L'IAT (Implicit Association Test) ou la mesure des cognitions sociales implicites : Revue critique de la validité et des fondements théoriques des scores qu'il produit

Christophe Blaison, Delphine Chassard, Jean-Luc Kop et Kamel Gana*
Laboratoire de Psychologie Clinique et Cognitive, Laboratoire de Psychologie Lorrain, et Groupe d'Analyse Psychométrique des Conduites – Université Nancy 2

RÉSUMÉ

Récemment, dans le but de remédier aux limites des mesures par questionnaires (biais d'auto-présentation, capacités introspectives limitées), plusieurs mesures dites indirectes ont été développées. Parmi elles, le test des associations implicites (IAT, Greenwald, McGhee et Schwartz, 1998) est celui qui a suscité le plus d'intérêt et le plus de travaux : il est suffisamment flexible pour mesurer des concepts variés (attitude, personnalité, stéréotypes…), il fournit des scores fidèles et offre des résultats encourageants en termes de validité critérielle. Mais la validité de construit des scores de l'IAT est contestée, notamment parce que leur interprétation n'est pas univoque. Il en est de même pour l'interprétation des dissociations observées entre mesures directes (questionnaires) et mesures indirectes qui, en l'absence d'un cadre théorique solide, donne lieu à de multiples débats. Ces défauts de jeunesse des mesures indirectes ont toutefois le mérite de stimuler la créativité des chercheurs et d'offrir de nouvelles perspectives qui devraient déboucher sur de nouveaux outils, mieux fondés théoriquement.

Implicit Association Test or the measure of implicit social cognition: a critical review of the validity and the theoretical basement of its scores

ABSTRACT

In order to remedy the limits of self-report measures (self-presentational biases, introspective limits…), several indirect measures have been developed. The Implicit Association Test (IAT, Greenwald, McGhee et Schwartz, 1998) is the one that has concentrated most interest and research: it is flexible enough to measure a broad range of constructs

*Pr. Kamel Gana, Université Nancy 2, 3 Place Godefroy de Bouillon, 54015 Nancy Cedex.
Kamel.gana@univ-nancy2.fr
Notes des auteurs : l'ordre entre des deux premiers auteurs a été déterminé par tirage au sort.

(attitude, personality, stereotypes…), and it produces reliable and encouraging criteria valid scores. But the construct validity of the IAT scores remains controversial because their interpretation is quite ambiguous. In the same vein, because of the lack of any strong theoretical background, the interpretation of the observed dissociations between direct (self-report) and indirect measures remains problematic. This leads to multiple discussions. These "youthful limits" however stimulate researchers' creativity and open new perspectives which should lead to more theoretically grounded measures.

1. INTRODUCTION

Les construits propres à la cognition sociale tels que les attitudes, les stéréotypes, l'estime de soi ou le concept de soi sont le plus souvent appréhendés par des mesures auto-rapportées (*i.e.*, directes). Les limites de ce type de mesures sont bien connues : elles dépendent à la fois de ce que les sujets veulent bien nous dire (*i.e.*, biais d'auto-présentation) et de ce qu'ils peuvent nous dire (*i.e.*, limites des capacités introspectives). Récemment, dans le but de remédier à ces limites, plusieurs mesures de type indirect, dont *l'Implicit Association Test* (ou IAT) (1) (Greenwald, McGhee et Schwartz, 1998), ont été développées et l'intérêt à leur égard n'a cessé d'augmenter.

Ces mesures s'inscrivent dans le domaine de la cognition sociale implicite (Devos et Banaji, 2003) dont on verra les principaux paradigmes dans la première partie de cet article, ce qui nous permettra de mieux comprendre le contexte dans lequel l'IAT a été créé.

Dans les deux parties suivantes, notre objectif sera de présenter une synthèse des travaux portant tant sur les caractéristiques méthodologiques de l'IAT (partie 2) que sur ses fondements théoriques (partie 3), avec pour objectif de fournir aux lecteurs francophones une vue d'ensemble des débats soulevés par l'IAT, ce qui devrait aider à une utilisation appropriée de ce nouvel instrument.

Dans la seconde partie, nous présenterons donc en détail l'IAT parce que c'est celui qui a suscité le plus d'intérêt, d'enthousiasme mais aussi le plus de critiques dès sa première publication. Nous évoquerons le principe sur lequel repose l'IAT ainsi que ses propriétés psychométriques**.** Les limites potentielles de l'IAT seront ensuite abordées à travers les critiques théoriques et méthodologiques qui lui ont été adressées avant de présenter les nouveaux instruments développés pour tenter d'y remédier.

Dans la troisième et dernière partie, nous reviendrons plus spécifiquement sur les problèmes théoriques posés par l'IAT et notamment sur les difficultés soulevées par les études de validité, celles-ci conduisant à s'interroger sur la nature même des construits appréhendés par cet outil.

