Using Implicit Tasks in Attitude Research: a Review and a Guide

Alexa Spence
Department of Psychology and the Institute for the Study of Genetics, Biorisks and Society, University of Nottingham

Implicit measures of attitudes may overcome many biases associated with explicit measures of attitudes e.g. self presentation effects, however implicit measures of attitudes are themselves problematic in different ways. Issues raised in association with implicit measures of attitude range from methodological issues of unreliability to basic theoretical questions about what implicit measures of attitude really examine. For this reason it is important to clarify that the terms 'implicit tasks' and 'implicit measures' will be used interchangeably throughout this review in referring to implicit measures of attitude.

The aim of this review is to bring together and clarify points of controversy associated with implicit measures of attitudes in the hope that this will help to overcome the problems related to implicit tasks. This review will also be useful for those seeking to use implicit measures of attitudes as tools rather than as the focus of research. More specifically, this review will clarify the concept of 'implicit attitudes' and examine how this relates to the concept of 'explicit attitudes' as well as behaviour. The low convergent validity of implicit measures of attitude will be examined and the processes underlying different implicit tasks will be used to analyse the differences between tasks in more depth. Finally those implicit tasks emergent as most promising will be reviewed in order to give suggestions as to which tasks are the most useful in which circumstances.

Introduction

Traditionally, attitudes have been examined using explicit measures, which are direct methods of assessing attitudes. Explicit measures often take the form of statements to be responded to on semantic differential scales or questions with Likert scale response measures. The use of qualitative methods such as interviews or focus

groups can also be used and provide a richer source of information although interpretations are subjective and the procedure is more labour intensive. Explicit attitudes are those that are evident when an individual has more time to provide a considered response. The problem with explicit measures is that they may be influenced or contaminated by extraneous factors such as self-presentation biases (Rosenberg, 1969) or demand characteristics (Orne, 1962).

Implicit measures of attitudes are often considered as a less bias prone method of measuring an individuals preferences; a way of avoiding the confounding factors associated with explicit tasks[1]. These range from physiological measures, to evaluations of non-verbal behaviour, to the more widely used reaction time tasks. Here focus will be on implicit reaction time tasks including affective priming and the Implicit Association Test (IAT) as these tasks now dominate the field.

Affective priming (Fazio, Sanbonmatsu, Powell and Kardes, 1986) requires the categorisation of a target stimulus as positive or negative when it is preceded by a valenced prime stimulus. If the valence of the prime stimulus and the target stimulus match, categorisation is facilitated and can be completed faster. In this way, the valence of particular stimuli can be determined by presenting the stimulus in question with prime stimuli of different valences to examine when responses are facilitated and when these are inhibited. The IAT was developed by Greenwald, McGhee and Schwartz (1998) and examines the differential association of target stimuli with chosen attribute dimensions. The task requires a target category (e.g. insects) and a contrasting category (e.g. flowers) and utilises two contrasting attributes (e.g. pleasant and unpleasant) as two further categories. In the critical task, the participant is asked to categorise individual exemplars of each category to one of two combined target-

[1] As we shall see this is a contentious point, it may be that implicit measures and explicit measures of attitudes are actually measuring different underlying constructs.

attribute pairs requiring the same response key (e.g. flowers and pleasant). Responses are facilitated when the target is related to the attribute it is paired with and this finding enables the IAT to be used to investigate differential associations between stimuli and attributes.

Characteristics of implicitly measured attitudes

Development

It is thought that implicitly measured attitudes can develop through two main routes, a non-associative route and an associative route (Hermans, Baeyens and Eelen, 2003). The non-associative route occurs when a person simply experiences repeated exposure to a stimulus; this has been found to result in increased liking for that stimulus, referred to as the mere exposure effect (Zajonc, 1968). The associative route occurs through evaluative conditioning methods in which attitude objects are paired with other particular valence objects (or feelings) over time so that they then acquire an association with that valence. In fact it has been found that implicitly measured attitudes can be developed through a conditioning process without awareness of this happening (Olson and Fazio, 2001).

Awareness

The term 'implicit' was borrowed from cognitive psychology in which individuals are described as having implicit memory for a prior event when their performance shows evidence of the influence of this prior event although they have no explicit memory and report no awareness of the event. Characterised in this way, then, implicitly measured attitudes would be considered as attitudes for which individual's lack awareness. However, results of current implicit tasks indicate that

individuals may have some awareness of their implicitly measured attitudes. In the IAT, some participants have been found to be aware of their performance (Monteith, Voils and Ashburn-Nardo, 2001) and of what this indicated (although in some cases this might be explained by a high correspondence between implicitly and explicitly measured attitudes).

Controllability

Research has previously emphasised that implicitly measured attitudes are not able to be deliberately influenced and indeed this has been proposed as one of the advantages of this type of measure. Supporting this assertion, several researchers (e.g. Banse, Seise, and Zerbes, 2001; Asendorpf, Banse, and Mucke, 2002) have found that it is not possible to fake responses in the IAT when this is normally administered. However, participants more fully informed about the IAT and its processes may be able to influence IAT responses. Indeed Kim and Greenwald (1998) found that participants informed about the IAT were able to slow responses on certain blocks within the IAT. Steffens (2004) also found that the IAT is susceptible to faking by participants but only slightly. It seems that when participants have had experience of the IAT and when they are informed about the task, obtaining faking effects becomes more likely.

Affective in nature?

In various theoretical models, authors have characterised implicitly measured attitudes as measuring something more affective than explicitly measured attitudes (e.g., Epstein & Pacini, 1999; Marsh, Johnson, & Scott-Sheldon, 2001). It is intuitively quite appealing to consider implicitly measured attitudes as basic

'upstream' evaluations which are more emotional and only influenced by cognitive information further 'downstream' in the thought process. This has parallels with Zajonc's (1968) concept of emotion without cognition and is thought to be useful in explaining phobias and other logic defying behaviour. The possibility of implicit measures correlating with just one component of explicit attitudes might also help to explain low correlations between implicit measures and explicit measures. It would be quite elegant to be able to align the affective/cognitive attitude component distinction with the implicit/explicit attitude component distinction. However, the situation is likely to be more complex than that.

