RUNNING HEAD: IMPLICIT MEASURES

Implicit Measures:

Similarities and Differences

Jan De Houwer

and

Agnes Moors

Ghent University, Belgium

mailing address: Jan De Houwer
Department of Psychology
Ghent University
Henri Dunantlaan 2
B-9000 Ghent
Belgium
Email: Jan.DeHouwer@UGent.be
Tel: ++32 9 264 64 62
Fax: ++32 9 264 64 89

During the past 15 years, an increasingly large number of procedures and effects have been referred to as “implicit measures”. These include affective priming (e.g., Fazio, Jackson, Dunton, & Williams, 1995), the name letter effect (e.g., Koole, Dijksterhuis, & van Knippenberg, 2001; Nuttin, 1985), semantic priming (Wittenbrink, Judd, & Park, 1997), the implicit association test (IAT; Greenwald, McGhee, & Schwarz, 1998), the affective Simon effect (De Houwer & Eelen, 1998), the Go-NoGo Association Test (GNAT; Nosek & Banaji, 2001), stereotypic explanatory bias (Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003), the single-target IAT (Wigboldus, 2001; Karpinski & Steinman, 2006), the extrinsic affective Simon (EAST) effect (De Houwer, 2003a), the stimulus response compatibility task (Mogg, Bradley, Field, & De Houwer, 2003), the implicit association procedure (Schnabel, Banse, & Asendorpf, 2006), the affect misattribution procedure (Payne, Cheng, Govorun, & Stewart, 2005), the single association test (Blanton, Jaccard, Gonzales, & Christie, 2006), the word association test (Stacy, Ames, & Grenard, 2007), the approach-avoid task (Rinck & Becker, 2007), the implicit relational assessment procedure (e.g., Barnes-Holmes, Murtagh, & Barnes-Holmes, in press), the sorting paired features task (Bar-Anan, Nosek, & Vianello, 2009), and the brief IAT (Sriram, & Greenwald, in press). Several of these implicit measures are described in other chapters of this book. At present, so many implicit measures are available that it becomes difficult to come to grips with why a particular procedure or effect qualifies as an implicit measure and how it differs from other (implicit) measures.

Recently, we have provided a detailed analysis of what it means to say that something is an implicit measure (De Houwer, 2006; De Houwer & Moors, 2007; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009a). In the first part of the present chapter, we briefly summarize the results of this analysis. In the second and main part of this chapter, we focus on classification criteria that could be used to clarify how different implicit measures differ from each other (also see De Houwer, 2009). Making these criteria explicit not only has a heuristic value for organizing different implicit measures but might also help us to improve the measures themselves and the way in which they are applied. In the third and final part, we propose a number of meta-criteria that can be used to evaluate the merits of classification criteria.

What are implicit measures?

In line with the ideas of Borsboom, Mellenbergh, and van Heerden (2004, p. 1061), and De Houwer (2006; De Houwer & Moors, 2007), De Houwer et al. (2009a) defined an implicit measure as the outcome of a measurement procedure that is causally produced by psychological attributes in an automatic manner. Figure 1 provides a graphical representation of this definition. The definition has several important implications (see De Houwer et al., 2009a, for a more in-depth discussion). First, a measure is the outcome of a measurement procedure that is applied to a certain person. For instance, a racial IAT score (i.e., the difference in performance during the Black-positive task and the White-positive task of a racial IAT) is derived from how a person responds when submitted to the racial IAT task. The outcome is assumed to be a measure of the attribute (e.g., racial attitudes) in that variations in the outcome are supposed to capture differences in this attribute (top arrow in Figure 1). The measurement procedure, on the other hand, is simply an objective list of instructions on how to obtain a measurement outcome. It specifies, amongst other things, the way in which stimuli should be presented, responses should be registered, and the outcome should be derived from the responses.

Second, the definition clarifies that research on the validity of a measure should focus on obtaining evidence for the causal relation between the to-be-measured attribute and the measure. Ideally, this research should not only reveal that variations in the attribute cause variations in the measurement outcome but should also uncover the processes by which the attribute has this causal effect (also see also Wentura & Rothermund, 2007). As Borsboom et al. (2004, p. 1067) point out, in exact sciences, “nobody starts constructing measurement instruments without the faintest idea of the processes that lead to measurement outcomes”. Although correlational research can provide useful information for detecting possible causal relations, experimental research generally offers more guarantees for the validity of causal inferences. Hence, experimental studies should be an essential part of validation research.

Finally, whether a measure is implicit is determined by whether the processes by which the to-be-measured attribute causes the measurement outcome, are automatic in a certain manner (see Figure 1). In line with a decompositional view of automaticity (Bargh, 1992; Moors & De Houwer, 2006; see Moors, Spruyt, & De Houwer, this volume), De Houwer et al. (2009a) assumed that processes can have different features of automaticity that do not always co-occur. For instance, some processes are automatic in the sense that they still operate even when participants do not have particular goals (e.g., the goal to engage in the process). Other processes are automatic in the sense that they operate even when participants are unaware of the stimulus that instigates the process. Different implicit measures can thus be implicit (i.e., automatic) in different ways. It therefore makes little sense to simply say that a measure is implicit. It is always necessary to specify the automaticity features that characterize the (processes underlying the) measure (see De Houwer, 2006, and De Houwer & Moors, 2007, for a detailed analysis; also see Moors et al., this volume).

Like all definitions, the definition of implicit measures that is provided by De Houwer et al. (2009a) is a matter of convention and thus to a certain degree arbitrary. Nevertheless, the definition does have a number of advantages. First, it has a strong conceptual basis in the work of Borsboom et al. (2004; Borsboom, 2006) and De Houwer (2006; De Houwer & Moors, 2007). Second, it is broad enough to allow for a variety of opinions about the nature of the attributes that are captured by implicit measures, the nature of the processes that underlie implicit measures, and the way in which measures are implicit. Third, it is detailed enough to clarify the abstract core of what all implicit measures have in common and to specify the properties that an ideal implicit measure should have. With regard to the latter point, De Houwer et al. specified three normative criteria: (1) The what criterion: It should be clear which attributes causally produce the measurement outcome. (2) The how criterion: It should be clear by which processes the attributes cause the measurement outcome. (3) The implicitness criterion: The way in which the processes underlying a measure are automatic needs to be specified and demonstrated empirically. Whereas De Houwer et al. focused on what it is that implicit measures (should) have in common, in the remainder of this chapter, we focus on how implicit measures can differ from each other. Of course, implicit measures differ with regard to many superficial characteristics. The challenge is not to find differences as such but to find patterns in all these differences, that is, commonalities in how measures can differ. We will use the definition of implicit measures as a guideline to find such commonalities.

Criteria to differentiate implicit measures

The definition of implicit measures as provided by De Houwer et al. (2009a) refers to elements internal to the person who is being measured (what is inside the box in Figure 1, i.e., mental attributes and processes) and elements external to the individual (what is outside the box in Figure 1, i.e., the measurement procedure). It leaves many degrees of freedom with regard to the precise properties of these internal and external elements. The outcome of the measurement procedure needs to be influenced by attributes of the individual and the processes by which this occurs need to be automatic in some sense, but apart from that, implicit measures can involve any type of measurement procedure, attribute, or process. Implicit measures can thus be compared with regard to the properties of these internal (attribute and process) and external (measurement procedure) elements. This insight provides the first step toward a heuristic system for classifying implicit measures. The second step involves making explicit how implicit measures can differ with regard to internal and external properties. This involves the specification of criteria that some implicit measures meet but others do not. In the following paragraphs, we propose a number of criteria that are related to the internal and external properties of implicit measures. An overview of these criteria can be found in Table 1.