2. LA COGNITION SOCIALE IMPLICITE ET SON OPÉRATIONNALISATION

Les avancées scientifiques dans le domaine de la cognition humaine ont permis à la psychologie de commencer à étudier de façon prometteuse, d'une part, les processus mentaux inconscients (Jacoby et Kelley, 1987 ; Jacoby, Toth et Yonelinas, 1993), voire, d'autre part, l'"inconscient psychologique" des individus (Kihlstrom, Barnhardt et Tataryn 1992 ; Kihlstrom, Mulvaney, Tobias et Tobis, 2000). Derrière cette expression d'"inconscient psychologique", se cache l'idée selon laquelle nos pensées, notre expérience et nos actes conscients pourraient être influencés par des perceptions, des souvenirs et autres contenus mentaux dont nous n'aurions pas conscience et qui seraient indépendants de tout contrôle volontaire. Si Kihlstrom *et al.* (2000) parlent volontiers d'« inconscient psychologique », d'autres préfèrent à ces termes ceux de cognition implicite, postulant que nos comportements ainsi que nos jugements peuvent être influencés par des expériences passées sans que nous n'ayons conscience de cette influence ni que nous nous rappelions cette expérience (Greenwald et Banaji, 1995). La notion de cognition implicite marque une avancée supplémentaire dans le domaine de la cognition par rapport à celle – plus ancienne – d'automaticité selon laquelle certains de nos savoir-faire moteurs et cognitifs peuvent, grâce à une pratique intensive, devenir automatiques et donc rendre les procédures ou les opérations sous-jacentes inaccessibles à l'introspection (Kihlstrom *et al.*, 2000). En effet, alors que la notion d'automaticité fait exclusivement référence à des *processus* mentaux dont nous n'aurions pas conscience, celle de cognition implicite laisse entrevoir la possibilité d'étendre cette idée aux *contenus* mentaux associés à ces processus.

Depuis ces dernières années, le concept de cognition implicite a largement été diffusé. En témoigne le nombre important de travaux publiés portant sur la mémoire implicite, l'apprentissage implicite, la perception implicite ou sur les pensées implicites (cf. Kihlstrom *et al.*, 2000).

En 1995, Greenwald et Banaji proposent de s'intéresser plus spécifiquement à l'étude des différences individuelles dans le domaine de la cognition sociale implicite, notamment aux attitudes, aux stéréotypes, à l'estime de soi et, plus tard (Greenwald *et al.*, 1998), au concept de soi. De façon générale, ces auteurs définissent un construit implicite comme une trace de notre expérience passée non identifiable de façon introspective (ou identifiée de manière imprécise, voire incorrecte) capable d'influencer nos sentiments, nos pensées et nos actions envers divers

objets sociaux (Greenwald et Banaji, 1995, p. 5). Dès lors, l'étude des différences individuelles dans le domaine de la cognition sociale implicite se heurte au problème de la mesure des construits implicites. Si l'on considère effectivement que ces derniers sont des traces en mémoire non identifiables de façon introspective, il n'est pas envisageable de les appréhender à l'aide de mesures auto-rapportées (*i.e.*, directes) puisque celles-ci dépendent des limites des capacités introspectives des sujets (Nisbett et Wilson, 1977). Il faut donc avoir recours à des instruments qui ne requièrent pas de réponses auto-rapportées (Greenwald et Banaji, 1995), autrement dit à des mesures appelées indirectes.

S'inspirant de toute sorte de paradigmes expérimentaux, Greenwald (cf. Dasgupta, Greenwald et Banaji, 2003) s'est alors lancé dans une série d'études ayant pour but de mettre au point un instrument de mesure susceptible d'appréhender des construits implicites tout en produisant des scores satisfaisants d'un point de vue psychométrique. De cette série d'études est né l'*Implicit Association Test* (ou IAT ; Greenwald *et al.*, 1998). L'IAT emprunte le paradigme expérimental classique des temps de réponse (« *response time paradigm* ») dont le *Stroop*, le *Simon Task* ou l'amorçage sémantique sont l'illustration (Chassard et Kop, 2003), et qui stipule que les temps de réponse d'un sujet à un certain matériel dépendent des structures cognitives spécifiques qu'il entretient à propos du monde ou de lui-même. Plus tard, ce paradigme a été adapté pour être utilisé dans des champs d'étude plus conatifs avec des stimuli chargés émotionnellement, permettant ainsi l'étude des processus automatiques d'évaluation (Musch et Klauer, 2003) comme, par exemple, le *Stroop* émotionnel (Pratto et John, 1991) ou le paradigme d'amorçage affectif (Fazio, Sanbonmatsu, Powell et Kardes, 1986). Dans sa forme originale, le paradigme d'amorçage affectif montre que les sujets traitent plus rapidement une cible affectivement polarisée quand elle est précédée d'une amorce de même valence (*e.g.*, « amour » précédé de « joie ») que lorsqu'elle est précédée d'une amorce de valence opposée (*e.g.*, « amour » précédé de « torture »). Enfin, certains chercheurs (*e.g.*, Banse, 1999) ont eu l'idée d'utiliser la variabilité interindividuelle relevée dans ces nouvelles tâches pour en faire des mesures – au sens psychométrique du terme – indirectes d'attitudes, la logique étant que si les sujets ont une attitude positive envers un objet présenté comme amorce, ils doivent identifier plus rapidement la valence de cibles positives que la valence de cibles négatives et inversement s'ils ont une attitude négative. Cependant, la fidélité observée pour de telles mesures restant très faible (Bosson, Swann et Pennebaker, 2000), les tentatives de mesure des différences individuelles sont restées peu satisfaisantes.

C'est donc principalement pour remédier à ce problème que l'IAT a été développé (Greenwald *et al.*, 1998).

