

**Digging for the real attitude: Lessons from research on implicit and explicit self-
esteem**

Ap Dijksterhuis

Luuk W. Albers

Karin C.A. Bongers

University of Amsterdam

Running head: Archeology and architecture

This research was supported by NWO-Vernieuwingsimpuls 016.025.030. Address for
correspondence: Ap Dijksterhuis, Social Psychology Program, University of Amsterdam.
Roetersstaat 15, 1018 WB, Amsterdam, The Netherlands. E-mail: a.j.dijksterhuis@uva.nl

It has once been argued that attitude formation and attitude expression are more reminiscent of architecture than of archeology. Rather than uncovering true, deeper beliefs and values, people's attitudinal expressions are the result of often distorted, temporary constructions created on the spot (Bettman, Luce & Payne, 1998; see also Dijksterhuis & Nordgren, 2006). Researchers administer questionnaires aimed at measuring how people think about themselves, about George W. Bush, or about Chocolate Chip cookies, and people construct what we call attitudes. Just as there is nothing wrong with architecture, there is nothing inherently wrong with measuring attitudes with questions, that is, *explicitly*. Usually, the goal behind measuring attitudes is to predict behavior, and indeed, explicitly measured attitudes often do: We are fairly positive about ourselves helping us to navigate life reasonably well, we vote against George W. Bush, and we eat way too many Chocolate Chip cookies.

Still, there is something about this practice that makes it somewhat unsatisfactory. If an old manuscript suggests an undiscovered tomb in the Egyptian desert, we send archeologists to find it, rather than ask architects to recreate the tomb on the basis of some vague descriptions. Have you ever been in a museum, staring in awe at some beautiful piece of old art, only then to discover (by reading the brochure that was handed to you at the entrance) that you are looking at a replica? The real statue created by Michelangelo is in an area inaccessible to the public, and you are looking at a copy made in 1987 in some Florentine factory. Even in such cases, when a replica looks exactly like the original, looking at it just does not feel quite right. It is nothing more than a minor nuisance, and it certainly does not spoil your entire day, but you had preferred to see the real thing.

One way to look at explicitly measured attitudes is to assume that what one measures is all there is. Explicitly measured attitudes are what they are, and there are no such things as underlying, “real” attitudes. This is unsatisfactory of course, as we know enough about unconscious affective and cognitive processes to assume that a “7” on a 9-point Likert-scale about Chocolate Chip cookies is not just something hovering in the air. It must come from somewhere, somehow. It may be the result of a construction process, but we hope it is, at least to some extent, influenced by deeper psychological forces. Hence, an alternative viewpoint that is more realistic (and much more exciting) is to assume that there are such things as “real” attitudes and that what we assess with attitude measures is at least in part based on this real thing. Sure, due to poor construction we routinely end up with very poor replica’s, indeed much more reminiscent of the work of architects working on the basis of vague descriptions than of contemporary Florentine artists who can make a detailed copy of the Michelangelo statue. After all, when we answer a questionnaire we often have not much to work with other than perhaps some vague hints, such as subtle affective reactions or old memories of past behavior. Still, somewhere in that Egyptian desert is the real thing we are looking for. Some process sparked by our millions of brain cells represents that real attitude.

The observation that attitudes are more the result of architecture rather than of archeology was made before psychologists started to develop *implicit* attitude measures. In our view, the creation and development of implicit measures is of paramount importance, because it fundamentally changes the way we think (at least it should) about attitudes. In the present chapter, we will argue that implicit attitude measurement is not just another style of architecture. Instead, we will review evidence (and present some new

evidence) strongly suggesting that implicit measurement reflects archeology. Amateur archeology with limited equipment perhaps, but archeology.

In this chapter, we focus on arguably the most important attitude we have: Self-esteem, or the attitude towards the self. However, our hope is that our thinking is generalizable to attitudes in general. Most of this chapter deals with research on the relation between three protagonists: Explicitly measured self-esteem, implicitly measured self-esteem, and that what we until now called “the real thing”. We start out by defining what we conceive of as this hypothesized “real” attitude and by proposing three hypotheses concerning the relationship between explicit and implicit self-esteem. We then discuss relevant research with the aim to differentiate between these hypotheses and to decide which of the three is the most plausible. Before we end with some conclusions, we present an experiment that we recently conducted.

Three alternative hypotheses on the relation between explicit and implicit self-esteem.

A chapter that features the term “the real thing” too often runs the risk of alienating a scientific audience (but perhaps attracting people from the music industry!), so a definition is in order. The “real” attitude, we propose, is the evaluative “tone” that is automatically activated upon the perception of the attitude object (Bargh, Chaiken, Govender & Pratto, 1992; Fazio, Sanbonmatsu, Powell & Kardes, 1986). With more multifaceted and/or important attitude objects, it is perhaps more appropriate (see e.g., Cacioppo, Crites, Berntson & Coles, 1993) to define the attitude as the sum of the various evaluative tones that are automatically activated. It is a proposed underlying construct constituting the core of the attitude, sitting there waiting to be excavated. It is itself

undisturbed by biasing processes that occur when the attitude is measured or verbalized, but at the same time it feeds such processes. From now on, we call it the core attitude or core self-esteem (see also Dijksterhuis, 2004).

What is the relation between this core attitude and implicitly and explicitly measured attitudes? Or, to turn to self-esteem, is there one core self-esteem that is related to both measures of implicit and explicit self-esteem? And if so, how are they related? Let us briefly discuss three possible hypotheses pertaining to this relation.