Differences with regard to internal properties

Based on their definition of implicit measures, De Houwer et al. (2009a) formulated three normative criteria that an ideal implicit measure should meet: (1) It should be clear which attributes the measure reflects (the what criterion). (2) The nature of the processes by which the attributes cause variations in the measure should be known (the how criterion). (3) It should be clear that the underlying processes are automatic in a certain manner (the implicitness criterion). These criteria not only set the ultimate standards against which each implicit measure can be compared, they also clarify the manner in which internal properties of the measures can differ. Implicit measures can differ with regard to (a) the attributes that cause the measure (e.g., attitudes, stereotypes), (b) the nature of the processes by which the attributes cause the measure (e.g., spreading of activation, response competition), and (c) the way in which these processes are automatic (e.g., unintentional, unconscious). As such, the what, how, and implicitness criteria allow for a classification of implicit measures. We now discuss each of these three criteria in more detail.

The what criterion. Classifying measures on the basis of what they are supposed to measure is perhaps the most common and intuitively appealing manner of classification. Many researchers and practitioners are inclined to characterize measures on the basis of their face validity, that is, on whether the measurement procedure looks suitable for capturing a certain attribute. One should be aware, however, that classification of a measure on the basis of the what criterion cannot be done on the basis of superficial, objective features of the measurement procedure. For instance, saying that racial IAT scores measure racial attitudes is a theoretical claim rather than something that can be verified by looking at the measurement procedure or the measurement outcome. Validity is not a property of the measurement procedure or the measurement outcome, it is a property of the claims about the measurement outcome (i.e., about what the outcome captures). Like all theoretical claims, claims about the validity of a measure need to be backed up by basic research (see De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009b). A measure can be considered as valid for measuring an attribute only when there is sufficient experimental and correlation evidence to support the conclusion that a measure indeed captures the to-be-measured attribute (see Borsboom et al., 2004; De Houwer et al., 2009a, 2009b). Hence, classification on the basis of the what criterion depends on the outcome of empirical research rather than on an inspection of the properties of the measurement procedure. As we noted above, conceptual analysis is also important. It makes little sense to characterize and classify a measure on the basis of the to-be-measured attribute if it is not clear what the attribute actually entails.

The risks of relying on face validity for deciding what a measure captures are well known (Anastasi, 1988). Nevertheless, even in recent studies on implicit measures, researchers sometimes rely exclusively on superficial properties of the measurement procedure when interpreting a measure. For instance, several implicit measures have been proposed in which participants are asked to approach or avoid stimuli in some way (e.g., move a manikin or a joystick toward or away from a stimulus, see Mogg et al., 2003; Rinck & Becker, 2007). Because the measure is derived from the manner in which stimuli influence approach and avoid responses, researchers tend to interpret the measures as indices of the motivational properties of stimuli. There is, however, little if any direct evidence to support this interpretation. It might well be that all the effects are driven exclusively by the valence of the stimuli independent of their motivational properties (De Houwer, 2003b, pp. 236-237; see Eder & Rothermund, 2008, for evidence supporting this hypothesis). At the conceptual level, it is also not entirely clear what is unique about the motivational properties of a stimulus and thus how one should assess these properties. Until more empirical and conceptual work on this matter has been completed, researchers should be very careful in interpreting approach-avoid measures as indices of motivational properties.

Even when there are good conceptual and empirical arguments to classify measures on the basis of the attribute that they reflect, one should not conclude that measures that are supposed to capture the same attribute always produce the same results. Different measures of the same attribute can, for instance, differ with regard to the impact of other attributes, the way in which the to-be-measured attribute influences the measure and the conditions under which it does so. There are thus plenty of potential reasons for why measures that are assumed to capture the same attribute often do not correlate (e.g., Bosson, Swann, & Pennebaker, 2000).

The how criterion. Measures can be classified also on the basis of the processes by which the to-be-measured attribute causes the measure. A prototypical example of a psychological process is spreading of activation (e.g., Collins & Loftus, 1975). It is assumed to operate on concepts represented as nodes in a semantic network. In this network, the representations of semantically similar concepts are connected by associations through which activation can spread. Activation of a concept that is due to the presentation of one specific stimulus can thus spread to other, semantically related concepts. This could facilitate the subsequent (semantic) processing of stimuli related to those concepts. It is beyond the scope of this manuscript to give an overview of all the possible processes that could operate in the various implicit measures that have been proposed so far (see De Houwer et al., 2009a, for a recent review).

We do want to emphasize that the how criterion is likely to be an important criterion for a number of reasons. First, knowing the processes underlying a measure can provide important information about the validity of a measure. As Borsboom et al. (2004) noted, it is important that we have a good understanding of the processes that underlie a measure because this provides important information about the validity of a measure: If we know how a psychological attribute produces measurement outcomes, this provides more certainty about the fact that the attribute causes the measurement outcome. It could also provide hints for optimizing the validity of the measure. Once a process has been identified through which the to-be-measured attribute can influence the measure, steps can be taken to strengthen the process and thus the impact of the attribute on the measures.

Second, measures that are based on different processes are also likely to produce different results. Hence, applying the how criterion could provide an important insight into why different implicit measures of the same psychological attribute are often uncorrelated (e.g., Bosson et al., 2000).

Third, merely acknowledging that implicit measures can differ with regard to the nature of the processes on which they are based is important for interpreting the results of (implicit) measures. It clarifies that all measures provide only an indirect reflection of psychological attributes. As depicted in Figure 1, a psychological attribute of the person can influence behavior only by virtue of certain processes (represented by the arrows inside the box in Figure 1). Many factors other than the to-be-measured psychological attribute can influence the responses from which the measure is derived. Because of this, it is difficult to be certain about the interpretation of a particular measurement outcome. When a measure provides evidence for the presence of a psychological attribute, this could be due to the effect of other (correlated) attributes. For instance, it has been argued that IAT effects do not reflect associations in memory (i.e., attitudes and stereotypes) but differences in the salience of concepts (e.g., Rothermund & Wentura, 2004). When a measure does not provide evidence for the presence of a psychological attribute, this could be due to other attributes or processes that counter the impact of the to-be-measured attribute on the measurement outcome. Before we can have certainty about how to interpret measures, it is thus vital that we learn more about the underlying processes. Until we have this knowledge, we should always keep in mind that a measure is the product not only of the to-be-measured psychological attribute but also of the processes intervening between the attribute and the behavior from which the outcome is derived (see Moors et al., this volume, and Gawronski, Deutsch, LeBel, & Peters, 2008, for an in-depth discussion of this point).

The implicitness criterion. In our previous work (De Houwer, 2006; De Houwer & Moors, 2007; Moors & De Houwer, 2006), we equated the concept “implicit” with the concept “automatic” and defined the latter according to a decompositional point of view (e.g., Bargh, 1992). From this perspective (see Moors et al., this volume), automaticity is not an all-or-none property that processes either possess completely or not at all. Rather, it is an umbrella concept that refers to a variety of automaticity features that do not necessarily co-occur. Each automaticity feature concerns a certain condition on which the operation of the process might depend. Many features such as the features uncontrolled, unintentional, goal independent, autonomous, and purely stimulus driven refer to the fact that the process can operate in the absence of certain goals. For instance, a process can be called unintentional if it can operate even when participants do not have the goal to engage in this process. Other automaticity features refer to the need of awareness (of the instigating stimulus, the process itself, or the output of the process), processing resources, and time. Features do not always co-occur. Some processes might be automatic in the sense that they operate in the absence of the goal to engage in the process at hand but nonautomatic in that they still require processing resources and awareness of the instigating stimulus. Other processes might be automatic in that they do not require substantial processing resources but nonautomatic in that they operate only when participants have the goal to engage in the process at hand.

Because processes can be automatic in different ways, it does not make sense to merely say that a process is automatic. Rather, one always needs to specify the way in which a process is automatic, that is, which automaticity features do and do not apply to the process. One can also not simply infer the presence of one automaticity feature from the presence of another automaticity feature. Instead, each feature needs to be examined separately. Based on this empirical work, measures can be classified according to the automaticity features of the processes on which they are based. We will refer to this criterion of classification as the implicitness criterion.