3. LE TEST DES ASSOCIATIONS IMPLICITES (IAT)

3.1. Principes de base et modalités d'administration

Nous empruntons à Greenwald et Farnham (2000) une illustration évocatrice du principe sur lequel repose l'IAT. Supposez que vous ayez à trier un jeu de 52 cartes en deux tas situés à votre gauche et à votre droite. Il vous sera probablement plus facile de classer les piques et les trèfles d'un côté et les cœurs et carreaux de l'autre, que de classer piques et cœurs d'un côté et trèfles et carreaux de l'autre. Cette facilité accrue est due au fait que les piques et les trèfles d'une part et les cœurs et les carreaux d'autre part "vont bien ensemble" : ils partagent respectivement un même attribut, ici la couleur. S'inspirant de cette idée, le principe général de l'IAT repose sur le fait qu'il est plus facile de classer ensemble des items cognitifs lorsque les groupements à réaliser sont cohérents avec ceux que l'on adopte spontanément du fait de notre manière particulière d'organiser l'information. Par exemple, par rapport à la population générale, un joueur de bridge s'empresserait d'infirmer l'observation faite plus haut, car pour lui ce sont d'une part les cœurs et les piques et de l'autre les carreaux et les trèfles qui vont le mieux ensemble ; pour ce joueur, l'association entre les familles de cartes n'est plus spontanément créée par la couleur mais par un autre attribut acquis à travers la pratique du bridge. L'IAT se veut donc être une *méthode de mesure indirecte de la force relative des associations entre différents concepts* (Greenwald, Banaji, Rudman, Farnham, Nosek et Mellott, 2002).

Afin de décrire plus précisément les applications de l'IAT, nous prendrons l'exemple d'un IAT censé mesurer l'estime de soi. Par son entremise, on mesure la force d'association entre le concept de soi et les deux valences affectives positive et négative. Cette application fait intervenir deux concepts dits cibles : « Moi » *vs.* « Pas moi » et deux concepts dits attributs : « Agréable » *vs.* « Désagréable ». Le sujet de l'expérience est assis en face d'un ordinateur muni d'un clavier dont il ne doit utiliser que deux touches, l'une à gauche, l'autre à droite. La procédure comporte cinq étapes dans lesquelles on lui demande de classer aussi vite que possible, tout en faisant le minimum d'erreurs, des stimuli exemplifiant les

concepts-cibles et les concepts-attributs. Le tableau I résume le contenu de ces cinq étapes (ou blocs d'items) détaillées ci-dessous.

Étape 1 – Des items appartenant aux deux concepts-cibles apparaissent les uns après les autres sur l'écran. Lorsque les items renvoient au concept-cible « Moi » (*e.g.*, « je », « mien », « mon »), le sujet doit appuyer sur une touche du côté droit du clavier ; lorsqu'ils renvoient au concept-cible « Pas moi » (*e.g.*, « ils », « leurs », « il »), le sujet doit appuyer sur une touche située du côté gauche du clavier.

Étape 2 – Les sujets doivent classer des items appartenant aux deux concepts-attributs. Lorsqu'ils renvoient au concept-attribut « Agréable » (*e.g.*, « sourire », « plaisir », « joie »), le sujet doit appuyer sur une touche du côté droit du clavier ; lorsqu'ils renvoient au concept-attribut « Désagréable » (*e.g.*, « mort », « poison », « vomi »), le sujet doit appuyer sur une touche située du côté gauche du clavier.

Étape 3 (1er bloc-test) – Les deux tâches précédentes sont présentées conjointement, les items à classer sont donc soit des exemplaires des concepts-cibles (*e.g.*, « je », « ils ») soit des exemplaires de concepts-attributs (*e.g.*, « sourire », « mort »). Lorsqu'ils renvoient soit au concept-cible « Moi », soit au concept-attribut « Agréable », le sujet doit appuyer sur une touche du côté droit du clavier ; lorsqu'ils renvoient soit au concept-cible « Pas moi », soit au concept-attribut « Désagréable », le sujet doit appuyer sur une touche située du côté gauche du clavier.

Étape 4 – Les sujets doivent à nouveau classer des items appartenant aux deux concepts-cibles (cf. étape 1) mais les touches à utiliser pour classer les exemplaires sont inversées (gauche pour « Moi », droite pour « Pas moi »).

Étape 5 (2e bloc-test) – Cette dernière étape repose sur le même principe que la 3e étape mais les sujets doivent cette fois appuyer sur une touche située du côté droit du clavier lorsque apparaissent des items renvoyant soit au concept-cible « Pas Moi », soit au concept-attribut « Agréable » ; et du côté gauche lorsque apparaissent des items renvoyant soit au concept-cible « Moi », soit au concept-attribut « Désagréable ».

La mesure résultante de la procédure (*IAT effect* ou effet IAT) repose sur la comparaison entre la force combinée des associations « Pas moi – Désagréable » et « Moi – Agréable » (1er bloc-test, étape 3) et la force combinée des associations « Moi – Désagréable » et « Pas moi – Agréable » (2e bloc-test, étape 5). Seules les étapes 3 et 5 sont donc prises en compte lors de l'analyse des résultats, *l'effet IAT s'exprimant par la différence de temps de réponses entre les deux blocs-tests.* Si un sujet répond plus rapidement aux items du premier bloc-test qu'aux items du second bloc-test, on en déduit qu'il associe plus fortement « Moi » et « Agréable » que « Moi » et « Désa-

Tableau I. Illustration des cinq étapes d'un IAT d'estime de soi
Table I. Five constitutive steps of a self-esteem IAT