1. *The independence hypothesis.* According to this hypothesis, implicit and explicit self-esteem are independent constructs. They happen to partly share their name, they happen to be about the same object, but they are unrelated. Both implicitly measured self-esteem and explicitly measured esteem are based on their own underlying core construct. Implicit self-esteem could be based on the automatically activated evaluative tone, whereas explicit self-esteem could be based on the evaluative tone that becomes apparent only when one explicitly or consciously reflects on the self.

2. *The equal relationship hypothesis.* This hypothesis states that implicit and explicit self-esteem are related because they are both related to the same core attitude, the one defined above. They are simply different manifestations of this core. In addition, the two manifestations do not differ as to how well they represent that core. They measure a different aspect, but generally do equally well.

3. *The hierarchy hypothesis.* This hypothesis also assumes that implicit and explicit self-esteem are related because they are both related to the same core. However, here implicit measures of self-esteem better represent core self-esteem than explicit measures of self-esteem. That is, implicit self-esteem digs deeper and more closely

approaches the hidden Egyptian tomb. This hypothesis also implies that what we measure explicitly is partly based (and can be partly predicted by) what we find when we measure implicitly¹.

In what follows, we will make a (stepwise) comparison between the plausibility of the hypotheses by reviewing evidence. We will start with comparing the independence hypothesis with the remaining two, the equal relationship hypothesis and the hierarchy hypothesis, whereby no distinction will be made between latter two yet.

Before we move on, it should be noted that although we only review evidence on self-esteem, the three hypotheses encompass possible relations between implicit and explicit measures of attitudes in general. Indeed, versions of both the independence hypothesis and the equal relationship hypothesis shine through in work on racial attitudes (i.e., prejudice, Dovidio, Kawakami, Johnson, Johnson & Howard, 1997). Likewise, the hierarchy hypothesis is in part based on, and fully in line with, the work by Fazio and colleagues on the MODE-model (see e.g., Fazio, 1990; Fazio, this volume). The MODE-model also views an attitude as a “core” that can be automatically activated, whereby implicit measures are more proximal indicators of these automatically activated attitudes than more downstream explicit measures.

Are implicit and explicit self-esteem related?

If we find evidence for the notion that implicit and explicit self-esteem are related, this implies that both the equal relationship hypothesis and hierarchy hypothesis are more plausible than the independence hypothesis. In our view, there are currently three relevant sets of research findings. First, quite a number of researchers have directly investigated the relation between implicit and explicit self-esteem by assessing correlations between

the two. A second fruitful avenue is to investigate whether the same specific levels of implicit and explicit self-esteem have the same or comparable consequences for other psychological processes. A third way to shed light on the relationship between implicit and explicit self-esteem is to see if there are experimental manipulations that affect both implicit and explicit self-esteem in comparable ways.

Are implicit and explicit self-esteem correlated? The answer is “sort of”. Some researchers did not find correlations (Baccus, Baldwin & Packer, 2004; Bosson, Swann & Pennebaker, 2000; Jordan, Spencer, Zanna, Hoshino-Browne & Correll, 2003; Spalding & Hardin, 1999), others did find significant correlations (DeHart, Pelham & Tennen, 2006; Greenwald & Farnham, 2000), yet others found significant correlations in some experimental conditions or in some samples and not in others (Jones, Pelham, Mirenberg, & Hetts, 2002; Koole, Dijksterhuis & van Knippenberg, 2001; Pelham, et al., 2005). Various people have concluded that implicit and explicit self-esteem correlate “weakly at best”. There is no arguing with that conclusion, and on the basis of the current state of affairs we cannot say much about the plausibility of the independence hypothesis. Rejecting the independence hypothesis would have required more consistent correlations between explicit and implicit self-esteem. On the other hand, given that some researchers did find significant correlations, the correlational data cannot be interpreted as support for the independence hypothesis either.

It may be noted that the generally low correlations between measures of implicit and explicit self-esteem are at least in part caused by the fact that implicit measures are still in a developing stage. Their reliability is often low (Bosson, Swann & Pennebaker, 2000), and it is not fully understood yet what exactly drives the effects of some of the

implicit measures. Recently, various researchers have proposed improvements to various measures of implicit self-esteem. Both Karpinski (2004) and Albers, Dijksterhuis and Rotteveel (2006) suggested improvements to implicit measures of self-esteem that will likely result in more meaningful correlations between implicit and explicit measures of self-esteem. Wentura, Kulfanek and Greve (2005) even proposed an interesting new measure that alleviates some problems of other measures. Such initiatives to strengthen implicit measures of self-esteem give rise to optimism and it is likely that researchers will obtain higher and more consistent correlations between implicit and explicit self-esteem in the future.

Do implicit and explicit self-esteem have comparable consequences? Explicit self-esteem is known to be predictive of many things, but arguably the best known fact is that it is related to how people cope with negative experiences: High levels of explicit self-esteem help people cope with negative feedback or negative experiences in general. High explicit self-esteem forms a “buffer” against stress and experiences of failure (see e.g., Dodgson & Wood, 1998; Shrauger & Rosenberg, 1970; Steele, 1988). For instance, it has been observed that people with low explicit self-esteem exhibit stronger emotional reactions after failure than people with high explicit self-esteem (Brown & Dutton, 1995) and that people with low explicit self-esteem demonstrate impaired motivation after failure whereas individuals with high self-esteem generally do not (e.g., DiPaula & Campbell, 2002; Shrauger & Rosenberg, 1970).