Given that there are so many automaticity features, the number of possible combinations of features is very large. Also, many processes will have one or more automaticity feature and will thus be automatic in some way and nonautomatic in another way. One might wonder how useful it is to classify measures on the basis of the implicitness criterion if there is a different box in the classification system for almost each different measure. We believe that there are at least four reasons why the implicitness criterion nevertheless might prove to be useful.

First, at the very least, it offers a vocabulary for describing implicit measures. That is, it allows one to make explicit and examine similarities and differences among measures.

Second, the same vocabulary can be used to describe behaviors in daily life. In fact, one could say that both the measurement outcome and the real-life behavior have certain automaticity features (i.e., the to-be-measured attribute influences behavior under certain conditions). It is possible that the predictive value of the measure depends on the extent to which the automaticity features of the processes underlying that measure are the same as the automaticity features of the processes underlying the behavior that one wants to predict. For instance, real-life behavior that is influenced by attitudes when people do not have the time to evaluate stimuli in the environment (e.g., buying products in a supermarket under time pressure) might be related most strongly to measurement outcomes that are observed when participants are under time pressure. The closer the overlap between the conditions under which the measurement outcome is observed and the conditions under which to-be-predicted behavior occurs, the more the measurement outcome might be able to predict the behavior (also see Vargas, 2004; De Houwer, 2006, 2009).

Third, the fact that there are many different possible combinations of automaticity features does not exclude the possibility that only a few features or combinations of features will turn out to be important for the added value of an implicit measure. The main reason for the introduction of implicit measures was that they might explain variance of to-be-predicted behaviors on top of what can be predicted on the basis of traditional, explicit measures. Future research might show that a measure has incremental predictive validity only when (apart from being reliable and valid) it possesses some automaticity features (e.g., uncontrolled) but not others (e.g., unaware). The only way to examine this issue is by making explicit for each measure what automaticity features it possesses.

Finally, the automaticity features of a measure specify the conditions under which the processes underlying the measure operate. If it is true that different types of processes operate under different conditions, knowledge about these conditions provides information about the type of processes that are at stake. Applying the implicitness criterion can increase understanding of the processes underlying a measure and thus help construct the measurement theory that is necessary to establish the validity of a measure (Borsboom et al., 2004; De Houwer et al., 2009a).

Differences with regard to external properties.

The external properties of a measure are related to the measurement procedure from which the measurement outcome is derived. As we pointed out earlier, a measurement procedure specifies which stimuli must be presented in what manner, which responses should be registered in what manner, and how a measurement outcome must be derived from those responses. This definition of a measurement procedure clarifies two important points. First, a distinction needs to be made between the responses that are observed (i.e., the raw data) and the outcome that is derived from these responses (see Figure 1). Responses are multi-featured entities. For instance, pressing a key during an IAT can be characterized by the time at which it was executed (i.e., reaction time) but also by its accuracy or even the pressure with which it was exercised. In principle, an outcome (e.g., IAT score) can be derived from each response feature or each combination of features. Moreover, several algorithms or statistical techniques can be used to derive the measurement outcome from the raw response data (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Greenwald, Nosek, & Banaji, 2003). How the outcome is derived from observed responses is a crucial and integral part of the measurement procedure. All of the external classification criteria that we discuss in this section deal with differences in the features of the responses that are used to derive the measurement outcome. Second, as De Houwer (2006) pointed out, there is nothing implicit about a measurement procedure. Hence, it does not make sense to use the term “implicit measure” to refer to a measurement procedure. Only measurement outcomes can qualify as implicit measures. Measurement outcomes and thus implicit measures can, however, be characterized and compared with regard to the nature of the measurement procedure that is used to obtain the outcome. We now discuss a number of possible criteria that can be used to classify measures on the basis of the measurement procedure. An overview of these criteria can be found in Table 1. The relations between the various external criteria are depicted in Figure 2.

The self-assessment criterion: Direct versus indirect measures. A first criterion relates to whether the measurement procedure requires the participant to self-assess the to-be-measured attribute. Many measures are based on such a measurement procedure. For instance, one can measure racial attitudes by asking participants to report the degree to which they hold a positive or negative attitude toward White and Black people. Such self-assessments are often registered using Likert-scales on which the participant selects a numerical value that expresses his/her assessment of the degree to which he/she possesses the attribute. We refer to these measures as direct measures. More formally, direct measures are characterized by two properties: (1) The measurement outcome is derived from a self-assessment by the participant. (2) The target of the self-assessment is the attribute that the measurement outcome is assumed to capture. If a measure does not have both of these properties, it can be called indirect.

We have chosen the qualifications “direct” and “indirect” because they convey information about the nature of the relation between the measurement outcome and the responses on which it is based. In direct measurement procedures, the outcome is based directly on the responses of the participant (i.e., his or her self-assessment of the to-be-measured attribute) without the need for further interpretation of the responses. In indirect measurement procedures, on the other hand, there is an additional step of interpretation by the researcher. The outcome is based not on the responses as such but on how the responses are interpreted by the researchers. [1]

Take the example of a measurement procedure in which participants are asked to self-assess for each letter of the alphabet how much they like or dislike that letter. When these self-assessments are used to construct a measurement outcome that aims to assess attitudes towards letters, the outcome can be described as a direct measure of attitudes towards letters. In this case, the measurement outcome is derived directly from self-assessments of the to-be-measured attributes (i.e., attitudes toward letters). The same self-assessments of attitudes toward letters can, however, also be used to indirectly measure self-esteem. Typically, this has been done by calculating name letter effects, that is, the difference between the mean evaluation of letters that are part of the name of the participant and the mean evaluation that these letters receive from participants who do not have those letters in their name (see Koole et al., 2001; Nuttin, 1985). Such name letter effects are indirect measures of self-esteem because they are not based directly on a self-assessment of self-esteem but are inferred indirectly by the researcher from another behavior, in this case, the self-assessment of attitudes toward letters.

The example of the name letter ratings also clarifies the difference between direct measures and self-report measures. Whereas a direct measure is based on a self-assessement of the to-be-measured attribute (i.e., the specific attribute that the outcome is supposed to capture), a self-report measure is based on a self-assessment of any kind of attribute, the to-be-measured one or another one. Direct measures are thus a subclass of self-report measures (also see Figure 2). This analysis clarifies that self-report measures can also be indirect measures. De Houwer (2006) illustrated this point by referring to the Minnesota Multiphasic Personality Inventory (MMPI) which is a widely used questionnaire that was designed to measure various personality traits (e.g., Butcher, Derksen, Sloore, & Sirigatti, 2003). The procedure of administrating the MMPI does not involve asking participants to answer questions about the extent to which they believe that they possess a certain personality trait. Rather they are asked to indicate whether statements about feelings and behaviors apply to them. For instance, people who endorse the item "I have a good appetite" will receive a lower score on the depression scale. Although the MMPI is a self-report measure (because participants are asked to self-assess attributes), it is an indirect measure of personality.

The self-assessment criterion does not overlap with the internal classification criteria that we discussed above. Regarding the what criterion, many if not all attributes can be measured directly or indirectly. Hence, knowing whether a measure is direct or indirect provides little or no information about what attribute is being measured. Regarding the how criterion, it is likely that all direct measures are at least in part based on similar processes because all involve the common goal of self-assessing a psychological attribute. Nevertheless, direct measures can differ in the conditions under which the self-assessment is registered and can thus differ with regard to some of the underlying processes. For instance, self-assessment under time pressure is likely to involve different processes than self-assessment without time pressure. Hence, simply classifying a measure as direct does not allow one to conclude how the attribute causes the measurement outcome. This is even more true for indirect measures. Indirect measures can be derived from self-assessments of attributes other than the to-be-measured attribute as well as from a variety of other responses and response features. Classifying a measure as indirect therefore provides little if any information about the underlying processes.