Étapes/Blocs	Presser Touche gauche	Presser Touche droite
1	Pas moi	Moi
2	Désagréable	Agréable
3 – 1er bloc-test	*Pas moi ou Désagréable*	*Moi ou Agréable*
4	Moi	Pas moi
5 – 2e bloc-test	*Moi ou Désagréable*	*Pas moi ou Agréable*

gréable » ; et on en infère qu'il possède une estime de soi plus élevée par rapport à un sujet qui est plus rapide au second bloc-test qu'au premier.
L'IAT est capable d'évaluer d'autres caractéristiques psychologiques, car son mode de construction en fait un outil très flexible. D'une part, les stimuli servant à sa construction peuvent être langagiers, picturaux, sonores ou des combinaisons de ces différentes modalités. D'autre part, le couplage judicieux de certains concepts-cibles et de certains concepts-attributs, permet théoriquement d'évaluer des attitudes, des stéréotypes ou des caractéristiques de personnalité (Greenwald *et al.*, 2002). Par exemple, si l'on définit une attitude comme l'association entre un objet social et un attribut valencé (positif ou négatif) (Fazio, Williams et Powell, 2000 ; Greenwald *et al.*, 2002), il est possible grâce à l'IAT de mesurer l'attitude envers des individus de couleur en sélectionnant les concepts-cibles « Blancs » *vs.* « Noirs » et les concepts-attributs « Agréable » *vs.* « Désagréable ». Si un sujet est plus rapide dans le bloc où les concepts « Blancs » et « Agréable » (respectivement, « Noirs » et « Désagréable ») sont associés à une même touche de réponse que dans celui où sont associés « Noirs » et « Agréable » (respectivement, « Blancs » et « Désagréable »), alors on en infère qu'il possède une attitude positive envers les Blancs et négative envers les Noirs. En poursuivant cette stratégie, on peut envisager un stéréotype comme l'association entre un groupe social et un ou plusieurs attributs non-valencés (*e.g.*, cibles : « Hommes » *vs.* « Femmes », attributs : « Mathématiques » *vs.* « Arts »). Enfin, une caractéristique personnelle de personnalité peut être assimilée à l'association entre le concept de soi et un attribut non-valencé (*e.g.*, cibles : « Moi » *vs.* « Pas moi », attributs : « Introverti » *vs.* « Extraverti »).

Cette flexibilité a permis d'étendre l'utilisation de l'IAT à de nombreux domaines. Il existe par exemple des études sur les attitudes (*e.g.*, Karpinski et Hilton, 2001 ; Jelenec et Steffens, 2002 ; Palfai et Ostafin, 2003 ; DeJong, Van den Hout, Rietbroek et Huijding, 2003), les stéréotypes (*e.g.*, Nosek, Banaji et Greenwald, 2002a ; Nosek, Banaji et Greenwald, 2002b), la catégorisation sociale (Piontkowski, Blanz, Rohman, Schmermund et Florack, 2001), l'estime de soi (*e.g.*, Bosson *et al.*, 2000 ; Greenwald et Farnham, 2000), la timidité (*e.g.*, Asendorpf, Banse et Mücke, 2002), l'affectivité positive et négative (Blaison et Gana, 2004), les cinq dimensions de personnalité (OCEAN) (Steffens, 2004), la dépression (*e.g.*, Gemar, Segal, Sagrati et Kennedy, 2001), l'anxiété (*e.g,* Egloff et Schmukle, 2002), l'anxiété sociale (De Jong, 2002), la phobie des animaux (Teachman, Gregg et Woody, 2001), le conditionnement évaluatif (Mitchell, Anderson et Lovibond, 2003a).

3.2. Fidélité

Développé principalement dans le but de remédier au manque de fidélité des scores obtenus par des outils tels que le *Stroop* émotionnel ou l'amorçage affectif, l'IAT a réussi son pari. Comme il permet d'obtenir des tailles d'effet importantes – dès 1998, Greenwald *et al.* rapportent une taille d'effet IAT ($d \approx 1.21$)(2) deux fois supérieure à celle de l'amorçage ($d \approx 0.62$) – il devient possible d'envisager une mesure offrant des scores fidèles des différences individuelles, ce que l'on constate effectivement avec l'IAT. Quel que soit l'objet mesuré, l'IAT montre une bonne consistance interne (α de Cronbach ≈ 0.80), comparable à celle des mesures directes correspondantes (Bosson *et al.*, 2000 ; Banse, Seise et Zerbes, 2001 ; Cunningham, Preacher et Banaji, 2001 ; Egloff et Schmukle, 2002) (3). Sa stabilité temporelle (fidélité test-retest), bien que modérée ($r \approx 0.60$), dépasse largement celle d'autres mesures indirectes (Bosson *et al.*, 2000 ; Dasgupta et Greenwald, 2001 ; Greenwald et Nosek, 2001).

