. First, the validity of the physiological measure would be determined by the validity of the hypotheses included in the theory. If all the hypotheses of the model find empirical support, the measure is considered to be valid. The quality of myocardial sympathetic activity as a measure of effort would depend on the question whether one can show that myocardial sympathetic activity shows all the effects that motivational intensity theory predicts for effort investment. By demonstrating that myocardial sympathetic activity follows the patterns predicted by the theory for tasks with fixed, unfixed, and unclear task difficulty (e.g., Richter, 2013; Richter et al., 2016; Wright, 2008), one could demonstrate that myocardial sympathetic activity is a valid indicator of effort.

The second implication of Clark's approach is that the motivation-related state is completely defined by the theory's hypotheses. For instance, effort would be defined by the network of task demand and success importance effects on myocardial sympathetic activity predicted by motivational intensity theory. Increases in myocardial sympathetic activity as a function of increases in task demand under conditions of clear and fixed task difficulty as well as increases in myocardial sympathetic activity as a function of increases in success importance under conditions of unclear task difficulty would be considered to constitute effort. Observations or phenomena that are not covered by the theory's scope would not be considered to reflect effort. This second implication might constitute the main inconvenient of Clark's approach. Given that the motivation-related state would be completely defined by the theory's network of hypotheses, there would be no justification to apply the physiological measure as an indicator of the motivation-related state outside of the theory. The physiological measure and the motivation-related construct would only be meaningful within the scope of the theory.

4. Conclusions and recommendations

There is no single best solution to demonstrate that a certain physiological measure constitutes a valid measure of a motivation-related state. Each one of the discussed approaches offers advantages but also comes with disadvantages. The goal that researchers pursue determines which approach is the most useful for them. Cacioppo and colleagues' approach offers the best solution for researchers aiming at employing physiological measures to draw inferences about motivation-related states without being restricted by a theoretical framework. The successful application of this approach requires 1) a definition of the motivation-related state that enables the detection of the absence, presence, or magnitude of the motivation-related state, 2) the demonstration that the physiological measure changes as a function of the motivation-related state, and 3) the demonstration that it is not related to other relevant psychological states. Only if these three criteria are fulfilled, it is warranted to interpret changes in the physiological measure as reflecting changes in the motivation-related state.

The more economic construct validation approach constitutes the appropriate strategy for researchers mainly interested in testing theories. This approach enables researchers to waive the time-consuming demonstration that their physiological measure is not affected by other psychological states. It only requires 1) a clear and explicit definition of the motivation-related state and 2) the demonstration the physiological measure is related to the motivation-related state. Independent of the approach that researchers adopt, an explicit definition of the motivation-related state that is independent of the theoretical hypotheses that are empirically examined avoids the problem of circular reasoning. If researchers would like to avoid an explicit definition of the motivation-related state that they are interested in, Clark's approach might constitute a suitable framework. However, this solution limits the application of the physiological measure to the scope of the theory. Independent of the specific approach that the individual researcher favors, we hope that our article helps researchers to understand and avoid the major pitfalls associated with the use of physiological measures in motivation-related research.

References

Alhanbali, S., Dawes, P., Simon, L., & Munro, K. J. (2016). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear & Hearing*. doi:10.1097/AUD.0000000000000361

Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers, 16,* 857-878. doi:10.1016/j.intcom.2004.08.001

Anderson, K. J., Revelle, W. & Lynch, M. J. (1989). Caffeine, impulsivity, and memory scanning: A comparison of two explanations for the Yerkes-Dodson effect. *Motivation and Emotion, 13,* 1-20. doi:10.1007/BF00995541

Andrew, R. J. (1974). Arousal and the causation of behaviour. *Behaviour, 51,* 135-165.

Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., … van der Molen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 34*, 623-648. doi:10.1111/j.1469-8986.1997.tb02140.x

Blascovich, J. (1992). A biopsychosocial approach to arousal regulation. *Journal of Social and Clinical Psychology, 11,* 213-237.

Blascovich, J., & Tomaka, I. (1996). The biopsychosocial model of arousal regulation. *Advances in Experimental Social Psychology, 28,* 1-5. doi:10.1016/S0065-2601(08)60235-X

Bonnet, M., Carley, D., Carskadon, M., Easton, P., Guilleminault, C., … Jordan, B. (1992). EEG arousals: Scoring rules and examples. A preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorder Association. *Sleep*, *15*, 173-184.

Beauchaine, T. P. (2009). The role biomarkers and endophenotypes in prevention and treatment of psychopathological disorders. *Biomarkers in Medicine*, *1*, 1-3. doi:10.2217/17520363.3.1.1.x

Beauchaine, T. P., Gatzke-Kopp, L., & Mead, H. K. (2007). Polyvagal theory and developmental psychopathology: Emotion dysregulation and condcut problems from preschool to adolescence. *Biological Psychology, 74*, 174-184. doi:10.1016/j.biopsycho.2005.08.008

Bechtoldt, H. (1959). Construct validity: A critique. *American Psychologist, 14*, 619-629.