Finally, the self-assessment criterion does not overlap with the implicitness criterion. A first issue to emphasize is that the distinction between direct and indirect measures differs from the distinction between explicit and implicit measures. Whereas the qualification "direct" or "indirect" refers to a property of the measurement procedure on which the measure is based, the terms "explicit" and "implicit" describe the conditions under which the psychological processes underlying the measure operate. This implies that direct measures are not necessarily explicit and that indirect measures are not necessarily implicit. Whether a measure is implicit needs to be examined empirically (see De Houwer, 2006; De Houwer & Moors, 2007; Moors et al., this volume).

It is true, however, that indirect measures are more likely to be implicit in certain ways than direct measures. Direct measurement procedures by definition include the instruction to adopt the conscious goal to self-assess the to-be-measured attribute in a certain manner. These instructions do not ensure that participants will actually adopt the goal or that the goal will have any impact on performance, but they do increase the likelihood. Indirect measurement procedures do not encourage the adoption of a conscious goal to express attributes, rendering it less likely that indirect measures depend on such a goal. Asking participants to adopt the goal to self-assess the to-be-measured attribute also renders it highly likely that participants will become aware of the crucial stimuli, the attribute that is being measured, and the fact that the attribute influences performance. Hence, it is unlikely that direct measures will be implicit in the sense of unaware. It is more likely that indirect measures will be implicit in the sense of unaware but this should be examined empirically.

Whereas the procedural distinction between direct and indirect measures is relevant for whether a measure is likely to be implicit in the sense of dependent on goals and awareness, it is less relevant for whether a measure is likely to be implicit in the sense of minimally dependent on resources (i.e., efficient) and time (i.e., fast). Direct measures can be implicit in that expression of the attribute occurs quickly (e.g., self-assessment under time pressure) or independently of the presence of considerable cognitive resources (e.g., self-assessments while performing a difficult secondary task). Hence, it is important to realize that direct measurement procedures can generate outcomes that are implicit in some sense of the word (see Ranganath, Smith, & Nosek, 2008, for evidence supporting this hypothesis).

The response system criterion: Behavioral, physiological, and neurological measures. In principle, measurement outcomes can be derived from any type of response. Typically, a distinction is made between (a) behavioral responses (e.g., spoken or written answers, keypress responses), (b) physiological responses (e.g., skin conductance), (c) and neurological responses (e.g., brain activity as registered by EEG and fMRI). Both currently available and yet-to-be-introduced measures can be classified by examining whether the responses from which the measurement outcome is derived belong to the behavioral, physiological, or neurological response system.

As is the case with the self-assessment criterion, it is important to realize that the response system criterion refers to a more or less objective external property, namely a feature of the measurement procedure that is used to arrive at a measurement outcome. As such, it clearly differs from and does not overlap with the internal criteria that we discussed earlier. First, many if not all attributes can be measured on the basis of behavioral, physiological, and neurological responses. Hence, classifying a measure on the basis of the response system criterion does not allow for a classification on the basis of the what criterion.

Second, although some processes are likely to be unique to certain response systems, knowing the type of response from which the measurement outcome is derived says little about how the attribute causes the outcome. Hence, the response system criterion does not overlap with the how criterion.

The response system criterion also does not overlap with the implicitness criterion. One might be tempted to believe that measures derived from physiological and neurological responses are by definition fully implicit. Such a belief could result from the fact that (some) physiological and neurological responses appear to arise independently of goals, awareness, substantial processing resources, or substantial time. This line of reasoning is not entirely valid. Most importantly, the implicitness of measures does not depend only on the automaticity features of the responses themselves. What also matters are the automaticity features of the processes by which the attributes influence the responses. Imagine a measurement procedure involving a device that provides a continuous and highly accurate index of the activity of the amygdala. Because the amygdala has been implicated in the processing of threatening stimuli (e.g., Öhman, Carlsson, Lundqvist, & Ingvar, 2007), the extent to which a stimulus activates the amygdala of an individual could be seen as a measure of how threatening the stimulus is for that individual. It is reasonable to assume that, without special training, people have little voluntary control over the activation of the amygdala. That is, the actual level of activity of the amygdala will most likely be independent of goals to promote or counteract activity in the amygdala. People are probably also unaware of the activation level of their amygdala and whether a stimulus activates the amygdala. This does not imply, however, that amygdala activity provides a measure of threat that is fully implicit. For instance, people might be able to exert indirect control over amygdala activity by avoiding to process the (threat value of the) stimulus or engaging in other voluntary emotion regulation strategies (e.g., Cunningham, Van Bavel, & Johnsen, 2008). As is the case with behavioral measures, empirical research is needed before it can be concluded that a physiological or neurological measure is an implicit measure. In sum, neurological (and physiological) measures are not by definition more implicit than behavioral measures.

We would also like to note that physiological and neurological measures are not by definition more valid measures than behavioral measures. As De Houwer et al. (2009a) pointed out, the validity of a measure of psychological attributes can go only as far as the validity of the assumptions about the to-be-measured attributes. Detailed conceptual analyses and empirical research are needed to validate also physiological and neurological measures. Let us return to the example of amygdala activation as a measure of the threat value of stimuli. This idea is based on research showing that the presentation of threatening stimuli leads to a higher activation of the amygdala (e.g., Öhman et al., 2007). More recent research has shown, however, that also highly relevant positive stimuli (e.g., erotic pictures or rewards) lead to an increase in amygdala activation, giving rise to the hypothesis that threatening pictures activate the amygdala not because of their threat value but because of their relevance (e.g., Sander, Grafman, & Zalla, 2003). If this new hypothesis would turn out to be correct, it would invalidate amygdala activation as a measure of threat value. One could of course reinterpret the measure as an index of relevance but the validity of this interpretation would also depend on the validity of the empirical evidence on which this interpretation is based. Moreover, this would raise important conceptual questions about what it actually means to say that a stimulus is relevant. In sum, as is the case with behavioral measures, the validity of neurological (and physiological) measures depends on the outcome of conceptual analyses and empirical research.

The response content criterion: Symbolic and nonsymbolic measures. All responses, whether behavioral, physiological, or neurological, have several physical features. Examples of such physical features are the time at which a behavioral response is emitted (which provides the basis for reaction time measures), the amount and time course of sweating (which provides the basis for skin conductance measures), and the location of increased blood flow in the brain (which provides the basis for fMRI measures). In principle, measurement outcomes can be derived from any physical feature of any type of response. Some responses not only have physical features; they also have a meaning. Such symbolic responses represent certain objects, concepts, or ideas. In the case of symbolic responses, measures can be based on the physical features of responses (e.g., the speed with which they are emitted, the pitch with which words are spoken), but they can also be derived from the symbolic properties of responses, that is, from their representational content. According to the response content criterion, a measure can be classified as symbolic if it is derived from the meaning of symbolic responses (e.g., the content of what someone says) and as nonsymbolic if it is derived merely from the physical properties of (symbolic or nonsymbolic) responses (e.g., the speed or amplitude with which something is said).

To determine whether a measure is symbolic or nonsymbolic, it does not suffice to check whether it is derived from a symbolic response or a nonsymbolic response. This is because nonsymbolic measures can be derived also from the nonsymbolic properties of symbolic responses (see Figure 2). For instance, one could imagine a measure of racial attitudes that is derived not from what someone says about Black people (i.e., the symbolic meaning of the description) but from the intonation with which the person speaks about Black people. Because the intonation with which words are spoken is a physical, nonsymbolic property of words, measures that are based on this property are nonsymbolic.

Well-known examples of symbolic measures are the Rorschach test (Exner, 2003) and the Thematic Apperception Test (TAT, Tuerlinckx, De Boeck, & Lens, 2002) in which psychological traits are inferred from the content of the description that participants give of pictorial stimuli. A more recent example is the Affect Misattribution Procedure (AMP; see Payne et al., 2005) in which the attitude toward a prime stimulus is derived from what people convey about how much they like a target stimulus that rapidly follows the (briefly presented) prime stimulus.