3.3. Éléments de validation

Les mesures utilisant l'IAT permettent généralement d'obtenir, au niveau du groupe, des différences d'attitude conformes à ce qui est attendu. Ainsi, un favoritisme envers l'endogroupe a pu être détecté, et ce, aussi bien avec des groupes réels (Greenwald *et al.*, 1998) qu'en utilisant le paradigme des groupes minimaux (Ashburn-Nardo, Voils et Monteith, 2001). Hummert, Gartska, O'Brien, Greenwald et Mellott (2002) ont

montré que les personnes âgées avaient, comparativement à de jeunes adultes, une attitude plus favorable envers les jeunes et une estime de soi plus élevée. Nosek *et al.* (2002a) ont montré que les garçons avaient une attitude plus favorable envers les disciplines scientifiques que les filles. Gemar *et al.* (2001) ont montré que des sujets récemment sortis de dépression avaient une attitude plus négative envers eux-mêmes que des sujets sains. Dans un autre domaine, Teachman *et al.* (2001) ont pu différencier des sujets arachnophobes de sujets ayant la phobie des serpents : les sujets arachnophobes ont une attitude plus négative envers les araignées et plus positive envers les serpents alors que c'est l'inverse chez les sujets ayant la phobie des serpents. Citons aussi DeJong (2002), qui a montré que des sujets anxieux avaient une estime de soi plus faible que des sujets non-anxieux, Swanson, Rudman et Greenwald (2001) qui ont établi que des végétariens avaient une attitude plus positive envers les légumes et les fruits et plus négative envers la viande que des non-végétariens ou Banse *et al.* (2001) qui ont mis en évidence, chez des sujets homosexuels, comparativement à des hétérosexuels, une attitude plus positive envers l'homosexualité, etc.

Les résultats concernant la validité critérielle des scores des différentes applications de l'IAT sont, eux aussi encourageants (Poehlman, Uhlman, Greenwald et Banaji, 2005) même s'ils restent peu nombreux. Ainsi, Phelps *et al.* (2000) ont pu mettre en évidence une relation entre une mesure IAT d'attitudes raciales (« Blancs » *vs.* « Noirs ») et l'intensité de l'activation de l'amygdale (évaluée par IRMf) lorsqu'un visage blanc ou noir est présenté au sujet. Greenwald et Farnham (2000) ont montré qu'un IAT d'estime de soi prédisait les réactions des sujets après un feed-back positif ou négatif (les sujets ayant une forte estime de soi semblent moins affectés par un feed-back négatif). Asendorpf *et al.* (2002) ont, quant à eux, observé qu'un IAT de timidité prédisait davantage les comportements de timidité spontanée (*e.g.*, position et tension corporelle) qu'une mesure directe de timidité, alors que cette dernière est davantage corrélée aux comportements contrôlés de timidité (*e.g.,* durée d'une prise de parole en public).

Tout comme d'autres mesures élaborées essentiellement sur des bases empiriques, celles utilisant l'IAT connaissent de sérieuses difficultés lorsqu'il s'agit d'apprécier la validité de construit des scores qu'elles produisent. Dès les premières applications de l'IAT, on a cherché à établir le niveau de relation entre la mesure directe d'un concept et la mesure indirecte de ce même concept (voir, par exemple, Greenwald et Nosek, 2001 ; Nosek *et al.*, 2002b ; Fazio et Olson, 2003). Dans l'ensemble, les corrélations obtenues sont plutôt faibles ($r < .30$; Hofman, Gawronski,

Gschwendner, Le et Schmitt, 2005), même si l'on a pu constater quelques exceptions pour certains construits. L'interprétation de ces dissociations entre mesures directes et mesures indirectes reste toutefois ambiguë puisqu'on peut soit retenir ce résultat comme montrant la faillite du paradigme (pour absence de validité convergente), soit l'utiliser pour attester de l'absence de contamination des mesures indirectes par les biais inhérents aux mesures directes. Rapidement esquissée, cette controverse amène à s'interroger sur ce que mesure réellement un paradigme comme l'IAT. C'est à cette question qu'est consacrée la majeure partie de la suite de cet article.

3.4. Les limites : l'interprétation équivoque de l'effet IAT

À un niveau strictement opératoire, l'effet IAT n'est qu'une différence de temps de réponse entre les deux blocs-tests de la procédure. On a d'abord considéré que cette différence reflétait des associations privilégiées entre concepts et attributs pouvant s'interpréter, selon le cas, comme des attitudes, des stéréotypes ou encore des représentations de soi. Ces interprétations – que l'on désignera par l'expression « interprétation naïve » – sont loin d'être aussi univoques qu'elles ne le paraissaient, et plusieurs éléments théoriques et/ou empiriques incitent aujourd'hui à davantage de prudence. Dans les points qui vont suivre, nous présentons les questionnements majeurs quant à l'interprétation d'un score IAT, ainsi que les garde-fous et/ou les raffinements méthodologiques permettant d'y répondre.

Premièrement, la mesure IAT est une mesure d'associations relatives (De Houwer, 2002). Prenons pour illustrer un IAT censé appréhender les attitudes raciales (utilisant les cibles « Blancs » *vs.* « Noirs » et les attributs « Agréable » *vs.* « Désagréable »). Supposons qu'un sujet soit plus rapide dans le bloc où les catégories « Blancs » et « Agréable » (respectivement « Noirs » et « Désagréable ») partagent la même touche de réponse que dans le bloc où les catégories « Noirs » et « Agréable » (respectivement « Blancs » et « Désagréable ») sont associées. Tout ce que l'on peut dire, c'est que l'association « Blancs – Agréable » ou l'association « Noirs – Désagréable » est plus forte que l'association « Blancs – Désagréable » ou plus forte que l'association « Noirs – Agréable ». Et, pour inférer de ce résultat que le sujet a une attitude positive envers les Blancs et négative envers les Noirs, il faut supposer que les deux premières associations sont fortement établies et que les secondes sont faibles ou inexistantes. On observerait toutefois les mêmes résultats : a) si le sujet avait une attitude positive envers les Blancs et neutre vis-à-vis des Noirs (*i.e.*, l'association