A recent study by Giner-Sorolla and Wilson (2003) investigated these predictions using an implicit priming task and the IAT along with an explicit attitude measure that divided attitude into four component parts of cognitive, hedonic affective, self conscious affective and overall evaluative. Contrary to some expectations, results showed that attitudes measured by the implicit priming task correlated best with overall evaluations and those measured by the IAT correlated best with the cognitive attitude component. In fact the hedonic affective component of attitude did not significantly predict implicitly measured attitude although it did significantly predict explicitly measured attitude indicating that it may only be at the explicit level that emotion impacts upon evaluations. Differences found between implicit tasks may be explained at an operational level in that it may be that different processes are involved in each task (Giner-Sorolla and Wilson, 2003).

The relationship between implicit and explicit measures of attitudes

How the constructs of implicitly and explicitly measured attitudes relate to each other is far from clear. Empirical examinations of correlations between implicit

and explicit measures have been wide ranging, varying extensively between domains, but are typically quite low within topics of high sensitivity (Kawakami and Dovidio, 2001; Devine, Plant, Amodio, Harmon-Jones, and Vance, 2002).

Low correlations between implicit and explicit measures may of course be partly accounted for by the very reasons why implicit tasks are utilised, due to the elimination of extraneous factors such as self-presentation effects or demand characteristics from implicit tasks that are assumed to be present in explicit tasks. One might argue that this is unable to account completely for differences though as some overlap between implicitly and explicitly measured attitudes would still be expected. We must then look to the theoretical underpinnings of these measures to examine further why these differ.

Theoretical advances

The single attitude model

The dominant view of attitudes is the single attitude model and this is the stance taken by dual process theorists. Dual process theories explain behaviour as a joint function of deliberate and spontaneous processes (see Smith and Decoster, 2000, for a review). This group of theories generally agree on the characteristics of the two systems postulated. Associative processing is described as being learned over many experiences. It occurs automatically and without awareness (although there is awareness of the results of processing) and draws on associations that are grouped by similarity and contiguity. Rule-based processing in contrast can be learned in just one or several experiences. This occurs consciously and draws on symbolically represented rules that are grouped by language and logic.

Implicitly measured attitudes are thought to examine associative processes and explicitly measured attitudes, rule-based processes. To this end it is argued that explicit and implicit attitudes are actually just different ways of measuring the same things and, in fact, for this reason should be referred to as differing measurements of attitudes rather than different attitudinal constructs. One of the main differences among models within dual process theories regards the postulated relationship between associative and rule-based processing (Smith and Decoster, 2000). Some models (e.g. Chaiken, 1980) suggest that the two types of processes occur in parallel, so both determine attitudes (although one process may dominate). Other models (e.g. Fazio, 1986) suggest that the two processes are mutually exclusive so that only one process will determine attitudes at any one time. In addition, further models (e.g. Devine, 1989) suggest that processes operate sequentially, so one process will precede the other.

Dual process theories do generally agree on the fact that implicitly measured and explicitly measured attitudes will each dominate in influencing behaviour in different situations. Indeed it has been found that explicit attitudes are more predictive of deliberative behaviour and that implicitly measured attitudes are more predictive of spontaneous behaviour. Fazio's (1990) MODE (Motivation and Opportunity as DEterminants of processing) model was developed as an integrative framework to explain when behaviour would be spontaneous, and best predicted by implicitly measured attitudes, and when behaviour would be deliberative, and best predicted by explicitly measured attitudes. As the name of the model indicates, the two factors highlighted as important influences on the type of processing undertaken are motivation and opportunity. When an individual is motivated and when the opportunity to reason carefully about a decision exists, deliberative processing of

information will occur. If an individual doesn't have the time or resources to be able to consider decision alternatives in this manner then a spontaneous, implicitly driven approach is likely to be taken.

The model of dual attitudes

The single attitude model contrasts with the model of dual attitudes. The model of dual attitudes attempts to account for the contradictory findings from implicit and explicit measures of attitudes which indicate (from the implicit literature) that attitudes are enduring and will persist over time, and also (from the explicit literature) that attitudes are changeable and depend on the situation. This view of attitudes implies that individuals can hold two different evaluations of the same attitude object at the same time (and in fact possibly more than two). This model of dual attitudes proposed by Wilson, Lindsey, and Schooler (2000) suggests that individuals may hold an implicit attitude and an explicit attitude that differ from one another. The theory states that the attitude which is endorsed at any one time will depend on the situation; if the individual has the cognitive capacity to retrieve the explicit attitude and this overrides their implicit attitude then this will dominate.

The main difference between the model of dual attitudes and the single attitude model is that the model of dual attitudes suggests that differing implicit and explicit attitudes can coexist without tension. In this way when attitude change occurs, rather than the previous attitude being replaced, it will remain in the individual's mind. However it is thought that this is not simply just a memory that the previous attitude once existed because in certain circumstances this attitude will dominate.

The Reflective-Impulsive Model

A more general theoretical approach is that of the Reflective-Impulsive model (Strack and Deutsch, 2004). This builds on dual process theories (and similarly takes a single attitude approach) but takes this theoretical position a step further and relates these to behaviour. This model suggests that the reflective and impulsive systems operate in parallel, interacting at various stages of processing. Perceived information will always be processed in the impulsive system and, depending on its intensity and the attention it is given, it may also be processed in the reflective system. The processes may then combine to determine behaviour.

In fact a final common pathway, receiving inputs from both impulsive and reflective systems is postulated to exist, and is thought to consist of behavioural schemata that may vary in levels of abstractness (e.g. Norman and Shallice, 1986). Behavioural schemata are part of the impulsive system, and can be activated by perceptual or imaginative input, but are also linked to the reflective system through the process of intending (e.g. Gollwitzer, 1999). If a behavioural schema is activated above a critical threshold, the behaviour will be carried out.