The response content criterion does not overlap entirely with the other external criteria that we have considered so far (see Figure 2). First, indirect measures can be both symbolic and nonsymbolic. AMP measures, for instance, are indirect and symbolic. They are indirect because attitudes toward the prime stimulus are derived from what people say about the target stimuli. They are based on the meaning of what people say about those targets and thus are symbolic. IAT measures, on the other hand, are indirect and nonsymbolic. They are indirect because the attributes are estimated on the basis of categorization responses rather than self-assessment. They are nonsymbolic because the estimate is derived from the speed rather than the meaning of the responses. [2] All direct (and self-report) measures are, however, symbolic because the measurement outcome is based on the meaning of a self-assessment. Second, the response content criterion does not overlap with the response system criterion (see Figure 2). Whereas the latter refers to the type of response (behavioral, physiological, or neurological), the former refers to the specific property of the response from which the measurement outcome is derived (physical or symbolical). The independence of the two criteria is most clear when considering behavioral responses. Measures based on behavioral responses can be either symbolic (if the measures are based on the meaning of the responses) or nonsymbolic (if the measures are based merely on the physical properties of the responses). It is less clear whether physiological and neurological responses can have symbolic properties and thus give rise to symbolic measures. [3]

The response content criterion also does not overlap entirely with the internal classification criteria that we have discussed. First, it is likely that many attributes can be measured on the basis of both symbolic and nonsymbolic responses. Hence, knowing the symbolic nature of a measure does not reveal what attribute is being measured (what criterion). Nevertheless, symbolic measures might be more suitable to capture symbolic attributes. Many of the psychological attributes that psychologists want to measure, can be regarded as symbolic in that the attributes represent beliefs about events in the world (e.g., attitudes, stereotypes). It is possible that such symbolic attributes are expressed more accurately in the meaning of symbolic responses than in the physical properties of (symbolic or nonsymbolic) responses. Second, there may be processes that are common to all symbolic measures, but it is also likely that different symbolic measures differ with regard to some underlying processes. Nonsymbolic measures are so diverse that classifying them as nonsymbolic says little or nothing about the underlying processes (how criterion). Finally, the symbolic nature of the measure says little about its implicitness (implicitnes criterion). One exception is that many symbolic measures are based on responses that can be controlled (e.g., describing a picture; but see Footnote 3). This creates room for influences of factors other than the to-be-measured attribute (e.g., social desirability concerns). Nevertheless, regardless of whether a measure is symbolic or nonsymbolic, the only way to arrive at definite conclusions about its validity and implicitness is by conducting research.

The Stimulus-Response-Compatibility (SRC) criterion. Many of the implicit measures that have been introduced during the past twenty years are based on reaction time tasks. The best known examples of these reaction time measures are affective priming effects (e.g., Fazio et al., 1995) and IAT effects (e.g., Greenwald et al., 1998). On the basis of the criteria specified so far, most reaction time based measures can be classified as indirect, behavioral, and nonsymbolic: The psychological attribute is estimated not on the basis of a self-assessment of the to-be-measured attribute but the speed with which participants respond to certain stimuli (see Figure 2). This important subset of implicit measures can be classified further using the SRC criterion that was first described by De Houwer (2003b). He pointed out that reaction time measures are most often based on a comparison of trials that differ with regard to a certain type of SRC.

Take the example of an affective priming task in which participants respond on the basis of the valence of positive and negative target stimuli that are preceded by positive and negative prime stimuli. On some trials, the task-irrelevant valence of the prime and the task-relevant valence of the target are compatible (e.g., HAPPY - SUMMER); on other trials they are incompatible (e.g., HAPPY - CANCER). Hence, the procedure involves a manipulation of stimulus-stimulus (S-S) compatibility, that is, the match between an irrelevant and a relevant aspect of the stimuli presented on a trial. The affective priming task involves also a manipulation of irrelevant stimulus-response (S-R) compatibility. On some trials, the irrelevant valence of the prime and the valence of the response are compatible (e.g., say “GOOD” to the target SUMMER that is preceded by the prime HAPPY) whereas on other trials they are incompatible (e.g., say “BAD” to the target CANCER that is preceded by the prime HAPPY). In the affective priming task, there is even a confound between the manipulation of S-S compatibility and irrelevant S-R compatibility. Because participants respond to the valence of the target, whenever the valence of the prime and the target match (S-S compatible), the valence of the prime and the correct response also match (irrelevant S-R compatible). When the valence of the prime and target differ (S-S incompatible), the valence of the prime and the correct response also differ (irrelevant S-R compatible). Hence, the affective priming task can be classified on the basis of the SRC criterion as a task in which S-S and irrelevant S-R compatibility are manipulated in a confounded manner (De Houwer, 2003b). Other reaction time tasks involve a manipulation of other types of compatibility (e.g., relevant S-R compatibility) or different combinations of types of compatibility. Reaction time measures can thus be classified on the basis of the types of compatibility that are manipulated in the procedure from which the measure is derived (see De Houwer, 2003b).

The SRC criterion does in part overlap with the other external criteria that we have put forward. Knowing that a measure can be classified according to the SRC criterion allows one to infer that the measure is likely to be an indirect, nonsymbolic, behavioral measure. On the other hand, knowing that a measure is indirect, symbolic, or behavioral does not allow one to decide whether the SRC criterion applies (because it is possible in principle to create indirect, nonsymbolic, and behavioral measures that are not based on reaction times) or which type of compatibility is manipulated (because reaction time measures can be based on different types of compatibility).

The overlap between the SRC criterion and the internal classification criteria is also not perfect. First, knowing the type of compatibility on which a reaction time measure is based does not allow one to infer which attribute is being measured (what criterion). Second, different types of compatibility can have an effect because of different types of processes (how criterion). Nevertheless, information about the SRC criterion can provide clues about the processes by which the to-be-measured attribute causes the measurement outcome. Again take the example of the affective priming task. In many applications of the affective priming task, the measurement outcome corresponds to a difference in reaction times between trials on which a particular class of prime stimuli (e.g., faces of black persons) is followed by a positive target and trials on which the same prime stimuli are followed by a negative target. On the basis of the SRC criterion, we know that these trials differ not only in the compatibility between the prime and the target stimuli (S-S compatibility) but also in the compatibility between the primes and the responses (irrelevant S-R compatibility). Hence, the measurement outcome (i.e., the difference in reaction times) could be due not only to the fact that the primes influence the processing of the targets but also to the fact that the primes influence the selection of the target responses. Analyses such as these have inspired theories and research about the processes underlying various reaction time measures (e.g., De Houwer, 2001; Gawronski & Bodenhausen, 2005; Olson & Fazio, 2003). Nevertheless, a classification on the basis of the SRC criterion at best provides only clues about the underlying processes. These clues need to be verified with empirical studies.

The SRC criterion can also provide clues about the implicitness criterion. Most if not all measures to which the SRC criterion applies are reaction time measures. Effects in reaction time measures arise in a very short period of time (often just a few hundred milliseconds) and seem to be difficult to control. For instance, the well known Stroop color-word effect has long been considered to be the prototypical example of an automatic effect. Naming the color of an incongruent Stroop stimulus (e.g, the word RED in blue letters) takes much longer than naming the color of a congruent Stroop stimulus (e.g., the word BLUE in blue letters) even when participants try not to be influenced by the meaning of the words and when the words are presented only very briefly (see MacLeod, 1991, for a review). More recent research, however, has shown that even the Stroop color-word effect is not impervious to control (e.g. Kane & Engle, 2003). Hence, measuring psychological attributes by looking at their effects on reaction times does not guarantee that the measure is implicit in all possible ways. Moreover, little is known about whether certain types of compatibility are associated with particular features of automaticity. Empirical research about implicitness is therefore necessary also for measures that can be classified according to the SRC criterion.