« Noirs – Agréable » est du même ordre de grandeur que l'association « Noirs – Désagréable ») ; b) ou s'il avait une attitude neutre envers les Blancs (*i.e.*, même intensité des associations « Blancs – Agréable » et « Blancs – Désagréable ») et une attitude négative vis-à-vis des Noirs (*i.e.*, association « Noirs – Désagréable » plus intense que l'association « Noirs – Agréable »). Pour trancher entre ces différentes interprétations, il faudrait connaître le niveau absolu de l'une des associations, ce qui n'est pas possible à l'intérieur du paradigme IAT, même si l'on essaie de contourner la difficulté en intégrant une catégorie supposée neutre. Les résultats obtenus dans cette perspective par Brendl, Markman et Messner (2001) montrent toute l'importance de cette relativité. Dans un premier temps, ces auteurs répliquent le résultat princeps de Greenwald et al. (1998) : les associations « Fleurs – Agréable » et « Insectes – Désagréable » donnent lieu en moyenne à des réponses plus rapides que les associations « Fleurs – Désagréable » et « Insectes – Agréable », ce qui est d'habitude interprété comme indiquant une attitude favorable vis-à-vis des fleurs et négative envers les insectes. Dans un second temps, les noms de fleurs ont été remplacés par des mots sans signification (catégorie « Pseudo-mots »), supposés affectivement neutres, le reste du matériel étant identique à celui de la première expérience. On observe alors que les associations « Insectes – Agréable » et « Pseudo-mots – Désagréable » donnent lieu à des réponses plus rapides que les associations « Insectes – Désagréable » et « Pseudo-mots – Agréable », ce qui, si l'on s'en tient à l'interprétation naïve, signifierait cette fois, qu'en général, les individus ont une attitude positive vis-à-vis des insectes et négative envers les pseudo-mots ! Cette étude illustre ainsi la relativité des mesures d'association révélées par l'IAT, et insiste donc sur le fait que l'appréciation d'un concept-cible est fortement dépendante de l'autre concept-cible utilisé. Ainsi, même lorsque l'application d'une méthode analytique afin d'évaluer une association absolue entre une seule cible et ses attributs semble séduisante, il est fortement déconseillé de calculer un effet IAT sur la base des temps de réponse aux exemplaires d'un seul concept-cible et de ses attributs (*i.e.*, soustraction des moyennes des temps de réponse à « Fleurs » et « Agréable » dans le premier bloc, de la moyenne des temps de réponse à « Fleurs » et « Désagréable » dans le deuxième bloc) (Nosek, Greenwald et Banaji, étude 1, 2005a).

Néanmoins, le fait que l'IAT ne permettrait que la mesure d'associations relatives n'est pas forcément problématique. En fait, que cela pose problème ou non dépend avant tout de ce que l'on cherche à appréhender. Par exemple, si l'on recourt à un IAT « Fleurs/Insectes » dans le but de situer des sujets les uns par rapport aux autres selon leur attitude envers

les fleurs, le problème de la mesure d'associations relatives se pose alors dans le sens où les différences de scores entre les sujets ne reflètent pas forcément leur différence d'attitude envers les fleurs (*e.g.* des sujets ayant la même attitude envers les fleurs peuvent obtenir des scores IAT « Fleurs/Insectes » différents et des sujets obtenant le même score IAT peuvent avoir des attitudes différentes envers les fleurs). Par contre, si l'on cherche à situer des sujets les uns par rapport aux autres selon leur préférence pour les fleurs ou les insectes, alors le problème de la mesure relative ne se pose plus, en ce sens que les différences de scores entre les sujets devraient refléter leur niveau de préférence pour les fleurs ou les insectes (*e.g.* un effet IAT négatif indique une préférence pour les insectes par rapport aux fleurs, un effet IAT nul indique une absence de préférence, un effet IAT positif indique une préférence pour les fleurs par rapport aux insectes ; et plus l'effet est élevé, en valeur absolue, plus la préférence est marquée).

Deuxièmement, il n'est pas certain que ce soient toujours les associations entre concepts-cibles et concepts-attributs qui soient pertinentes dans l'effet IAT, alors que c'est sur celles-ci que repose l'interprétation naïve. Dans la plupart des applications habituelles, la valence des concepts-cibles est confondue avec la valence des exemplaires qui les représentent (dans un IAT « Fleurs *vs.* Insectes », tous les noms de fleurs utilisés évoquent des représentations positives, tous les noms d'insectes utilisés évoquent des représentations négatives). De Houwer (2001) a essayé de rendre indépendantes la valence des cibles et celle de leurs exemplaires. Ainsi, dans un IAT « Anglais *vs.* Étrangers » réalisé en Grande-Bretagne, les exemplaires de chaque concept-cible sont pour moitié des personnes appréciées et, pour moitié, des personnes supposées détestées (*e.g.*, respectivement « Ghandi » et « Hitler » pour la cible « Étrangers »). Les résultats obtenus semblent indiquer que les concepts-cibles priment sur les exemplaires, c'est-à-dire que l'on observe bien un effet IAT interprété comme une préférence envers l'endogroupe. Mais ce n'est pas toujours le cas : dans certaines situations, la valence des exemplaires peut contribuer à moduler l'effet IAT (Mitchell, Nosek et Banaji, 2003b), voire à l'expliquer totalement lorsque les concepts-cibles n'ont pas de valence marquée (De Houwer, 2001). L'effet IAT serait donc non seulement relatif aux concepts-cibles utilisés, mais aussi aux exemplaires choisis pour les représenter. Cette question étant néanmoins bien traitée dans la littérature, il est possible d'éviter les écueils d'interprétation qui lui sont liés pour peu que l'on respecte certaines règles établies de choix d'exemplaires. On peut trouver une belle synthèse méthodologique concernant le rationnel de ce choix dans Nosek, Greenwald et Banaji (2005b).