A motivational dimension is also included in this model (e.g. Cacioppo, Priester and Berntson, 1993) in that the valence of processing in the impulsive system may influence behaviour towards either approach or avoidance. In addition internal conditions may influence processing within the impulsive system so that in the occurrence of homeostatic disregulation, schemata associated with fulfilling required needs will become activated (e.g. Hull, 1943). Altogether the Reflective-Impulsive model forms an integrative framework, quite neatly combining elements from existing theories, including behavioural schemata, intentional mechanisms and motivational orientation, in order to relate mental processing with behaviour.

The relationship between implicit measures of attitudes and behaviour

Empirically, the examination of the predictive power of implicit measures of attitudes has produced mixed results with some studies indicating that implicit measures used predicted behaviour well (e.g. Frings and Wentura, 2003) and some indicating that implicit measures used did not predict behaviour well (e.g. Bosson, Swan and Pennebaker, 2000).

Generally it is found that implicit measures of attitude will predict spontaneous behaviour and explicit measures will predict deliberative behaviour (supporting Fazio's MODE model). For example, the examination of racial attitudes that utilised implicit priming measures of attitudes and self-report measures of explicit attitudes found that spontaneous, non-verbal behaviour was predicted by implicit measures and verbal behaviour was predicted by explicit measures (Dovidio, Kawakami and Gaertner, 2002). Similar results have been found with the IAT (McConnell and Leibold, 2000).

It is considered that dual process models and the dual attitude model differ in how they anticipate that implicit and explicit measures of attitude will combine in order to predict behaviour[2] (Perugini, in press). Dual process models, which postulate that spontaneous and reflective processes are mutually exclusive, imply an additive pattern of combining attitudes to predict behaviour. In contrast those dual process models (including the Reflective-Impulsive model) that postulate that processes operate in parallel through a common pathway to predict behaviour, would anticipate an interactive pattern of combining attitudes. Here implicitly and explicitly measured attitudes should combine multiplicatively to predict behaviour. Finally the model of

[2] Although each different theoretical model implies a different pattern of results, correspondence is only partial and each is flexible enough to explain each result pattern.

dual attitudes would forecast a double-dissociation pattern of combining attitudes in which implicitly measured attitudes would predict spontaneous behaviour and explicitly measured attitudes would predict deliberate behaviour.

As yet empirical results have not supported one pattern of results conclusively. Perugini (in press) examined these predictions within two different behaviours, smoking and eating snacks versus fruits. The first behaviour, smoking, was investigated with regard to two of these possible result patterns, the additive pattern and the interactive pattern, and it was found that the interactive pattern of results was supported. In this way, it was found that for neutral explicit attitudes the likelihood of smoking increases with an increasing positive implicitly measured attitude. For positive explicit attitudes, the likelihood of smoking increases sharply with an increasing positive implicitly measured attitude and for negative explicit attitudes the likelihood decreases even with an increasing positive explicit attitude.

All three possible result patterns were examined within the behaviour of eating snacks versus fruits (Perugini, in press) and for this behaviour it was found that the double dissociation pattern of results was supported whilst additive and interactive patterns were not. Here, implicitly measured attitudes clearly predicted spontaneous behaviour but not deliberative behaviour and explicitly measured attitudes predicted deliberative behaviour but not spontaneous behaviour. Results then remain inconclusive with regard to supporting one clear pattern of processes. It may be that different patterns of processes are more clearly supported in different behaviours and different contexts (Perugini, in press).

Reliability and convergent validity of different implicit tasks

One problem when measuring attitudes implicitly is deciding which task to use and because correlations between different implicit tasks have been found to be

extremely low this is a serious problem; the choice of task may completely alter findings. An investigation by Bosson et al. (2000) examined a series of implicit measures of self-esteem on their levels of validity and reliability. These included the IAT, supraliminal priming, subliminal priming, the Stroop colour-naming task and the initials and birthday preference tasks. Alarmingly it was found that the convergent validity of tasks examined, as indexed by inter-correlations, were extremely low and non-significant. There have also been several investigations of the convergent validity of the IAT with priming measures. An examination by Rudman and Kilianski (2000) of the relationship between gender and role status found that the IAT correlated with the priming measure on some, but not all, measures. Other investigations of correlations between priming measures and the IAT have typically yielded non-significant results (Fazio and Olson, 2003). However, one study that used structural equation analysis to control for measurement error in order to investigate the relationship between the IAT and priming measures found a highly significant almost perfect relationship between measures (Cunningham, Preacher and Banaji, 2001).

It seems that the lack of convergent validity between implicit tasks then is likely to be attributable, to some extent, to the low reliability of implicit tasks. Internal consistency may be a problem and this varies widely for implicit measures, alphas for conventional affective priming measures range from around zero to around 0.50 (Banse, 1999) and consistencies for IATs are generally reported at a higher level of around 0.80 (Banse, et. al. 2001). Only IATs are therefore found to be internally consistent to a satisfactory level indicating that the use of affective priming measures may be problematic in this way and that if affective priming is used, internal consistencies should always be examined.

Test re-test reliabilities are typically very low. For example, Bosson et al. (2000) found that they were unacceptably low in all implicit tasks apart from the IAT and the initials and birthday preference tasks in which these were still low, but acceptable at 0.69, 0.63 and 0.53 respectively. Low test re-test reliabilities may be explained if the construct that is being measured (e.g. attitudes) are highly changeable from one time point to the next. In this way differences may also be explained by slight changes in context or experience. However, one should consider that unstable attitude measures are problematic no matter what the cause. Measures must be consistent otherwise they have little use in predicting behavioural outcomes.

Overall, measurement error seems to be a serious problem in implicit tasks although this may be reduced by procedures such as increasing the number of trials used, introducing a response window and refining scoring techniques. Lack of convergent validity is unlikely to be completely due to measurement error though and the extremely low correlations between measures indicate they may not be examining the same psychological construct. Alternatively implicit measures may measure different aspects of the same thing and the lack of convergent validity may simply reflect the complexity of implicit attitudes.