What criteria should we use to classify measures?

The criteria that we have discussed are only a subset of all possible criteria that could be used to differentiate between implicit measures. For instance, all the external criteria that we put forward focus on just one aspect of measurement procedures, namely the type of responses or response features that are used to determine the measurement outcome. Undoubtedly, other external classification stimuli can be formulated on the basis of other aspects of measurement procedures (e.g., the kind of stimuli that are used). This raises the interesting question of how to determine which criteria are the best and should be used for classifying measures. In other words, what meta-criteria can we use to evaluate the classification criteria? In our opinion, at least two meta-criteria are important. The first concerns the applicability of the classification criterion. A classification criterion is useful only if there are ways to determine whether it applies to a certain measure. The extent to which a classification criterion meets this meta-criterion depends on how easy it is to apply it and on how many measures it can be applied to. The second meta-criterion concerns the functionality of the classification criterion. The main raison d’être of implicit measures is that they might allow one to predict variability in behavior above what can be predicted on the basis of other, explicit measures. An ideal classification criterion would be one that (a) distinguishes between implicit measures that do and do not have incremental predictive validity or (b) indicates the conditions under which certain measures will have incremental predictive validity. Having such classification criteria is important because it could lead to the discovery of variables that determine whether implicit measures will have added value. This could help researchers to optimize their measures. We will now evaluate the extent to which both meta-criteria apply to the external and internal classification criteria that we discussed earlier.

Regarding the meta-criterion of applicability, it seems safe to conclude that external criteria are generally more easily applicable than internal criteria. Applying the external criteria requires only an analysis of the objective properties of the measurement procedure. For instance, in order to determine whether a measure is a direct measure, it suffices to see whether the measurement outcome is derived from the responses that participants are asked to use to express their self-assessment of the to-be-measured attribute. There is no reason to do research about this; it can simply be verified by looking at the measurement procedure. Most external criteria can also be applied to a variety of different measures.

Although the internal criteria can in principle be applied to most if not all measures, they are more difficult to operationalize. Internal representations and processes cannot be observed directly. They are hypothetical constructs that psychologists have invented as a way to think about what happens in the black box that mediates the relation between the environment and behavior (see Figure 1). Although it seems more than reasonable to assume that there are representations and processes that determine behavior, their nature and properties can be inferred only indirectly from (behavioral, physiological, or neurological) responses that are observable. The applicability of internal criteria thus depends on the degree to which they can be made objective. This can be done by linking them to observable responses in objective situations. The chapter by Moors et al. (this volume) discusses in detail ways to verify whether processes possess certain automaticity features. For instance, in order to examine whether the processes underlying a measure are uncontrolled (in the sense of the goal to alter the outcome of the process), one can set up a situation in which participants are asked to try to fake a certain measurement outcome. If the instruction does not alter the validity of the measurement outcome, this provides evidence for the conclusion that the measure is uncontrolled.

As becomes clear from Moors et al.'s chapter, there are many problems and pitfalls when trying to link automaticity features to observable responses. This complicates the application of the implicitness criterion. Applying the how criterion poses an even more daunting task. At present, we know little if any observable responses or other phenomena for which there is strong evidence that they unambiguously reveal the operation of a particular process.[4] For instance, despite the popularity of the idea of spreading of activation, we do not know an unambiguous observable indicator of this process. In order to infer the operation of a particular process on the basis of a particular observable phenomenon, the process needs to be a necessary and sufficient condition of that phenomenon. This is often if not always an implausible assumption. Even if one could find a particular observable phenomenon that currently can be explained in terms of only one kind of process, there is no guarantee that the same phenomenon cannot result from other processes that have not yet been considered. For instance, whereas the phenomenon of priming (e.g., faster responses to targets in the context of a related prime) was for a long time regarded as an objective indicator of spreading of activation, it has become clear that priming effects can be due to a host of other processes (e.g., Ratcliff & McKoon, 1994). Attempts have been made to delineate when priming effects do provide evidence for spreading of activation, but the past has learned that specific process explanations of observable phenomena rarely remain unchallenged. The uncertainty about future theoretical developments constitutes also another problem for the applicability of internal criteria of implicit measures. Theories about the exact nature of psychological processes and representations change constantly (also see Footnote 4). If measures would be classified on the basis of current theories about the processes underlying the measures, this classification would become useless when theories change. In sum, the applicability of internal classification criteria is endangered not only by the lack of objective criteria for establishing the presence of certain processes but also by the (expected) lack of stability in theories of the processes that might underlie the measures. External classification criteria are by far superior to internal classification criteria in terms of the meta-criterion of applicability.

The opposite is likely to be true when considering the meta-criterion of functionality. We have argued that implicit measures differ from explicit measures with regard to the automaticity or nature of the processes by which psychological properties influence the measurement outcome. Hence, the added value of implicit measures probably depends on the automaticity or nature of the processes that underlie the measure. In other words, its seems reasonable to expect that classifying measures on the basis of their automaticity features or the nature of the underlying processes will provide more information about the incremental predictive validity of the measures than classification criteria based on objective properties of the measurement procedure.

Nevertheless, at least some external criteria are also likely to be functional. The properties of the measurement procedure do not give a perfect indication of the processes that produce the outcome, but they can put strong restrictions on the processes that can operate. For instance, if stimuli are presented so briefly that participants are not aware of their presence, then processes that require awareness of the instigating stimuli are unlikely to operate. Likewise, if the measurement outcome is based on a comparison of reaction times on trials that differ with regard to the similarity of a relevant and irrelevant stimulus, it is at least possible that the outcome is produced by processes akin to spreading of activation. Assuming that the added value of implicit measures depends on the processes underlying the measure, external classification criteria are likely to have at least some functional value in identifying which measures have predictive validity under which conditions.

Ideally, there should be an interplay between the further development of external and internal classification criteria. Examining empirically which processes are likely to operate in which measurement procedures could provide new information not only about which measurement procedures are likely to produce implicit measures with added predictive validity (which would increase the functionality of external classification criteria) but also about what observable phenomena are good indicators of specific processes (which would increase the applicability of internal classification criteria). Ultimately, external criteria are probably the most crucial ones. Because psychological processes are not observable, internal classification criteria will always have to be linked to external, observable phenomena. Once this has been done, the internal classification criteria de facto become external ones because they can be defined solely in terms of observable phenomena.

Another way to deal with the potential problems of external and internal classification criteria is simply forget about ways to classify measures. One could adopt a purely pragmatic approach that focuses merely on assessing the reliability and (incremental) validity of individual measures without much consideration for regularities in the measurement procedures or processes underlying the measures (see Nosek & Greenwald, 2009). On the one hand, we agree that, in order to better understand and predict behavior, it would suffice to know whether a measure allows one to predict variability in behavior that cannot be predicted by other measures. On the other hand, we also realize that a purely pragmatic approach does have serious limitations. In the absence of useful classification criteria, there are no means for comparing various measures or for understanding why some measures do and other measures do not have incremental validity. It also becomes difficult to predict when a new measure will be related to a particular behavior. Progress in obtaining evidence for relations between measures and behavior will thus proceed slowly and in a haphazard manner. Likewise, there will be little guidance for attempts to improve the quality of the measures. In the end, a purely pragmatic approach might be less efficient than a conscientious conceptual, experimental, and theoretical approach to understanding implicit measures of psychological attributes (see De Houwer et al., 2009b).

The criteria put forward in this chapter could in fact provide a guiding framework for research on the psychometric properties of implicit measures. For instance, by characterizing various measures on the basis of the various criteria, one could uncover those properties that moderate the relation between the various measures (including reliability, i.e., the relation of a measure with itself). Our criteria could also provide guidance for uncovering those properties that determine the predictive validity of implicit measures. Importantly, analyses that are guided by this framework could already be conducted using the data that have been gathered in the past.