Troisièmement, les différences de temps de réponse mesurés aux deux blocs-tests de l'IAT pourraient refléter davantage la fréquence avec laquelle les sujets sont confrontés aux associations rencontrées dans leur environnement plutôt que l'attitude individuelle que l'on infère souvent. Si l'on suit, par exemple, le raisonnement de Karpinski et Hilton (2001), on peut considérer en effet que l'on a davantage d'occasions d'établir des associations stéréotypées « Blancs – Agréable » et « Noirs – Désagréable » que l'inverse (dans la culture américaine tout du moins). La familiarisation avec ces associations permettrait de traiter plus rapidement le bloc dans lequel elles sont présentées et l'on n'aurait alors nul besoin d'invoquer une quelconque attitude pour rendre compte des différences constatées. Certains scores d'attitude IAT sont compatibles avec cette hypothèse, notamment ceux qui montrent que les fumeurs ont une attitude aussi négative envers le tabac que les non-fumeurs (Swanson *et al.*, 2001), ou bien qu'approximativement la moitié des Noirs américains ont une attitude plus positive vis-à-vis des Blancs que vis-à-vis des Noirs (Nosek *et al.,* 2002b ; Banaji, 2001). En substance, l'idée qu'avancent Karpinski et Hilton (2001) ainsi que d'autres comme Olson et Fazio (2004), est que les associations « culturelles » ne nous appartiennent pas, et qu'elles parasitent de ce fait l'évaluation IAT des attitudes personnelles. Mais, avec Nosek et Hansen (2004), on pourrait réviser nos conceptions habituelles en considérant les attitudes automatiques (*i.e.*, les attitudes telles qu'évaluées par l'IAT) comme des construits multidimensionnels dont l'une des nombreuses facettes serait forgée par un conditionnement culturel. Ce qui compterait alors, ce ne serait pas tant l'origine de l'attitude considérée, mais sa disponibilité, son accessibilité et son applicabilité dans un certain contexte (Nosek et Hansen, 2004).

Quatrièmement, une autre explication de l'effet IAT met l'accent sur l'intervention de mécanismes de contrôle exécutif dont certains effets sur les temps de réponse seraient indépendants de la force d'association entre concepts et fausseraient par là l'interprétation des scores IAT (Mierke et Klauer, 2001, 2003 ; Klauer et Mierke, 2005). Comme le rappellent McFarland et Crouch (2002), on peut s'étonner que dès la première présentation de l'IAT (Greenwald *et al.*, 1998), les auteurs trouvent une corrélation anormalement élevée (r = .58) entre les scores d'un IAT portant sur les concepts-cibles « Fleurs *vs.* Insectes » et ceux d'un IAT portant sur les concepts-cibles « Instruments de musique *vs.* Armes » (les attributs étant « Agréable » et « Désagréable » dans les deux cas) : l'attitude envers les fleurs serait ainsi assez fortement liée à l'attitude envers les instruments de musique. Obtenant des résultats similaires avec d'autres cibles dont on s'attend à ce qu'elles génèrent des attitudes indépendantes

les unes des autres, McFarland et Crouch (2002) concluent alors que ces corrélations reflètent un effet de méthode qu'ils attribuent à la plus ou moins grande facilité avec laquelle les sujets sont capables de traiter le bloc-test dit incompatible, c'est-à-dire celui qui repose sur les associations les moins évidentes pour les sujets (et qui entraîne donc des temps de réponse en général plus longs). Mierke et Klauer (2001, 2003) proposent une explication cognitive élégante de cet effet faisant intervenir le coût de l'alternance entre deux tâches ("task switch cost"). Selon ces auteurs, confrontés au bloc compatible de l'IAT (celui dans lequel les associations sont les plus évidentes pour les sujets), les sujets pourraient simplifier la consigne qui leur demande de classer des stimuli soit en fonction des concepts-cibles (*e.g.*, « Fleurs » *vs.* « Insectes »), soit en fonction des concepts-attributs (*e.g.*, « Agréable » *vs.* « Désagréable »). Ainsi suffirait-il, dans le bloc où sont associés « Fleurs » et « Agréable » d'une part, et « Insectes » et « Désagréable » d'autre part, de ne classer qu'en fonction d'une seule dichotomie : « Agréable » *vs.* « Désagréable ». Cette simplification de deux tâches en une n'est plus possible dans le bloc incompatible (associations « Fleurs – Désagréable » et « Insectes – Agréable ») où la sélection de la réponse appropriée ne peut se faire qu'en fonction des deux dichotomies indiquées dans la consigne (*i.e.*, « Fleurs » *vs.* « Insectes » et « Agréable » *vs.* « Désagréable »). Par exemple, lors du passage d'un exemplaire attribut à un exemplaire cible, le sujet doit inhiber la tendance de réponse maintenant non pertinente (classer selon la dichotomie « Agréable » *vs.* « Désagréable ») pour activer le schéma de réponse maintenant pertinent (classer selon la catégorie sémantique « Fleur » *vs.* « Insecte »). Cette alternance entre deux tâches entraînerait un coût cognitif supplémentaire dont l'importance serait propre à chaque individu et ce, quels que soient les concepts-cibles ou attributs à classer. C'est ici que s'infiltrerait une part de variabilité interindividuelle non désirée car indépendante de la force d'association entre concepts. Afin d'appuyer leur théorie, Mierke et Klauer (2001) rapportent les résultats d'études expérimentales montrant notamment que l'effet IAT diminue lorsqu'on fournit aux sujets, avant chaque essai, des indices permettant de diminuer le coût de l'alternance entre les tâches (*cf.* aussi l'étude de Dasgupta, McGhee, Greenwald et Banaji (2000) établissant que l'effet IAT est plus faible lorsqu'il s'agit de classer des photos que lorsqu'il s'agit de classer des noms, le premier classement étant jugé cognitivement moins complexe que le second). D'autres résultats semblent, eux aussi, apporter quelque crédit à cette hypothèse. Ainsi, Chee, Sriram, Soon et Lee (2000) observent, par IRMf, une activation des centres neuronaux impliqués dans l'inhibition lorsque des sujets réalisent un IAT ; Hummert *et al.*