After only a few years of research on implicit self-esteem, we can safely conclude that implicit self-esteem has indeed comparable consequences. Spalding and Hardin (1999) demonstrated that low implicit self-esteem individuals show more anxiety during

a confronting interview than high implicit self-esteem individuals. Greenwald and Farnham (2000) showed that implicit self-esteem is negatively related to motivation after failure such that people with low self-esteem show a stronger decrease in motivation than people with high self-esteem. Baccus, Baldwin and Packer (2004) demonstrated that people with high implicit self-esteem show less aggression after an insult than people with lower implicit self-esteem. Finally, Dijksterhuis (2004) showed that people with high implicit self-esteem show no changes in mood after negative feedback, whereas people with lower implicit self-esteem reported a more negative mood after negative feedback. Indeed, high implicit self-esteem is a buffer against negative experiences, just as high explicit self-esteem is. These findings strongly suggest that implicit and explicit self-esteem are to some extent related, rendering the independence hypothesis less plausible.

Do (some) experimental manipulations have the same effects on implicit and explicit self-esteem? Currently, there are two areas of research that indeed suggest this to be the case. First, threats to the self have been known to decrease explicit self-esteem. For instance, both Dutton and Brown (1997), and Heatherton and Polivy (1991) found that people report lower explicit self-esteem after negative intelligence feedback. In recent years, various researchers have reported comparable consequences of threats to the self on implicit self-esteem. Jones, Pelham, Mirenberg and Hetts (2002) asked participants to write about a negative aspect of their personality and demonstrated that this lowered implicit self-esteem. In addition, Dijksterhuis (2004) gave participants (bogus) negative intelligence feedback and found that it lowered participants' score on an implicit measure of self-esteem. It is also known that people engage in self-affirming behavior in order to

repair “dents” in their self-esteem. And again, engaging in self-affirmation after threat has been shown to both restore explicit (Steele, Spencer & Lynch, 1993), as well as implicit self-esteem (Koole, Smeets, van Knippenberg & Dijksterhuis, 1999).

Secondly, it has been demonstrated that both explicit self-esteem and implicit self-esteem can be changed by evaluative conditioning (Baccus, Baldwin & Packer, 2004; Dijksterhuis, 2004; Riketta & Dauenheimer, 2003). Evaluative conditioning (see De Houwer, Thomas & Baeyens, 2001, for a review) is a technique in which is an attitude object (the Conditioned Stimulus or CS [plural CSi]) is repeatedly paired with either a positive or a negative stimulus (the Unconditioned Stimulus or US [plural USi]). After a number of pairings, the CS takes on the valence of the USi. In our view, evaluative conditioning is fascinating because it changes an attitude at its core. Earlier, we defined the core attitude as the evaluative tone (or tones) that become automatically activated upon the perception of the attitude object. And it is this evaluative tone evaluative conditioning directly works on.

The procedures used by the different researchers differed only subtly. Baccus et al., (2004) presented participants with self-relevant words (such as their own names) on a computer screen, and in the experimental condition the words were followed by smiling faces. In a control condition, self-relevant words were randomly paired with smiling, frowning and neutral faces. Dijksterhuis (2004) presented participants repeatedly with the word “I”, in the experimental condition followed by positive adjectives. In the control condition, neutral adjectives followed the word “I”. In some of the experiments, all this information was presented subliminally. Riketta and Dauenheimer (2003) followed almost exactly the same procedure, except that whereas in the experiments by

Dijksterhuis the adjectives immediately followed the word “I”, in the Riketta and Dauenheimer experiments the word “I” and the positive adjectives were presented simultaneously. Both Baccus et al., (2004) and Dijksterhuis (2004) assessed implicit self-esteem after the evaluative conditioning procedure, whereas Riketta and Dauenheimer (2004) measured self-esteem explicitly. Crucially, in all sets of studies it was found that evaluative conditioning increased self-esteem².

Where does this leave things? Although the findings on correlations between implicit and explicit self-esteem are inconclusive, other evidence is not. First, high (and low) implicit and explicit self-esteem have comparable consequences for how people deal with negative experiences. Second, various experimental manipulations (threat to the self, evaluative conditioning) have the same effect on implicit as on explicit self-esteem. In our view, this makes the independence hypothesis untenable. There is some sort of relation between implicit and explicit self-esteem. Put differently, they must at least to some extent represent the same underlying core.

Is implicit self-esteem closer to the core than explicit self-esteem?

Now that we have rejected the independence hypothesis, we can begin to analyze which of the two remaining hypotheses is the most plausible. Is the equal relationship hypothesis, whereby (measures of) explicit and implicit self-esteem represent the underlying core attitude equally well (or equally poorly) the best descriptor of the current state of affairs? Or are the relevant findings better described by hierarchy hypothesis, stating that implicit measures of self-esteem represents core self-esteem better?

In order for the hierarchy hypothesis to trump over the equal relationship hypothesis, it has to be proven that explicit self-esteem is more dissociated from the core

attitude than implicit self-esteem. If this is true, it should be possible to demonstrate why this dissociation is indeed more pronounced. As our opening lines suggested, explicit attitudes are often active constructions, more reminiscent of architecture than of archeology. A prediction one can derive from the conceptualization of explicit self-esteem as a construction process is that, since constructive processes are easier to change than underlying representations, explicit self-esteem must be easier to change than implicit self-esteem. Another way to differentiate between the two hypotheses is to examine the construction process itself. Is there evidence for the architectural aspect of explicit self-esteem? Can we find evidence for biasing psychological forces that leads explicit self-esteem away from its underlying core? In what follows, we first look at the changeability of explicit and implicit self-esteem. Later, we examine the evidence for the notion that explicit self-esteem is a construction partly based on biasing processes that are not related to the core attitude.