Summary and Conclusion

A large number of implicit measures have been introduced during the past twenty years. In line with De Houwer et al. (2009a), we argued that an implicit measure can be defined as the outcome of a measurement procedure that is caused by the to-be-measured psychological attribute (e.g., an attitude or stereotype) by means of automatic processes. In the present chapter, we focused on the ways in which implicit measures can differ. Based on our definition of implicit measures, we proposed that measures can differ with regard to internal properties (i.e., the properties of the attributes and processes underlying the measure) and external properties (i.e., the properties of the measurement procedure). With regard to the internal properties, we put forward the what criterion (i.e., What attributes influence the measure?), the how criterion (i.e., What is the nature of the processes underlying the measure?) and the implicitness criterion (i.e., What are the automaticity features of the processes underlying the measure?). With regard to the external properties, we discussed the self-assessment criterion (i.e., Is the measurement outcome derived from a self-assessment by the participant of the to-be-measured attribute?), the response system criterion (i.e., Does the observed response belong to the behavioral, physiological, or neurological response system?), the response content criterion (i.e., Is the measurement outcome based on the physical properties or on the meaning of the responses?), and the SRC criterion (i.e., How are different aspects of the stimulus display related to each other and to the responses?). Discussing these criteria also allowed us to make explicit our understanding of the terms implicit, explicit, direct, indirect, self-report, symbolic, and nonsymbolic measure and how these different terms relate to each other.

The classification criteria that we put forward are certainly not the only ones and perhaps not even the best ones. We argued that the quality of classification criteria can be determined on the basis of two meta-criteria: (a) Applicability (Can the classification criterion easily be applied to various measures?) and (b) Functionality (To what extent do measures that meet the same classification criterion have the same incremental predictive validity under the same set of conditions?). Whereas external classification criteria score good on applicability and poor on functionality, the reverse is true for internal classification criteria. Future research could help improve both kinds of criteria by examining empirically which processes are likely to operate given the presence of certain measurement procedures.

By making explicit various criteria that can be used to classify implicit measures and by specifying meta-criteria to evaluate these classification criteria, we hope to have provided new conceptual tools for improving communication and future research on implicit measures.

References

Anastasi, A. (1988). *Psychological testing*. New York, NY: Macmillan.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: A measure of association strengths. *Experimental Psychology, 56*, 329-343.

Bargh, J. A. (1992). The ecology of automaticity. Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology, 105*, 181-199.

Barnes-Holmes, D., Murtagh, L., & Barnes-Holmes, Y. (in press). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes towards meat and vegetables in vegetarians and meat-eaters. *The Psychological Record.*

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology, 42*, 192-212.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*, 425-440.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal Of Personality And Social Psychology, 79,* 631-643.

Butcher, J., Derksen, J., Sloore, H., & Sirigatti, S. (2003). Objective personality assessment of people in diverse cultures: European adaptations of the MMPI-2. *Behavior Research and Therapy, 41*, 819-840.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407-428.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The quad-model of implicit task performance. *Journal of Personality and Social Psychology*, *89,* 469–487.

Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activation. *Psychological Science, 19,* 152-159.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37*, 443-451.

De Houwer, J. (2003a). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77-85.

De Houwer, J. (2003b). A structural analysis of indirect measures of attitudes. In J. Musch & K.C. Klauer (Eds.), *The Psychology of Evaluation: Affective Processes in Cognition and Emotion* (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum.

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The Handbook of Implicit Cognition and Addiction* (pp. 11-28). Thousand Oaks, CA: Sage Publishers.

De Houwer, J. (2009). Comparing measures of attitudes at the procedural and functional level. In R. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the New Implicit Measures* (pp. 361-390). Mahwah, NJ: Lawrence Erlbaum.

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 12*, 45-61.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.). *Implicit measures of attitudes: Procedures and controversies* (pp. 179-194). New York: Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009a). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347-368.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009b). Theoretical claims necessitate basic research. *Psychological Bulletin*, *135,* 377-379.

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General*, *137*, 262-281.

Exner, J. E., Jr. (2003). *The Rorschach: A comprehensive system: Vol. 1: Basic foundations* (4th ed.). New York: Wiley.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69,* 1013-1027.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology, 89*, 672-685.

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment, 24* 218-225.

Greenwald, A. G., McGhee, D. E, & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74,* 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003) Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social*

Psychology, 85, 481-481.

Kane, M., J., & Engle, R. W. (2003). Working-Memory Capacity and the Control of Attention: The Contributions of Goal Neglect, Response Competition, and Task Set to Stroop Interference. *Journal of Experimental Psychology: General, 132*, 47–70.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16-32.

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology, 80*, 669–685.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect. An integrative review. *Psychological Bulletin, 109*, 163-203.

Mogg, K., Bradley, B. P., Field, M., & De Houwer, J. (2003). Eye movements to smoking-related pictures in smokers: Relationship between attentional biases and implicit and explicit measures of stimulus valence. *Addiction, 98*, 825-836.

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin, 132*, 297-326.

Moors, A., Spruyt, A., & De Houwer, J. (this volume). In search of a measure that qualifies as implicit: Recommendations based on a decompositional view of automaticity. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition, 19*, 625–666.

Nosek, B. A., & Greenwald, A. G. (2009). (Part of) the Case for a Pragmatic

Approach to Validity: Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin, 135*, 373-376.

Nuttin, J. M. (1985). Narcissism beyond gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15*, 353-361.

Öhman, A., Carlsson, K., Lundqvist, D., & Ingvar, M. (2007). On the unconscious subcortical origin of human fear. *Physiology & Behavior, 92*, 180-185.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*, 636-639.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277–293.

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement. *Journal of Experimental Social Psychology, 44*, 386-396.

Ratcliff R., & Mckoon, G. (1994). Retrieving information from memory - spreading-activation theories versus compound-cue theories. *Psychological Review, 101*, 177-184.

Rinck, M., & Becker, E. (2007). Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 105-120.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139-165.

Sander, D., Grafman, J., & Zalla, T. (2003). The Human Amygdala: an Evolved System for Relevance Detection. *Reviews in the Neurosciences, 14*, 303-316.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology, 39*, 75-82.

Schnabel, K, Banse, R., & Asendorpf, J. (2006). Employing automatic approach and avoidance tendencies for the assessment of implicit personality self-concept: The Implicit Association Procedure (IAP). *Experimental Psychology, 53*, 69-76.

Sriram, N., & Greenwald, A. G. (in press). The Brief Implicit Association Test. *Experimental Psychology*.

Stacy, A. W., Ames, S. L., & Grenard, J. L. (2007). Word associations tests of associative memory and implicit processes: Theoretical and assessment issues. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 75-90). Thousand Oaks, CA: Sage Publishers.

Tuerlinckx, F., De Boeck, P., & Lens, W. (2002). Measuring needs with the thematic apperception test: A psychometric study. *Journal of Personality and Social Psychology, 82*, 448-461.

Vargas, P. T. (2004). On the relationship between implicit attitudes and behavior: Some lessons from the past, and directions for the future. In G. Haddock & G. R. Maio (Eds.), *Contemporary Perspectives on the Psychology of Attitudes* (pp. 275-297). NY: Psychology Press.

Wentura, D., & Rothermund, K. (2007). Paradigms we live by. A plea for more basic research on the IAT. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 195-215). New York: Guilford Press.

Wigboldus, D. (2001, September). *The difference between relative and absolute implicit associations*. Paper presented at the third meeting of the European Social Cognition

Network, Houffalize, Belgium.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262-274.