Underlying processes

It may be that different implicit tasks are completed by utilising different mental processes. The spreading of activation account (Fazio et. al., 1986) of processes involved is one way of thinking about these tasks. This account suggests that the presentation of a valenced stimulus will activate other concepts with the same valence. If a further stimulus with a congruent valence is then presented, it will then take less time for its activation level to reach its threshold and the stimulus will be

identified more quickly than others. A stimulus with an incongruent valence on the other hand will not have received any activation and will take longer to be identified.

Conversely, a Stroop-like response conflict may also be responsible for results obtained in implicit tasks (Klauer, Robnagel, and Musch, 1997; Wentura, 1999). In this account the associated critical stimulus of a particular valence is assumed to automatically produce a tendency to give the response that is associated with it's valence, particularly when the choices of response are themselves valent in nature (i.e. 'Pleasant' or 'Unpleasant'). Therefore, when an incongruent stimulus is presented, the response activated will differ from the correct response and time will be required to resolve this response conflict. Of course, it may be that in many implicit tasks both of these processes plays a part and the process that dominates may vary from task to task. Depending on the mental processes operating to produce responses, it is quite conceivable that the results of implicit tasks may differ. Thus it would be expected that those tasks operating due to a spreading of activation process should correlate with others operating in this way and less well with those operating due to response conflict processes. This hypothesis remains to be tested experimentally.

One way of examining underlying processes involved in a task is to examine its structural make-up. De Houwer (2003a) provides a taxonomy of compatibility tasks that can be applied to implicit measures of attitudes. This distinguishes three types of compatibility: a/ relevant Stimulus-Response (S-R) compatibility, b/ irrelevant S-R compatibility and c/ Stimulus-Stimulus (S-S) compatibility. Relevant S-R compatibility refers to a task when the relevant feature of the stimulus that the participant is required to respond to is compatible, or incompatible, with the required response. Irrelevant S-R compatibility refers to a task in which an irrelevant task feature that the participant is not required to respond to is compatible, or

incompatible, with the required response. Finally S-S compatibility refers to the similarity between different features of the same stimulus or stimuli. In addition, a task may feature more than one of these types of compatibility either in a confounded, or an un-confounded, manner.

Table 1: A taxonomy of implicit tasks (adapted from De Houwer 2003a)

	Is there a manipulation of:		
Task	**S-S compatibility**	**Irrelevant S-R compatibility**	**Relevant S-R compatibility**
Affective priming	Yes	Yes	No
IAT / GNAT	No	Yes*	Yes
AST	No	Yes	No
EAST / Modified AST	No	Yes*	Yes

*But only on target concept trials

For example, affective priming features both S-S compatibility and irrelevant S-R compatibility. S-S compatibility can be noted in that the prime utilised may or may not relate to the critical stimulus presented, and irrelevant S-R compatibility is evident in that the prime may or may not have the same valence as the appropriate response required (see Table 1). With regard to the IAT (and its cousin the Go No-Go Association Task), there is a manipulation of both irrelevant S-R compatibility and relevant S-R compatibility. In the IAT, the category to which the individual stimulus belongs is the relevant feature and on some trials this will have a positive valence (e.g. flowers) and on others this will have a negative valence (e.g. insects). In addition responses are also imbued with an extrinsic valence because of the response assignments, e.g. a left key to be pressed for positive words and flowers, and in this way the relevant stimulus feature and the response can be compatible or incompatible. Individual stimulus valence is identified as an irrelevant task feature within the IAT

and similarly compatibility with response valence may vary. The Affective Simon Task (AST) and its relatives, the Extrinsic Affective Simon Task (EAST) and the modified AST will be described in further detail later. It is likely that tasks that are similar in structural make-up will produce more similar results than tasks that differ in structural make-up and this may also account for the lack of convergence between implicit measures of attitude.

Critical review of implicit tasks available

To summarise, until now we have seen that implicitly measured attitudes can be available to consciousness in some instances but are generally not controllable. They are not necessarily affective in nature as previously thought and are context dependent. It does seem that different types of implicit task may actually be measuring slightly different things, whether these be different aspects of the same attitudes or different constructs altogether which may help to account for low convergent reliabilities. Implicit tasks that have emerged as most valid and reliable are the priming measures and the IAT. It seems then that the most useful direction for research in this area is to concentrate on these two main types of task and offshoots from these.

Affective Priming Measures

Affective priming measures are one of the most common techniques to study attitudes and can be supraliminal or subliminal. Masked affective priming (Frings and Wentura, 2003), with subliminal presentation of attitude-related stimuli, is particularly useful in its unobtrusiveness. Primes in masked affective priming are presented at very fast speeds (around 28ms) and are then replaced by a mask to ensure that participants remain ignorant of the fact they have just viewed a prime stimulus.

This makes it less likely that participants will guess what is being examined in the task.

Types of priming task vary widely as do associated effect sizes and the internal reliability measures associated with the different types of task. Average effect sizes are of a medium level at around d = 0.6 (Greenwald, McGhee and Schwartz, 1998). The most reliable versions of affective priming measures seem to be variations with a response window of 200ms – 600ms within which responses must be provided, which has evidenced test re-test reliabilities of around 0.6 (Cunningham et. al., 2001). This version of evaluative priming is also known to significantly enhance the magnitude of effects found (Draine and Greenwald, 1998).