Brehm, J.W., & Self, E.A. (1989). The intensity of motivation. *Annual Review of Psychology, 40*, 109-131. doi:annurev.ps.40.020189.000545

Brenner, S. L., Beauchaine, T. P., & Sylvers, P. D. (2005). A comparison of psychophysiological and self-report measures of BAS and BIS activation. *Psychophysiology*, *42*, 108–115.

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, *42*, 898-903. doi:10.1016/j.aap.2009.06.001

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 173-191). New York, NY: Guilford.

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45*, 16-28.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). Psychophysiological science. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 3-26). New York, NY: Cambridge University Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Choi, J., Ahmed, B., & Gutierrez-Osuna, R. (2012). Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE Transactions on Information Technology in Biomedicine,* 16, 279-286. doi:10.1109/TITB.2011.2169804

Clark, A. (1983). Hypothetical constructs, circular reasoning, and criteria. *The Journal of Mind and Behavior, 4,* 1-12.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 32*, 281-302. doi:10.1037/h0040957

de Morree, H. M., & Marcora, S. M. (2010). The face of effort: Frowning muscle activity reflects effort during a physical task. *Biological Psychology, 85*, 377-382. doi:10.1016/j.biopsycho.2010.08.009

Derefinko, K. J., Eisenlohr-Moul, T. A., Peters, J. R., Roberts, W., Walsh, E. C., Milich, R., & Lynam, D. R. (2016). Physiological response to reward and extinction predicts alcohol, marijuana, and cigarette use two years later. *Drug and Alcohol Dependence, 163,* S29-S36. doi:10.1016/j.drugalcdep.2016.01.034

Edelberg, R., & Wright, D. J. (1964). Two GSR effector organs and their stimulus specificity. *Psychophysiology, 1*, 39-47. doi:10.1111/j.1469-8986.1964.tb02619.x

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers, 21*, 133-145. doi:10.1016/j.intcom.2008.10.011

Groeppel-Klein, A. (2005). Arousal and consumer in-store behavior. *Brain Research Bulletin, 67*, 428-437. doi:10.1016/j.brainresbull.2005.06.012

Gendolla, G. H. E., & Richter, M. (2010). Effort mobilization when the self is involved: Some lessons from the cardiovascular system. *Review of General Psychology, 14,* 212-226. doi:10.1037/a0019742

Gergelyfi, M., Jacob, B., Olivier, E., & Zénon, A. (2015). Dissociation between mental fatigue and motivational state during prolonged mental activity. *Frontiers in Behavioral Neuroscience, 9*, 176. doi:10.3389/fnbeh.2015.00176

Harmon-Jones, E. (2003). Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology, 40,* 838-848. doi:10.1111/1469-8986.00121

Jones, M. V., Meijen, C., McCarthy, P. J., & Sheffield, D. (2009). A theory of challenge and threat states in athletes. *International Review of Sport and Exercise Psychology, 2,* 161-180. doi:10.1080/17509840902829331

Kelley, T. L. (1927). *Interpretation of educational measurements.* Yonkers-on-Hudson, NY: World Book.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: The University of Chicago Press.

Kuipers, M., Richter, M., Scheepers, D., Immink, M. A., Sjak-Shie, E., & van Steenbergen, H. (in press). How effortful is cognitive control? Insights from a novel method measuring single-trial evoked beta-adrenergic cardiac reactivity. *International Journal of Psychophysiology.*

Lenzenweger, M. F. (2013a). Endophenotype, intermediate phenotype, biomarker: Definitions, concept comparisons, clarifications. *Depression and Anxiety, 30,* 185-189. doi:10.1002/da.22042

Lenzenweger, M. F. (2013b). Thinking clearly about the endophenotype-intermediate phenotype-biomarker distinctions in developmental psychopathology research. *Development and Psychopathology, 25*, 1347-1357. doi:10.1017/S0954579413000655

Marcora, S. (2009). Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs. *Journal of Applied Physiology, 106*, 2060-2062. doi:10.1152/japplphysiol.90378.2008

Meyer, W.-U., & Hallermann, B. (1977). Intended effort and informational value of task outcome. *Archive für Psychologie, 129*, 131-140.

National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (2016). *Behavioral assessment methods for RDoC constructs*. Bethesda, MD: National Institute of Mental Health.

Neiss, R. (1988). Reconceptualizing arousal: Psychobiological states in motor performance. *Psychological Bulletin, 103*, 345-366.

Obrist, P. A. (1981). *Cardiovascular psychophysiology.* New York, NY: Plenum.