Is explicit self-esteem easier to change than implicit self-esteem? We already discussed evidence that shows that both implicit and explicit self-esteem can be changed, at least for a brief period of time, by various experimental manipulations. However, what can we say about more enduring changes as a result of major life events?

There is indeed some evidence for a greater flexibility of explicit self-esteem. Hetts and Pelham (2003) found people whose birthday was overlooked (!) reported low implicit self-esteem, whereas their explicit self-esteem was on a normal level. One could assume that when one's birthday is overlooked, this initially has negative consequences for both implicit and explicit self-esteem. However, due to the assumed nature of explicit self-esteem as more of an active construction process, explicit self-esteem can be easier

brought to more normal levels than implicit self-esteem. Although we concede that this interpretation of the findings of Hetts and Pelham (2003) is somewhat speculative, other research from Hetts, Pelham and colleagues (1999) more firmly support the hierarchy hypothesis. Hetts, Sakuma and Pelham (1999) assessed implicit and explicit self-esteem among Asian-Americans who immigrated relatively recently. They reasoned that such major life events would affect both explicit and implicit self-esteem, but that it is more likely, due to the nature of explicit self-esteem measures, that explicit self-esteem changes more quickly than implicit self-esteem. This is exactly what they found. Whereas recent immigrants still demonstrated low implicit self-esteem, their explicit self-esteem soon appeared to be back to normal levels. Fully in line with the hierarchy hypothesis, they concluded that conscious constructions are more malleable than “deeper” unconscious representations (see also DeHart, Pelham & Tennen, 2006).

Is explicit self-esteem a construction process? One could argue that explicit self-esteem must largely be a construction process, simply because people do not have much conscious access to deeper, unconscious processes. Following Nisbett and Wilson (1977) one could reasonably assume that explicit self-esteem (or explicit measures in general) relies on introspective processes to an extent that is unwarranted and perhaps even unrealistic. We simply do not know how we truly feel about ourselves, so we have no choice but to engage in construction. We are architects working with poor and vague instructions.

Pelham et al., (2005) recently reported evidence supporting this idea. They reasoned that some people may have better access to how they truly feel about themselves than others. Now the better people have access to core self-esteem, the less

need there is for construction. That means that, assuming implicit self-esteem reflects core self-esteem better than explicit self-esteem, implicit self-esteem and explicit self-esteem should correlate higher among people who have better access to their core self-esteem. Pelham et al., (2005) reasoned this could well mean that gender moderates the correlation between implicit and explicit self-esteem. After all, aren't women generally better at accessing their deeper feelings than men? Socialization processes make women trust their feelings and intuitions more (Pacini & Epstein, 1999) and we know that women are generally better than men in expressing their emotions (e.g., Lakoff, 1990). Pelham et al., (2005) compared six samples from three different countries and indeed confirmed their prediction. Among men, explicit and implicit self-esteem did not correlate in any of the samples, whereas significant correlations were found in all samples for women (ranging in size from .11 to .51).

Other evidence for explicit self-esteem as a construction comes from studies suggesting that explicit self-esteem assesses factors other than the core attitude towards the self. For example, various researchers have found that explicit self-esteem correlates significantly with style of self-presentation, impression management, and self-deception (Greenwald & Farnham, 2000; Jordan, Spencer, Zanna, Hoshino-Browne & Correll, 2003; Raskin, Novacek, & Hogan, 1991). These findings support the hierarchy hypothesis. Explicit self-esteem may be assessing a mixture of core self-esteem, and various essentially unrelated motives. Especially the finding that explicit self-esteem is correlated with self-deception is interesting. The higher one's explicit self-esteem, the greater the possibility that people's construction work reflects an attempt to fool oneself.

If one is willing to assume that motivated construction takes effort, one can derive a straightforward prediction from the notion that explicit self-esteem correlates with various motives. Obviously, the people whose explicit self-esteem does not reflect core self-esteem are the ones who have to engage in effortful strategies to maintain this inconsistency. Specifically, people with high explicit self-esteem but low implicit self-esteem are the true construction workers. They engage in self-presentation and self-deception, thus, they expend most effort. Conversely, given that implicit self-esteem represents this core attitude quite well, people with comparable explicit self-esteem and implicit self-esteem (both high or both low) do not engage in much construction and hence, do not expend much effort.

There is indeed some evidence that maintaining high self-esteem in the face of negative experiences requires work. People have to “explain things away,” for instance by changing the way they interpret experiences or by making self-serving attributions (e.g., Crocker & Major, 1989; Pelham, DeHart & Carvallo, 2003). Importantly however, there is evidence that this is especially true for people who want to maintain high explicit self-esteem in the face of low implicit self-esteem (Bosson, Brown, Ziegler-Hill and Swann, 2003; Jordan, et al., 2003; McGregor & Marigold, 2003; see also Jordan, Logel, Spencer, Zanna & Whitfield, this volume).

Bosson et al., (2003) investigated two groups of participants. They compared people with both high explicit and implicit self-esteem and people with high explicit but low implicit self-esteem (often called fragile or defensive self-esteem, see e.g., Kernis, 2003). They found that people with low implicit self-esteem engaged more in unrealistic optimism. In addition, they found evidence that supports the notion that high explicit self-

esteem can be related to self-deception. Participants were presented with four personality profiles about themselves ranging from highly unflattering to highly flattering. They were then asked to rate the accuracy of each of the profiles (that were allegedly written by clinical psychology students) and as it turned out, participants with low implicit and high explicit self-esteem rated the very flattering profile as more descriptive of themselves than participants with both high explicit and implicit self-esteem. The different profiles are given in the Appendix to the Bosson et al., (2003) article and this makes the data even more interesting, as the flattering profile is indeed rather extreme, including the phrase “knows that affection and admiration from others are well-deserved” (p. 183).