The processes involved in affective priming were originally considered to involve spreading of activation within a semantic network in a similar way to associative priming. However, more recently it has been suggested that the dominant process impacting on responses is actually a Stroop-like response conflict (De Houwer, Hermans, Rothermund and Wentura, 2002). Overall it seems that either mechanism may operate in this task depending on a number of conditions. If there are several possible response alternatives and the choice of response depends on the semantic encoding of possible targets, a spreading of activation is likely to occur. If there is a limited selection of possible responses and all targets of the same valence require the same response, a Stroop-like response conflict is likely to occur. A third postulated mechanism for the affective priming task is the affective-matching mechanism (Klauer and Musch, 2003). This assumes that the prime and the target are automatically evaluated and spontaneously compared for evaluative consistency. If the prime and target are consistent (e.g. flower and pleasant), this facilitates the

production of an affirmative response, whereas if the prime and target are inconsistent a negative response is facilitated.

It must be noted that priming measures examine evaluations at the level of individual stimuli. In this way, if several exemplars of a category are utilised within the task in order to examine the evaluation of that category, priming measures will provide an estimate of the average rating of all stimuli presented. The IAT differs in this respect in that it examines evaluations at a category level[3] and will provide an estimate of the rating of the category label itself (De Houwer, 2001, 2003a). This idea is supported by a recent study by Olson and Fazio (2004) that compared a race IAT and two versions of a traditional priming measure of racial attitudes, one when race was made a salient feature and one when this was not. A significant correlation was found between the race IAT and the priming measure when race was salient but this was non-significant when race was not made salient. This supports the theory that priming measures produce results based on individual exemplars as when participants were encouraged to consider stimuli at a category level, results corresponded with results from the IAT. For this reason priming measures have been described as more realistic than the IAT as it is asserted that category level evaluations will often not be noted in real life and it may make more sense, and may be more ecologically valid, to examine evaluations of individual exemplars (Olson and Fazio, 2004)

IAT

Since its development in 1998, the IAT has become the most widely used and widely researched implicit technique used to investigate attitudes. Its proliferation

[3] In relation to De Houwers (2003a) taxonomy, affective priming effects are mainly due to the impact of irrelevant S-R compatibility whereas IAT effects are mainly due to the impact of relevant S-R compatibility.

was largely due to its flexibility, its large effect sizes and its ease of use. Nonetheless the deluge of research that has investigated the IAT and processes involved has uncovered various possible problems with the task. These include both practical problems and conceptual problems. One practical difficulty is the fact that it is not possible to evaluate one concept in isolation; this has to be done in the context of some other concept because the task requires the categorisation of pairs of objects (Greenwald and Farnham, 2000). In this way the IAT only measures relative evaluations of concepts. For instance the repeated finding that flowers are perceived as more pleasant than insects may not mean that flowers are perceived positively and insects negatively. It may be that both flowers and insects are perceived positively but that flowers are perceived more positively than insects or it may be that both flowers and insects are perceived negatively but that insects are perceived more negatively than flowers.

Another possible problem is that participants may recode the task at hand in order to make it easier. The so-called figure-ground problem may be an issue here in that participants may only focus on one category. This would reduce the classification task to a single search task in which they respond to one category and do not respond to the other (Rothermund and Wentura, 2001; Rothermund and Wentura, in press). This of course is more likely when categories differ significantly in their salience.

An obvious consideration is the possible impact of differential familiarity of concept exemplars. Various studies were carried out by Greenwald and colleagues (Dasgupta, McGhee, Greenwald and Banaji, 2000; Greenwald, McGhee and Schwartz, 1998) that indicated that IAT effects remain when familiarity was controlled for. However these results do not rule out the possibility that a familiarity

effect occurs with this task; it proves that the IAT effect occurs over and above a familiarity effect but one cannot conclude that familiarity does not exert an effect at all.

IATs primarily measure category associations rather than associations at the level of the exemplar (so that the relevant stimulus feature carries more weight than the irrelevant stimulus feature) which indicates that differential familiarity at the level of the exemplar will not exert an effect (De Houwer, 2001). However, it was found that individual exemplar associations did exert a significant effect when category level associations were neutral (De Houwer, 2003a). Extrapolating from this it seems that differential familiarity of exemplars (the irrelevant feature) may well exert an influence on response times when included in a concept category that is neutral in valence (relevant feature). This poses a problem as IATs are often used to investigate the valence of a category; results are then difficult to interpret as these may be attributable to overall category valence or individual stimuli valence.

The use of unfamiliar categories by Brendl, Markman and Messner (2001), in their investigation of the IAT, resulted in negative evaluations of these categories. This occurred even when the unfamiliar categories were actually described as neutral or when paired with an opposing category of 'insects' (known to be perceived negatively). This may be due to the well-known finding that familiar stimuli are perceived more positively than unfamiliar stimuli (Zajonc, 1968). However, it may also be attributable to the possibility that participants used the salience of the category to recode the task. Overall it seems that familiarity is a possible confound that should be considered both at the category level and at the level of the exemplar when categories used are neutral in valence.

It is suggested that response conflict processes underlie the IAT effect noted (De Houwer, 2001). It seems that the categorisation of a stimulus will activate response representations that are associated with the same valence as the target concept. When response assignments of the target concept and attribute are compatible only one (the correct) response representation is activated. When response assignments are incompatible both the correct and the incorrect response representation will be activated and it will take longer to select the correct response.

In relation to response conflict processes, Mierke and Klauer (2001) provide evidence that suggests that on compatible tasks where attributes and concepts correspond in valence (and there is no response conflict), individuals may neglect to switch mental task set between attribute and response because basing responses on attribute features alone can allow for fast and accurate responding. Further to this it is suggested that shifts in response criteria may account for part of the IAT effect noted (Brendl, et. al. 2001). It is argued that the response threshold may be raised on incompatible trials, containing response conflicts, because these will be perceived as more difficult due to the fact that they take longer and produce more errors. These possible response strategies pose serious problems for the IAT in that these processes may contribute to what the IAT is actually measuring in addition to response conflict processes.