Author Note

Jan De Houwer and Agnes Moors, Ghent University, Belgium. Agnes Moors is a postdoctoral fellow of the Scientific Research Foundation – Flanders (FWO). Preparation of this chapter was supported by GOA Grant BOF/GOA2006/001 and Methusalem Grant BOF09/01M00209 of Ghent University

Correspondence concerning this paper can be addressed to Jan De Houwer, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. E-mail: Jan.DeHouwer@UGent.be

Footnotes

1. As Malte Friese pointed out to us during a conference, the terms “direct” and “indirect” are somewhat ambiguous in that they could also be understood as referring to other properties of measures. For instance, they could be interpreted as referring to the complexity of the processes underlying the expression of the to-be-measured attribute with “direct” implying less or less complex processes than the term “indirect”. Nevertheless, in the remainder of the manuscript, we will continue to use the terms direct and indirect measures but only in the sense of whether the measurement outcome is derived from a self-assessment of the to-be-measured attribute.
2. Reaction time based measures such as IAT effects can also be calculated on the basis of whether responses are correct or incorrect. Like response latency, response accuracy can be regarded as a physical response property. It can be determined solely on the basis of whether certain relevant physical properties of the responses (e.g., location of the keypress) match the physical properties of the response that should have been emitted according to the instructed response rules. Hence, even outcomes that are (co-)determined by response accuracy can be regarded as nonsymbolic measures. Nevertheless, like any other physical response property, response accuracy can be given a meaning. Assume that participants are instructed to self-assess the to-be-measured attribute and express this self-assessment via the accuracy of their responses. For instance, participants could be informed about response rules but then asked to press the incorrect key whenever they see an item they like and the correct key whenever they see an item they dislike. Attitudes toward the items can then be derived from calculating the number of correct and incorrect responses. Such a

measure would be a direct attitude measure because the to-be-measured attribute is estimated on the basis of behavior (i.e., accuracy of responses) that participants were asked to use to self-assess the to-be-measured attribute. Importantly, this measure would qualify as symbolic because the measurement outcome depends on the fact that an incorrect response is interpreted as symbolizing "I like" and a correct response is interpreted as symbolizing "I dislike". The measurement outcome would thus be derived from the meaning assigned to the accuracy of the responses, not from the accuracy as such.

3. In order to intentionally convey a meaning via a certain response, a person must have at least some degree of conscious control over that response. Because (most) physiological and neurological responses are probably difficult to control in this manner, it seems unlikely that meaning can be conveyed through these responses in an intentional manner. However, it is possible that responses can convey meaning in an unintentional manner, that is, in a manner that is not due to a goal to convey meaning. For instance, the behavioral response of avoiding eye contact with a superior at work can be interpreted as a sign of submissiveness even though the person does not have the intention to communicate his submissiveness via the direction of gaze. These and other types of body language can indeed be regarded as symbolic responses (hence the name "body *language*") that are unintentional. In a similar vein, certain physiological responses (e.g., increased sweating when seeing a superior at the workplace) or neurological responses (e.g., increased activation of the amygdala) can be seen as conveying a meaning.
4. Neurological activity does not provide a direct reflection of psychological processes and representations. As with the interpretation of behavioral and physiological

responses, assumptions need to be made about which aspects of neurological activity are linked to which psychological processes and representations. The example of amygdala activity as an indicator of relevance rather than threat (see section on response system criterion and Sander et al., 2003) illustrates that theoretical assumptions can change over time. The validity of these assumptions is also limited by the validity of the available theories about the nature of processes and representations. Although progress has been made, one would be hard pressed to argue that current psychological theories provide the ultimate truth about the nature of psychological processes and representations. Identifying specific neurological responses with specific processes would even hamper theoretical development because it would detract attention away from possible alternative explanations of those neurological responses.

Table 1. Overview of different criteria that can be used to classify implicit measures.

Name	Criterion	Types
	Internal Criteria	
What	What is the to-be-measured attribute?	Racial attitudes Gender stereotypes, …
How	By which process does the to-be-measured attribute cause the outcome?	Spreading of activation Response conflict …
Implicitness	Which automaticity features characterize the process by which the to-be-measured attribute causes the outcome?	Unintentional Unconscious Efficient …
	External Criteria	
Self-assessment	Is the measure derived from a self-assessment of the to-be-measured attribute?	No: Direct Yes: Indirect
Response system	Is the measure derived from behavioral, Physiological, or neurological responses?	Behavioral Physiological Neurological
Response content	Is the measure derived from the symbolic or physical properties of responses?	Symbolic Nonsymbolic
SRC	Is the measure derived from a comparison of trials that differ with regard to a type of SRC?	Relevant SRC Irrelevant SRC S-S …

Note: "…" signifies that other types are possible.

Figure Captions

Figure 1. A schematic representation of the definition of an implicit measure. Certain aspects of the responses that a person shows when submitted to a certain situation are assumed to be mediated by a mental attribute of the person (e.g., an attitude) by means of automatic mental processes (arrows). The responses are used to determine an outcome (i.e., the measure) which is then used to make inferences about the attributes of the person.

Figure 2: A graphical representation of how the different extrinsic criteria are related. SRC stands for “stimulus-response compatibility”. See Footnote 3 for a discussion of whether all symbolic measures are behavioral. Additional (combinations of) SRC types are possible.

Figure 1.

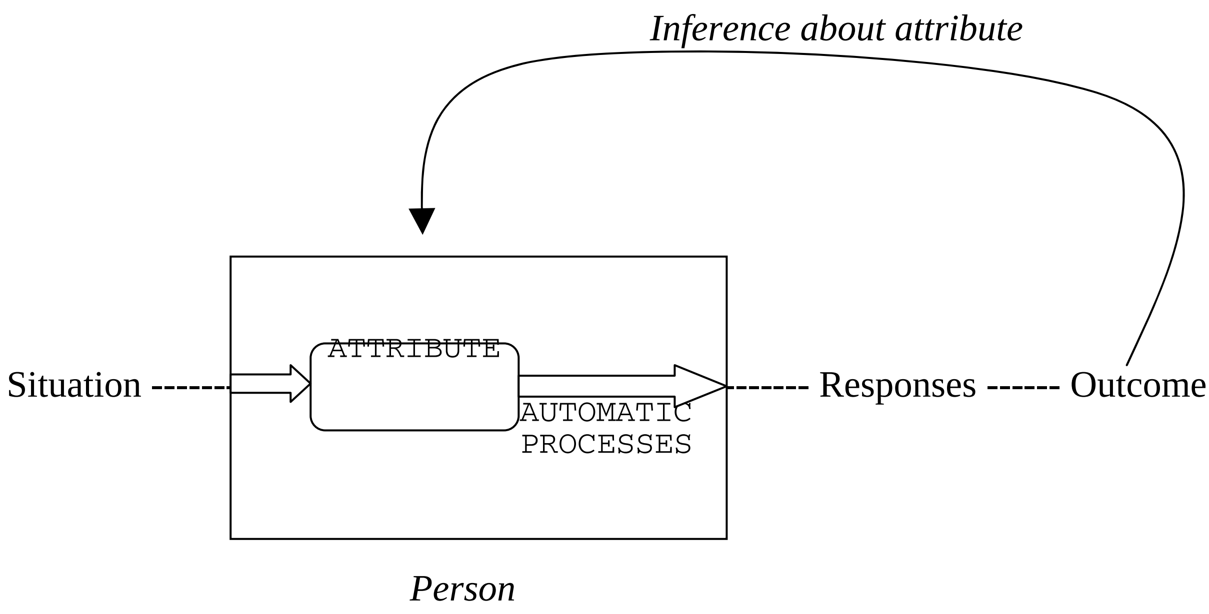
Inference about attribute
Situation
ATTRIBUTE
AUTOMATIC
PROCESSES
Responses
Outcome
Person

Figure 2.

SELF ASSESSMENT?	RESPONSE CONTENT?	RESPONSE SYSTEM?	TYPE OF SRC?
Yes: Direct	Symbolic	Behavioral *	
No: Indirect	Symbolic	Behavioral *	
	Nonsymbolic	Behavioral	Relevant S-R
			Irrelevant S-R
			S-S …
		Physiological	
		Neurological	

* Self-report measures include all direct and indirect symbolic behavioral measures that are derived from self-assessments of attributes