Jordan et al., (2003) distinguished between the same two groups: People whose explicit and implicit self-esteem are high versus people whose explicit self-esteem is high, while their implicit self-esteem is low. They first established that people with low implicit self-esteem showed more narcissistic behavior. In later experiments, they obtained more direct evidence for the idea that maintaining high explicit self-esteem based on low implicit self-esteem takes effort. They demonstrated that individuals with high explicit but low implicit self-esteem showed much more defensive behavior. They engaged more in ingroup bias, they demonstrated more prejudice when threatened, and they put more effort in dissonance reduction.

McGregor and Marigold (2003) investigated effects of personal uncertainty on “compensatory conviction”. Conviction refers to the extremity and certainty of important personal attitudes and compensatory conviction is the tendency to increase the extremity of such attitudes and the commitment with which such attitudes are held. People under uncertainty generally show compensatory conviction, but McGregor and Marigold (2003)

showed that this is especially true for people with low implicit and high explicit self-esteem. That is, relative to people with both high implicit and explicit self-esteem, people with low implicit and high explicit self-esteem engaged in more compensatory conviction regarding such moral topics as the death penalty or abortion.

The work by Jordan and colleagues (this volume) on people with low implicit and high explicit self-esteem is consistent with our reasoning. Moreover, their analysis sheds some more light on why people with low implicit and high explicit self-esteem have to engage in defensive effort. Jordan et al., reason that implicit self-esteem is not so much unconscious as it is preconscious. Sometimes, especially in the face of threats, people become aware of their (low level of) implicit self-esteem. This fleeting awareness is assumed to be aversive among people with low implicit and high explicit self-esteem, leading to what they call “nagging doubts”. These nagging doubts, in turn, will motivate defensive effort. Jordan et al., present some interesting first evidence for their reasoning in this volume.

To conclude, the research on people with high explicit and low implicit self-esteem clearly shows that maintaining high explicit self-esteem when implicit self-esteem is low is, at least sometimes, hard work – construction work. In general, the evidence for explicit self-esteem as a construction process with many inherent biases is strong, rendering the hierarchy hypothesis more plausible than the equal relation hypothesis.

Finding support for the hierarchy hypothesis

With the hierarchy hypothesis coming out as the most plausible, in the last part of this chapter we try to corroborate the hierarchy hypothesis by discussing (and to some extent testing) the support for a few hypotheses derived from the hierarchy hypothesis.

The first hypothesis following from the hierarchy hypothesis is that there should be an asymmetry in frequency of occurrence of different combination of implicit and explicit self-esteem. That is, we can predict that the combination high explicit/low implicit self-esteem is more common than the combination low explicit/high implicit self-esteem. The second hypothesis pertains to the fact that, if we assume explicit self-esteem is a construction, variations in the degree to which people engage in active construction should affect the relation between implicit and explicit self-esteem. That is, we expect implicit and explicit self-esteem to correlate higher when there is less construction. Both hypotheses will be further discussed, starting with the asymmetry hypothesis.

Is there an asymmetry? First, let us make the rather safe assumption that the construction process underlying explicit self-esteem biases explicit self-esteem more often in a positive rather than in a negative fashion. After all, people are known to be motivated to see themselves (and have others see them) in a positive light. This was already suggested earlier by the finding that explicit self-esteem is correlated with self-presentation style, self-deception and impression management (Greenwald & Farnham, 2000; Jordan et al., 2003; Raskin, et al., 1991). It may certainly be the case that people strategically report low explicit self-esteem (perhaps because they want to come across as modest), but this is likely to be relatively rare.

If this reasoning is correct we should be able to witness the following asymmetry: For people with incongruent implicit and explicit self-esteem (i.e., one is high, the other low), the combination low implicit/high explicit self-esteem should occur much more often, or among many more people, than the combination high implicit/low explicit self-esteem. People with high implicit self-esteem are seldom motivated to report low explicit

self-esteem, whereas low implicit self-esteem individuals may often report relatively high explicit self-esteem (even if, as we have seen, it is often hard work). Now is there such an asymmetry?

One problem is that whether one finds support for this asymmetry or not depends on where one draws the line. When do we categorize implicit self-esteem and explicit self-esteem as truly low or truly high? The evidence for the hypothesized asymmetry is, up to this point, only suggestive. First, various people have made the same prediction (Epstein, 1983; O'Brien & Epstein, 1988; Kernis, 2003). Others have argued that the combination of high implicit/low explicit self-esteem is indeed uncommon in Western cultures but not in Asian cultures (see Kitayama & Uchida, 2003, for a brief review). Kitayama and Uchida (2003) reported that the combination of high implicit/low explicit self-esteem can be found among Western participants but only (or at least mostly) under highly specified measuring circumstances. Concretely, Western participants only showed the high implicit/low explicit self-esteem combination in the context of close, interdependent relations. Perhaps also telling is the fact that the combination low implicit/high explicit has been named – as defensive or fragile self-esteem – and its consequences have been investigated by an increasing number of research groups (e.g., Bosson et al., 2003; Jordan et al., 2003; Kernis, 2003; McGregor & Marigold, 2003), whereas the combination of high implicit/low explicit self-esteem has received relatively little attention (for exceptions, see Jordan et al., this volume; Kitayama and Uchida, 2003). Still, we concede that research is needed to more strongly corroborate this hypothesized asymmetry.