A more fundamental conceptual problem with the IAT is that it may include environmental associations in responses along with personal evaluations. It has been argued though that it may be difficult to separate the two (Banaji, 2001) so depending on one's conceptual leanings this may be considered more or less of a problem. This problem has recently been addressed in a variant of the IAT developed by Fazio and Olson (2004). This IAT variant apparently reduces extrapersonal associations by

utilising the attribute labels ‘I like’ and ‘I dislike’ in place of the commonly utilised ‘Pleasant’ and ‘Unpleasant’ attribute labels. Consistent with the idea that the traditional IAT includes societal knowledge in the attitude evaluation, the IAT variant examining racial attitudes found significantly less prejudice amongst white people than did the traditional IAT (Fazio and Olson, 2004). In addition the IAT variant also displayed higher correlations with explicitly measured attitudes and behavioural intentions.

Although advances have thus been made with regard to discovering the underlying processes at work in the IAT, it is also associated with a variety of problems, both methodologically and conceptually. However the IAT is one of the most valid and reliable implicit tools available to measure attitudes and remains a useful tool due to its continuously high predictive performance. One proposed modification of the IAT is the Go No-Go Association Task (GNAT) (Nosek and Banaji, 2001), which has been advanced as possibly solving some of the methodological problems associated with the IAT.

The Go No-Go Association Task (GNAT)

The GNAT essentially requires the categorisation of single attitudinal objects against different contextual backgrounds. This enables the examination of how that attitudinal object is evaluated in different contexts. A context free version is also proposed in which the contextual background used is composed of attributes only. The procedure in the context free version would require an individual to respond to target concepts and one type of attribute (e.g. positive) against a background consisting solely of the opposing attribute (e.g. negative). This would then be reversed and the individuals overall evaluation is taken as the difference in

responding[4] between the two conditions. Different contexts can be added by including other stimuli, related to the particular context required, as part of the background which the participant is not required to respond to.

The context free version of the GNAT solves one of the main practical problems associated with the IAT. This is that the IAT can only examine evaluations of a concept in the context of some other opposing concept whereas the GNAT is able to evaluate a single concept in isolation. However, the possibility that individuals may utilise certain response strategies when responding during this task remains. The figure-ground problem raised in association with the IAT is still a concern with the context free version of the GNAT, if not a greater concern. This is due to the format in which consideration of the 'figure' involves keeping two pieces of information in mind (target and attribute required to respond to) whereas consideration of the 'ground' involves keeping just one piece of information in mind (opposing attribute not required to respond to). In addition, problems with participants neglecting to shift task set and possibly shifting their response criteria on incompatible trials are likely to remain. The conceptual problems associated with the IAT also remain. Concepts are still examined at the level of the category and this may lack ecological validity. Further to this, the possibility still exists that environmental associations may be being measured by the task, rather than personal evaluative associations. However as with the IAT (Olson and Fazio, 2004) it may be possible that a GNAT variant using attribute labels of 'I like' and 'I dislike' rather than the previously used attribute labels 'Pleasant' and 'Unpleasant' may overcome this problem.

Research remains sparse on the GNAT, with little information available on its predictive validity. The little research that has so far been carried out indicates that

[4] This can be evaluated in two ways, either using errors and signal detection theory, or through differences in response latencies

the GNAT is reliable and, similarly to the IAT, more powerful than most counterparts (Nosek and Banaji, 2001). Its flexibility in being able to measure context free associations with particular concepts makes this task a potentially very useful tool.

The Extrinsic Affective Simon Task (EAST)

The EAST was developed by De Houwer (2003b) and builds on both the IAT and the Affective Simon Task. This task is similar to the GNAT in that it can examine evaluations of individual stimuli but seems to be able to further resolve some of the issues associated with these tasks. The Affective Simon Task (AST) (De Houwer and Eelen, 1998) is also a reaction time task that requires participants to respond to the stimuli on the basis of a non-affective stimulus feature (e.g. grammatical category). The response required has a positive or negative valence itself, for example, participants may be asked to say 'Good' when a noun is presented and 'Bad' when an adjective is presented. The finding is that the time that participants take to select their response is influenced by the valence of the stimuli (De Houwer, Crombez, Baeyens and Hermans, 2001). In this way it is found that the required affirmative response to the noun 'flower' would be easier to make than the required negative response to the adjective 'happy'.

Relating this to De Houwer's (2003a) taxonomy of compatibility tasks, the AST features irrelevant S-R compatibility, as the irrelevant feature of valence influences responses, but no relevant S-R compatibility or S-S compatibility (see Table 1). Structurally this task is similar to an IAT with neutral categories in which the valence of individual stimuli can be inferred. One problem of the AST is that effects can be quite small. These have been shown to be stronger when presented stimuli have a clear valence (Duschere, Holender and Molenaar, 2002) however this is

a problem, particularly if investigating the valence of stimuli in which this is not immediately clear.

The EAST builds on the AST because the responses required are not intrinsically valent; that is participants do not have to respond verbally, saying 'Good' or 'Bad'. Response valence instead is introduced by requiring participants to complete various evaluation trials in which obviously valent words have to be categorised according to their valence by pressing a key. This then results in those keys acquiring the valence of those words. This same key is also used to denote responses to the attitudinal objects on the basis of some other relevant feature but it is found that the acquired valence of the response keys will interfere with responses. Therefore, if responses are faster when denoted by the key that has acquired the positive response rather than when denoted by the key that has acquired the negative response, it is assumed that the stimulus is more strongly associated with positive attributes. For example if the stimulus 'flower' is required to be categorised as a noun using the same response key as is used for denoting a positive valence, responses will be facilitated.

Effect sizes in the EAST are greatly increased from that of the AST although still smaller in size than those of the IAT and GNAT (De Houwer, 2003b). Although the EAST seems valid, reliabilities measured so far have also been low. There are various ways in which this can be improved such as increasing the number of trials and keeping trial order constant and further testing is required to examine these possibilities.

A slightly different modification of the AST was proposed by Voss, Rothermund and Wentura (2003). This is similar to the EAST in that valence relevant trials are included which emphasise valence, however in addition responses given are

intrinsically valent as in the original AST. In this way participants have to respond by saying 'Good' or 'Bad' whether responding on the basis of valence or not. This variation may be useful in increasing effect sizes but also may make the purpose of the task more obvious. In addition, the requirement of vocal responses is more difficult to implement.