Padoa-Schioppia, C., & Assad, J. (2006). Neurons in orbitofrontal cortex encode economic value. *Nature, 441*, 223-226. doi:10.1038/nature04676

Pendleton, D. M., Sakalik, M. L., Moore, M. L., & Tomporowski, P. D. (2016). Mental engagement during cognitive and psychomotor tasks: Effects of task type, processing demands, and practice. *International Journal of Psychophysiology, 109*, 124-131. doi:10.1016/j.ijpsycho.2016.08.012

Platt, J. R. (1964). Strong inference. *Science, 146*, 347-353.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron, 72*, 692-697. doi:0.1016/j.neuron.2011.11.001

Puntmann, V. O. (2009). How-to guide on biomarkers: Biomarker definitions, validation and applications with examples from cardiovascular disease. *Postgraduate Medical Journal, 85*, 538-545. doi:10.1136/pgmj.2008.073759

Richter, M. (2013). A closer look into the multi-layer structure of motivational intensity theory. *Social and Personality Psychology Compass, 7*, 1–12. doi:10.1111/spc3.12007

Richter, M. (2016). The moderating effect of success importance on the relationship between listening demand and listening effort. *Ear and Hearing, 37*, 111S–117S. doi:10.1097/AUD.0000000000000295

Richter, M., & Gendolla, G. H. E. (2009). Mood impact on cardiovascular reactivity when task difficulty is unclear. *Motivation and Emotion, 33*, 239–248. doi:10.1007/s11031-009-9134-4

Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know. In A. J. Elliot (Ed.), *Advances in motivation science* (pp. 149-186). Cambridge, MA: Academic Press. doi:10.1016/bs.adms.2016.02.001

Richter, M., Baeriswyl, E., & Roets, A. (2012). Personality effects on cardiovascular reactivity: Need for closure moderates the impact of task difficulty on engagement-related myocardial beta-adrenergic activity. *Psychophysiology, 49*, 704–707. doi:10.1111/j.1469-8986.2011.01350.x

Richter, M., & Knappe, K. (2014). Mood impact on effort-related cardiovascular reactivity depends on task context: Evidence from a task with an unfixed performance standard. International Journal of Psychophysiology, 93, 227-234. doi:10.1016/j.ijpsycho.2014.05.002

Rith-Najarian, L. R., McLaughlin, K. A., Sheridan, M. A., & Nock, M. K. (2014). The biopsychosocial model of stress in adolescence: Self-awareness of performance versus stress reactivity. *Stress, 17*, 193-203. doi:10.3109/10253890.2014.891102

Sideridis, G. D., Kaplan, A., Papadopoulos, C., & Anastasiadis, V. (2014). The affective experience of normative-performance and ouctome goal pursuit: Physiological, observed, and self-report indicators. *Learning and Individual Differences, 32*, 114-123. doi:10.1016/j.lindif.2014.03.006

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology, 5,* 1-25. doi:10.1146/annurev.clinpsy.032408.153639

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., & Okubo, Y. (2009). When your gain is my pain and your pain is my gain: Neural correlates of envy and schadenfreude. *Science, 323*, 937–939. doi:10.1126/science.1165604

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology, 73*, 63-72.

Trochim, W. (2016). *Research methods: The essential knowledge base* (2nd ed.). Cincinnati, OH: Atomic Dog Publishing.

van ’t Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research, 169*, 564-568. doi:10.1007/s00221-006-0346-57

Weinberg, A., Riesel, A., & Hajcak, G. (2012). Integrating multiple perspectives on error-related brain activity: The ERN as a neural indicator of trait defensive reactivity.

Motivation and Emotion, *36*, 84-100. doi:10.1007/s11031-011-9269-y

Wright, R. A. (1996). Brehm's theory of motivation as a model of effort and cardiovascular response. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 424–453). New York: Guilford.

Wright, R. A. (2008). Refining the prediction of effort: Brehm's distinction between potential motivation and motivation intensity. S*ocial and Personality Psychology Compass, 2*, 682–701. doi: 10.1111/j.1751-9004.2008.00093.x.

Table 1

Psychophysiological relations, possible inferences, and validation

	Outcome	Concomitant	Marker	Invariant
Psychophysiological relation	Context-limited one-to-many relationship	General one-to-many relationship	Context-limited one-to-one relationship	General one-to-one relationship
Application				
Test of hypotheses	Possible	Possible	Possible	Possible
Comparison of theories	Possible	Possible	Possible	Possible
Inferences to changes in the psychological state	Not possible	Not possible	Possible	Possible
Validation				
Demonstration that the psychological state of interest influences the psychological measure	Required	Required	Required	Required
Demonstration that other psychological states do not influence the psychological measure	Not required	Not required	Required	Required
Demonstration that the relationship is context-independent	Not required	Required	Not required	Required