Variations in construction and the relation between implicit and explicit self-esteem. The second hypothesis derived from the hierarchy hypothesis is much easier to test. If explicit self-esteem is partly a construction process guiding people away from core self-esteem and therefore also from implicit self-esteem, it means that the less construction there is, the more explicit self-esteem should correlate with implicit self-esteem. After all, the less construction there is, the better explicit self-esteem should represent core self-esteem.

Koole, Dijksterhuis and van Knippenberg (2001) have reported supportive evidence for this hypothesis. In one experiment, they first assessed people's implicit self-esteem. Later they measured explicit self-esteem and they measured the time it took participants to complete the explicit self-esteem items. They hypothesized that the longer people would take to complete the measure of explicit self-esteem, the more they engaged in active construction. By measuring response times they assessed natural variations in people's degree of construction. They then divided the participants in two groups: Fast responders and slow responders. In support of the hierarchy hypothesis, for fast responders the correlation between explicit and implicit self-esteem was high (.51), whereas for slow responders there was no correlation at all (-.06). In sum, the less people engaged in active construction during assessment of explicit self-esteem, the more explicit self-esteem correlated with implicit self-esteem.

In another experiment, Koole et al., (2001) manipulated rather than measured construction. Again, they first measured participants' implicit self-esteem. Subsequently, explicit self-esteem was assessed and this was either done under cognitive load or not. Obviously, cognitive load prevents people from engaging in too much active construction

and the experimenters predicted that explicit self-esteem would correlate with implicit self-esteem under load, but not necessarily under normal conditions. Indeed, this is what they found. The correlation between explicit and implicit self-esteem was high under load (.48) and absent under normal conditions (-.15). This fully supports the hierarchy hypothesis.

Further support for the hierarchy hypothesis: An experiment

The experiment we report here extends the experiments reported by Koole et al., (2001). Again, we tried to manipulate the extent to which participants would engage in active construction processes biasing explicit self-esteem away from core self-esteem. Before participants' completed measures of implicit and explicit self-esteem, we subliminally primed half of our participants with the goal to be honest. This was done under the guise of a lexical decision task whereby experimental participants were subliminally presented with words such as honest, sincere, and true. Control participants were not presented with words related to honesty. We then measured implicit self-esteem by name-letter preferences and explicit self-esteem with Heatherton and Polivy's (1991) State Self-Esteem Scale³. We tested three hypotheses.

Hypothesis 1. The first hypothesis is the most important and the most straightforward. Assuming that explicit self-esteem is in part the result of a construction process biasing the view of the self in a positive way, the goal to be honest should lead people to engage in this biased construction to a lesser extent. The goal to be honest should lead reported explicit self-esteem to be more strongly related to core self-esteem and therefore to implicit self-esteem. This means that the correlation between the

measures of implicit self-esteem and explicit self-esteem should be higher for people with a primed honesty goal than for control participants.

Hypothesis 2. The advantage of the SSES scale is that explicit self-esteem is divided into three subscales, Appearance self-esteem, Performance self-esteem, and Social self-esteem. Hypothesis 2 pertains to these subscales. One could argue that the extent to which people can positively construct self-esteem differs for the different subscales. After all, reality constraints differ between the different domains they represent. Appearance self-esteem is probably the hardest to strategically bias in a positive way. We can maintain that our attractiveness is on par with that of Brad Pitt or Jennifer Lopez, but it does not make sense. It's a form of absurd self-deception and we know it. A mild form of self-deception is likely to be easier in the domain of Performance self-esteem. The truth is still to some extent objective, but at least one can easily switch between different domains ("Yes, I lost a game of Trivial Pursuit against friends, but I had an A+ for Intro Social Psychology!"). Social self-esteem is, with the exception of extreme cases perhaps, likely the easiest one to steer towards rosiness. One can think about many different relationships, and it is often possible to flexibly interpret social behavior ("His insult was not personal, he must have been in an awful mood"). If this reasoning is valid, this would mean that, for people without an honesty goal, appearance self-esteem correlates highest with implicit self-esteem, whereas social self-esteem correlates lowest with implicit self-esteem. For participants with an honesty goal, these differences should disappear.

Hypothesis 3. Assuming that explicit self-esteem, although partly constructed, does to some extent represent core self-esteem, this should also be true for the three

subscales. If participants primed with honesty engage less in construction, their self-esteem should better reflect core self-esteem (as reflected in Hypothesis 1). As this should be true for all subscales, it follows that the different subscales should “converge” towards core self-esteem and therefore also to each other. Hence, the correlations between the subscales should be higher among people primed with honesty.

In total, seventy-one undergraduate students participated in the experiment, 37 in the control condition, 34 in the honesty-prime condition. The correlations pertaining to the hypotheses are listed in the Table. As can be seen, hypothesis 1 was supported, although the difference just failed to reach significance. As predicted, we found a high correlation between explicit and implicit self-esteem after honesty priming, and no such correlation for control participants. Hypothesis 2 also received support in that, for control participants, the correlation between implicit self-esteem and appearance self-esteem was highest, whereas the correlation between implicit self-esteem and social self-esteem was lowest. The honesty prime significantly increased the correlation between implicit self-esteem and social self-esteem and between implicit self-esteem and performance self-esteem (although this latter effect was marginally significant), whereas the honesty prime did not affect the correlation between implicit self-esteem and appearance self-esteem. Finally, hypothesis 3 also received support. Correlations between subscales were generally higher under honesty conditions than under control conditions, with one of them reaching conventional levels of significance. As predicted, the different subscales converged because the honesty prime decreased the amount of construction work.