The AST, the EAST and the modified AST solve several of the problems associated with the IAT. Methodologically these tasks do not compare performances on different tasks as does the IAT but instead compare trials within the same task. This makes these less likely to be influenced by effects that may impact upon participant's response strategies including figure-ground effects, response criterion shift and task set shifting and also removes any concerns regarding familiarity effects. In addition the AST and its offshoots enable the researcher to examine single associations by themselves rather than in the context of something else (De Houwer, 2003b). Further to this stimuli can be examined at the individual level which can be argued to make the task more ecologically valid. As with each of these implicit tasks, the possibility that environmental associations are influencing results remains although again the use of more personal attribute labels such as 'I like/dislike' rather than 'good/bad' may help to overcome this.

The development of implicit tasks and which one to use?

The continual development and improvement of implicit tasks in the area of attitude evaluation has meant that these tools are increasingly valid and reliable, reinforcing the usefulness of these tasks as a tool in other areas of research rather than as a point of investigation in themselves. Of course there are still uncertainties and points of disagreement which have yet to be resolved in order to provide a clearer

understanding of processes involved. However the speed of progress in this area indicates that a more complete understanding is not far off.

In future, research should be concentrated on improving the reliability of implicit measures and on increasing the understanding the conceptual underpinnings of implicitly measured attitudes. Affective priming and the IAT are currently the most useful tasks in this area, due to their high reliability and validity, despite the many criticisms levelled at them. The GNAT and the AST and its relatives also seem like very promising tasks in the area although as yet further research is required to establish their usefulness.

With regard to investigating particular attitudes that people hold, it is suggested that a combined approach be taken utilising both explicit and implicit tasks to gain a more complete picture. Explicit tasks utilised can also be improved by minimising demand characteristics and including measures of social desirability in order to eliminate these influences as sources of variation.

Conclusion

Overall it seems that at the most basic conceptual level, disagreements remain as to the nature of implicitly measured attitudes. The prevailing view seems to be that implicit attitudes and explicit attitudes are actually different measures of the same thing. However, the model of dual attitudes suggests that implicit and explicit attitudes can be separate constructs that exist alongside each other. This view is coherent and as yet there is little evidence that conclusively proves or disproves either model. The relationship between explicit and implicit attitudes is therefore hugely thought provoking and much researched. Empirical reports of correlations between explicitly and implicitly measured attitudes are varied but in some circumstances are

very low. There are a variety of reasons suggested for this, including of course the very reason that implicit tasks are heralded, namely the assertion that these can bypass confounding factors of self-presentation effects. Further to this it is suggested that low correlations between explicit and implicit tasks may be largely due to the low reliabilities of implicit tasks. In addition, there are concerns over the convergent validity of implicit tasks and the possibility that different tasks are measuring different things; possibly different aspects of the same construct, or even different constructs. Implicit tasks that have emerged as most valid and reliable are priming measures and the IAT. It seems then that the most useful direction for research in this area is to concentrate on these two main types of task and offshoots from these. It is recommended that future research should concentrate on improving the reliability of implicit measures and clarifying their conceptual basis. This will ultimately then clarify attitudinal structure and may help to relate attitudes and behaviour.

References

Asendorpf, J. B. Banse, R. and Mucke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behaviour. *Journal of Personality and Social Psychology*. 83: 380-393.

Banaji, M. (2001). Implicit attitudes can be measured. In H. Roediger and J. Nairne. *The Nature of Remembering: Essays in Honour of Robert G. Crowder*. Washington DC: American Psychological Association.

Banse, R. (1999). Automatic evaluation of self and significant others: Affective priming in close relationships. *Journal of Social and Personal Relationships*. 16: 803-821.

Banse, R., Seise, J., and Zerbes, N. (2001). Implicit attitudes towards homosexuality: reliability, validity and controllability of the IAT. *Zeitshrift fur Experimentelle Psychologie.* 48 (2): 145-160.

Bosson, J., Swann, W., and Pennebaker, J. (2000). Stalking the perfect measure of implicit self-esteem: the blind men and the elephant revisited? *Journal of Personality and Social Psychology*. 79: 631-43.

Brendl, C., Markman, A., and Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association test. *Journal of Personality and Social Psychology*. 81: 5: 760-773.

Cacioppo, J., Priester, J. and Berntson, G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*. 65: 5-17.

Chaiken, S. (1980). Heuristic versus Systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*. 39: 752-766.

Cunningham, W., Preacher, K., and Banaji, M. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*. 12: 163-70.

Dasgupta, N., McGhee, D., Greenwald, A. and Banaji, M. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*. 36: 316-328.

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). A structural analysis of indirect measures of attitudes. In J. Musch and D. C. Klauer (Eds.). *The psychology of evaluation: Affective processes in cognition and emotion*. NJ: Laurence Erlbaum.

De Houwer, J. (2003b). The Extrinsic Affective Simon Task. *Experimental Psychology*. 50 (2): 77-85.

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

De Houwer, J., Crombez, G., Baeyens, F., and Hermans, D. (2001). On the generality of the affective Simon effect. *Cognition and Emotion*. 15: 189-206.

De Houwer, J., Hermans, D., Rothermund, K., and Wentura, D., (2002). Affective priming of semantic categorisation responses. *Cognition and Emotion*. 16: 5: 643-666.

Devine, P. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology*. 56: 5-18.

Devine, P., Plant, E., Amodio, D., Harmon-Jones, E., Vance, S. (2002). The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*. 82: 835-848.

Dovidio, J. Kawakami, K. and Gaertner, S. (2002). Implicit and explicit prejudice and interracial interactions. *Journal of Personality and Social Psychology*. 82: 62-68.

Draine, S. and Greenwald, A. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*. 127: 286-303.

Epstein, S. and Pacini, R. (1999). Some basic issues regarding dual process theories from the perspective of cognitive-experiential self-theory. Automaticity and control in stereotyping. In S. Chaiken and Y. Trope (Eds.), *Dual process theories in social psychology*. New York: Guilford.

Fazio, R. (1986). How do attitudes guide behaviour? In R. M. Sorrentino and E. T. Higgins. (Eds.) *Handbook of motivation and cognition*. New York: Guilford

Fazio, R. (1990). Multiple processes by which attitudes guide behaviour: the MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology.* 23: 75-109. New York: Academic Press.

Fazio, R., Sanbonmatsu, D., Powell, M., and Kardes, E. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*. 50: 229-238.

Fazio, R and Olson, M. (2003). Implicit measures in social cognition research: their meaning and use. *Annual review of psychology*. 54: 297-327.

Fazio, R. and Olson, M. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*. 86: 5: 653-667.

Frings, C. and Wentura, D. (2003). Who is watching Big Brother? TV consumption predicted by masked affective priming. *European Journal of Social Psychology*. 33: 779-791.

Giner-Sorolla, R. and Wilson, T. (2003). Implicit attitudes and explicit attitude components. *Presentation given at BPS Social Section Annual Conference. September*.

Gollwitzer, P. (1999). Implementation Intentions. Strong effects of simple plans. *American Psychologist.* 54: 493-503.

Greenwald, A., McGhee, D., and Schwartz, J. (1998). Measuring individual differences in implicit cognition. *Journal of Personality and Social Psychology*. 74: 1464-1480.

Greenwald, A. and Farnham, S., (2000). Using the Implicit Association Task to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*. 79: 1022-1038.

Hermans, D., Baeyens, F. and Eelen, P. (2003). On the Acquisition and Activation of Evaluative Information in Memory: The study of evaluative learning and affective priming combined. In J. Musch and D. C. Klauer (Eds.). *The psychology of evaluation: Affective processes in cognition and emotion*. NJ: Laurence Erlbaum.

Hull, C. (1943). *Principles of behaviour*. New York: Appleton-Century-Crofts.

Kawakami, K. and Dovidio, J. (2001). The reliability of implicit stereotyping? Or confronting prejudice: effects of prejudice labelling on stereotype activation. *Personality and Social Psychology Bulletin*. 27: 212-225.

Kim, D. and Greenwald, A. (1998). Voluntary controllability of implicit cognition: can implicit attitudes be faked? In A. Greenwald and B. Nosek. Health of the Implicit Association Test at Age 3. *Zeitschrif fur Experimentelle Psychologie*. 48: 2: 85-93.

Klauer, K. and Musch, J. (2003). Affective priming: Findings and Theories. In J. Musch and D. C. Klauer (Eds.). *The psychology of evaluation: Affective processes in cognition and emotion*. NJ: Laurence Erlbaum.

Klauer, K., Robnagel, C., and Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 23: 246-255.

Marsh, K., Johnson, B., Scott-Sheldon, L. (2001). Heart versus reason in condom use: implicit versus explicit attitudinal predictors of sexual behaviour. *Zeitshrift fur Experimentelle Psychologie. 48: 2: 161-75.*

McConnell, A. and Liebold, J. (2001). Relations between the Implicit Association Test, explicit racial attitudes and discriminatory behaviour. *Journal of Experimental and Social Psychology*. 37: 435-442.

Mierke, J. and Klauer, K. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift fur Experimentelle Psychology*. 48: 107-122.

Monteith, M., Voils, C., and Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting and reacting to implicit racial biases. *Social Cognition*: 19 (4): 395-417.

Norman, D. and Shallice, T. (1986). Attention to action. Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, and D. Shapiro (Eds.). *Consciousness and self regulation: Advances in research*. New York: Plenum Press.

Nosek, B., and Banaji, M. (2001). The go/no-go association task. *Social Cognition*. 19: 625-666.

Orne, M. T. (1962). On the social psychology of the psychology experiment: With particular reference to demand characteristics and their implications. *American Psychologist*. 17: 776-83.

Olson, M and Fazio, R. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*. 12: 5: 413-417.

Olson, M. and Fazio, R. (2004). Trait inferences as a function of automatically-activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology*. 26. 1-11.

Perugini, M. (in press). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*.

Rosenberg, M., J., (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal and R. L. Rosnow (eds.). *Artefact in Behavioural Research*. New York: Academic Press.

Rothermund, K. and Wentura, D.(2001). Figure-ground asymmetries in the Implicit Association Test. *Zeitschrift fur Experimentelle Psychologie*. 48: 94-106.

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

Rudman, L. and Kilianski, S. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social psychological bulletin*. 26: 1315-1328.

Smith, E. R. and Decoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*. 4: 108-131.

Steffens, M. C. (2004). Is the Implicit Association Test Immune to Faking? *Experimental Psychology.* 51(x): 1-15.

Strack, F. and Deutsch, R. (2004). Reflective and Impulsive Determinants of *Social Behaviour*. *Personality and Social Psychology Review*. 8 (3): 220-247.

Voss, A., Rothermund, K., and Wentura, D. (2003). Estimating the valence of single stimuli: A new variant of the Affective Simon Task. *Experimental Psychology*. 50 (2): 86-96.

Wentura, D. (1999). Activation and inhibition of affective information: Evidence for negative priming in the evaluation task. *Cognition and Emotion*. 13: 65-91.

Wilson, T. Lindsay, S. and Schooler, T (2000). A model of dual attitudes. *Psychological Review*. 107 (1): 101-126.

Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology. 9: 1-27.*

Author Note

This research was supported by the Economic and Social Research Council and the Natural Environment Research Council. I am grateful to Dr. Ellen Townsend, Dr. Marco Perugini and Prof. Jan De Houwer for their comments.

Correspondence concerning this article should be addressed to the author:

Alexa Spence

School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD

E-mail: lpxas@psychology.nottingham.ac.uk

Fax: +44 (0)1159515324

Tel: +44 (0)1158467402