In sum, although some of the evidence was statistically somewhat weak, the results support the hierarchy hypothesis. The effects of priming of the honesty goal were exactly as predicted.

Conclusions

To conclude, the hierarchy hypothesis best describes the relation between implicit and explicit self-esteem. Both explicit and implicit self-esteem are in part based on the same underlying construct, that what we called core self-esteem. However, due to the fact that explicit self-esteem is often the consequence of active and biased construction processes, it represents core self-esteem less well than implicit self-esteem. Furthermore, as implicit self-esteem represents core self-esteem better, explicit self-esteem can be partly predicted by implicit self-esteem. In addition, it was shown that (in line with Koole et al., 2001) the correlation between explicit and implicit self-esteem can be increased by interfering with the active construction process explicit self-esteem is partly based on.

Before ending we would like to remark that we do not see implicit self-esteem as an infinitely better construct than explicit self-esteem. In addition, we certainly do not argue that we should stop using the latter. Such a claim would clearly be unwarranted. The relation between explicit and implicit self-esteem is often so weak that it clearly pays off to investigate both and to scrutinize combinations of consistent and inconsistent combinations, as interesting recent research clearly shows. Another reason for not solely relying on implicit self-esteem is the fact that measures of implicit self-esteem are in a sense still in a developing stage. For some measures, the underlying processes driving its effects are not fully understood. Although some important improvements have been proposed recently (Albers, Dijksterhuis & Rotteveel, 2006; Karpinski, 2004; Wentura,

Kulfanek & Greve, 2005), there is still quite some work to be done to optimize implicit measurement.

However, we do want to maintain that explicit self-esteem is a less pure form of self-esteem. In addition to being less pure though, it is also more rich and multifaceted. It is in part construction rather than excavation work. It only weakly reflects core self-esteem and it is affected by self-deception, impression management, and self-presentation style. However, this inherent richness is not in itself problematic, after all, explicit self-esteem predicts quite a number of psychological processes very well.

To recapitulate, both explicit and implicit self-esteem clearly have their value, also in an Egyptian desert. Measuring explicit self-esteem may be architecture, but it is pretty good architecture with means we are familiar with. Measuring implicit self-esteem, on the other hand, is sincere archeology, but with equipment that still leaves things to be desired.

Notes

1. One could raise the reverse hierarchy hypothesis, namely that explicit self-esteem is closer to the core than implicit self-esteem. However, such a hypothesis is at odds with so much psychological knowledge that it cannot be seriously defended. Conscious processes are by necessity preceded by unconscious processes (at least when one maintains that consciousness resides in the brain). Hence, one cannot be conscious of an attitude (“I really like Chocolate Chip cookies”) without preceding unconscious attitudinal processes (such as positive affective reactions upon the perception of Chocolate Chip cookies). One way out would be to say that attitudes are only attitudes when they are conscious and that the core is to be found in consciousness. Such a conceptualization is possible, but it would have some undesirable consequences, the least problematic being that the current chapter would be superfluous (as implicit attitudes would not exist). However, it would also render the attitude concept rather limp as we are not that often consciously aware of our attitudes, except perhaps during communication. Of course, we are very often aware of attitude objects of course (“Ah, cookies”) but not of the attitude. In addition, the reverse hierarchy hypothesis would severely constrain the number of cases where attitudes can predict behavior, because even if we are consciously aware of an attitude, this very often happens only after we act (such as when one mindlessly reaches for Chocolate Chip cookies, and only then thinks “I’m fond of them!”).

2. It should be noted that we take the liberty here to interpret the Ricketta and Dauenheimer findings in terms of evaluative conditioning. The authors themselves favor a different explanation for their findings.

3. The order in which implicit self-esteem and explicit self-esteem were administered was counterbalanced. Order did not affect the results.

References

- Albers, L.W., Dijksterhuis, A., & Rotteveel, M. (2006). Towards Optimizing the Name Letter Test as a Measure of Implicit Self-Esteem. *Manuscript submitted for publication.*
- Baccus, J.R., Baldwin, M.W., & Packer, D.J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science, 15*, 498-502.
- Bargh, J.A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic evaluation effect. *Journal of Personality and Social Psychology, 62*, 893-912.
- Bettman, J.R., Luce, M.F., & Payne, J.W. (1998). Constructive consumer choice processes. *Journal of Consumer Research, 25*, 187-217.
- Bosson, J.K., Brown, R.P., Zeigler-Hill, V., & Swann, W.B. (2003). Self-enhancement tendencies among people with high explicit self-esteem: The moderating role of implicit self-esteem. *Self and Identity, 2*, 169-187.
- Bosson, J.K., Swann, W.B.Jr., & 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.
- Brown, J.D., & Dutton, K.A. (1995). The thrill of victory, the complexity of defeat: Self-esteem and people's emotional reactions to success and failure. *Journal of Personality and Social Psychology, 68*, 712-722.
- Cacioppo, J.T., Crites, S.L. Jr., Berntson, G.G., & Coles, M.G.H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential? *Psychological Science, 4*, 108-112.

- Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, *96*, 608-630.
- DeHart, T., Pelham, B.W., Tennen, H. (2006). What lies beneath: Parenting style and implicit self-esteem. *Journal of Experimental Social Psychology*, *42*, 1-17.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853-869.
- Di Paula, A., & Campbell, J.D. (2002). Self-esteem and persistence in the face of failure. *Journal of Personality and Social Psychology*, *83*, 711-723.
- Dodgson, P.G., & Wood, J.V. (1998). Self-esteem and the cognitive accessibility of strength and weaknesses after failure. *Journal of Personality and Social Psychology*, *75*, 178-197.
- Dovidio, J.F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510-540.
- Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology*, *86*, 345-355.
- Dijksterhuis, A., & Nordgren, L.F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, *1*, 95-109.
- Dutton, K.A., & Brown, J.D. (1997). Global self-esteem and specific self-views as determinants of people's reactions to success. *Journal of Personality and Social Psychology*, *73*, 139-148.

- Epstein, S. (1983). The unconscious, the preconscious, and the self-concept. In J. Suls & A. Greenwald (Eds.), *Psychological perspectives on the self* (Vol 2, pp. 219-247). Hillsdale, NJ. Lawrence Erlbaum.
- Fazio, R.H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M.P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 75-109). New York: Academic Press.
- Fazio, R.H., this volume
- Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., & Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.
- Greenwald, A.G., & Farnham, S.D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79, 1022-1038.
- Heatherton, T.F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, 60, 895-910.
- Hetts, J.J., Pelham, B.W. (2001). A case for the nonconscious self-concept. In G.B. Moskowitz (Ed.), *Cognitive Social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 105-124). Mahwah, NJ: Erlbaum.
- Hetts, J.J., Sakuma, M., & Pelham, B.W. (1999). Two roads to positive regard: Implicit and explicit self-evaluations and culture. *Journal of Experimental Social Psychology*, 35, 512-559.

- Hetts, J.J., & Pelham, B.W. (2003). *The ghosts of Christmas past: Reflected appraisals and the perils of near Christmas birthdays*. Manuscript in preparation.
- Jones, J.T., Pelham, B.W., Mirenberg, M.C., & Hetts, J.J. (2002). Name-letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology*, 38, 170-177.
- Jordan, Logel, Spencer, Zanna and Whitfield (this volume).
- Jordan, C.H., Spencer, S.J., Zanna, M.P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, 85, 969-978.
- Karpinski, A. (2004). Measuring self-esteem using the implicit self-association test: The role of the other. *Personality and Social Psychology Bulletin*, 30, 22-34.
- Kernis, M.H. (2003). Towards a conceptualization of optimal self-esteem. *Psychological Inquiry*, 14, 1-26.
- Kitayama, S., & Uchida, Y. (2003). Explicit self-criticism and implicit self-regard: Evaluating self and friend in two cultures. *Journal of Experimental Social Psychology*, 39, 476-482.
- 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.
- Koole, S.L., & Pelham, B.W. (2003). On the nature of implicit self-esteem: The case of the Name-letter effect. In. S. Spencer & M.P. Zanna (Eds.), *Ontario Symposium on Personality and Social Psychology* (Vol. 7).

- Koole, S.L., Smeets, K., van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology, 77*, 111-125.
- Lakoff, R.T. (1990). *Talking power: The politics of language*. New York: Basic Books.
- McFarlin, D.B., Baumeister, R.F., & Blascovich, J. (1984). On knowing when to quit: Task failure, self-esteem, advice, and nonproductive persistence. *Journal of Personality, 52*, 138-155.
- McGregor, I., & Marigold, D.C. (2003). Defensive zeal and the uncertain self: What makes you so sure? *Journal of Personality and Social Psychology, 85*, 838-852.
- Nisbett, R.E., & Wilson, T.D. (1997). Telling more than we can know: verbal reports on mental processes. *Psychological Review, 84*, 231-259.
- O' Brien, E.J., & Epstein, S. (1988). *The multidimensional self-esteem inventory*. Odessa, FL: Houghton Mifflin.
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ration-bias phenomenon. *Journal of Personality and Social Psychology, 76*, 972-987.
- Pelham, B.W., DeHart, T., & Carvallo (2003). *Implicit effects of stigmatizing names*. State University of New York at Buffalo, unpublished manuscript.
- Pelham, B.W., Koole, S.L., Hardin, C.D., Hetts, J.J., Seah, E., & DeHart, T. (2005). Gender moderates the relation between implicit and explicit self-esteem. *Journal of Experimental Social Psychology, 41*, 84-89.
- Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissism, self-esteem, and defensive self-enhancement. *Journal of Personality, 59*, 19-38.

- Riketta, M., & Dauenheimer, D. (2003). Manipulating self-esteem with subliminally presented words. *European Journal of Social Psychology, 33*, 679-699.
- Shrauger, J.S., & Rosenberg, S.E. (1970). Self-esteem and the effects of success and failure on performance. *Journal of Personality, 38*, 404-417.
- Spalding, L.R., & Hardin, C.D. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science, 10*, 535-539.
- Steele, C.M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 261-302). New York: Academic Press.
- Steele, C.M., Spencer, S.J., & Lynch, M. (1993). Self-image and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology, 64*, 885-896.
- Wentura, D., Kulfanek, M., & Greve, W. (2005). Masked affective priming by name letters: Evidence for a correspondence of explicit and implicit self-esteem. *Journal of Experimental Social Psychology, 41*, 654-663.

Table. Correlations between implicit and explicit self-esteem (and its subscales) and between the subscales as a function of condition.

| | Honesty | Control | Difference (<i>p</i> - one-tailed) |
|-------------------------------|---------|---------|--|
| Correlations with implicit SE | | | |
| Overall explicit SE | .54* | .21 | < .06 |
| Appearance | .41* | .33* | ns |
| Performance | .45* | .16 | < .09 |
| Social | .48* | .05 | < .03 |
| Correlations among subscales | | | |
| Appearance-Performance | .63* | .51* | ns |
| Appearance-Social | .70* | .42* | < .02 |
| Performance-Social | .77* | .68* | ns |

* $p = <.05$