(2002) parviennent à faire disparaître ou à atténuer certaines différences entre groupes d'âge lorsque la vitesse globale de traitement de l'information des sujets est contrôlée. Dans une perspective quelque peu différente, les observations mettant en évidence le rôle de l'apprentissage dans l'effet IAT peuvent aussi être lues à l'aune de l'hypothèse de Mierke et Klauer (2001). Ainsi, des études montrent que l'effet IAT est plus faible : lors d'une seconde passation comparativement à la première (Steffens et Buchner, 2003) ; quand on le calcule sur la seconde partie des items des blocs-tests par rapport au même calcul sur la première moitié des items (Marsh, Johnson et Scott-Sheldon, 2001) ; lorsque le bloc incompatible précède le bloc compatible (Greenwald et Nosek, 2001)(4). La présence de variabilité systématique de méthode dans les scores produits par l'IAT est donc bien établie. Ce pourrait être un coup sérieux porté à la validité des scores IAT si l'on n'avait pas récemment découvert une parade à ce problème. Signalons d'abord qu'il existe deux manières « canoniques » de calculer l'effet IAT : l'algorithme dit « amélioré » apparu en 2003 (Greenwald *et al.*, 2003) ; l'algorithme dit « conventionnel » (Greenwald *et al.*, 1998), considéré comme dépassé. Une des particularités de l'algorithme amélioré est de calculer l'effet IAT en unité d'écart-type des temps de réponse sur l'ensemble des deux blocs qui composent l'IAT. Mierke et Klauer (2003) ont montré que parce qu'il prenait en compte la variabilité intraindividuelle des temps de réponse, le nouvel algorithme contrôle en fait la variabilité interindividuelle de flexibilité cognitive dont provient la part de variance systématique de méthode des scores IAT. Pour cette raison, et parmi d'autres (*cf.* Greenwald *et al.*, 2003), il est fortement conseillé d'utiliser maintenant l'algorithme amélioré pour tout calcul d'effet IAT.

Cinquièmement, en raison de la relation asymétrique entre une association et une attitude, il n'est pas toujours certain que les associations observées dans les IAT d'attitudes renvoient aux attitudes correspondantes. Pour Fiedler, Messner et Bluemke (2005)(5), le fait de définir une attitude comme une association entre un objet donné et une valence (positive ou négative) (Fazio, 1986 ; Greenwald *et al.*, 2002) ne doit pas faire oublier qu'il n'y a aucune raison de postuler, comme on le fait habituellement, qu'il existe une relation symétrique entre ces deux concepts : si l'on peut effectivement modéliser une attitude par l'association entre un objet donné et une valence, *toute* association entre un objet donné et une valence observée dans un IAT n'est pas forcément le reflet d'une attitude. Il suffit pour cela que les sujets utilisent d'autres critères de classification que ceux indiqués par l'expérimentateur. Comme illustration un peu extrême, prenons l'exemple de De Houwer, Geldof et De

Bruycker (2005) qui montre que les associations « Pizzas – Pièces de monnaie » et « Serpents – Rivières » donnent lieu à des réponses plus rapides que les associations « Pizzas – Rivières » et « Serpents – Pièces de monnaie ». Dans la première situation et afin de simplifier la tâche, les sujets utilisent probablement la similarité de forme entre les concepts, même si cette caractéristique n'a jamais été évoquée dans les consignes. D'autres caractéristiques non pertinentes du point de vue du chercheur peuvent être utilisées par les sujets. Par exemple, Rothermund et Wentura (2004) expliquent la relative facilité des associations « Insectes – Agréable » et « Pseudo-mots – Désagréable » décrite *supra* par un modèle de saillance figure-fond. Selon ce dernier, les pseudo-mots (du fait de leur caractère étrange) et les attributs négatifs (*e.g.*, « Désagréable ») (parce que les stimuli négatifs attirent davantage l'attention que les stimuli positifs, *cf.* Peeters, 1992) constituent des éléments saillants par rapport au fond constitué des attributs positifs et des exemplaires renvoyant à « Insectes ». Les associations peuvent se former ici suivant le degré de saillance du matériel à classer. D'où le fait que l'on observe une « attitude » favorable envers les insectes. En revanche, dans un IAT plus traditionnel (« Fleurs » *vs.* « Insectes » et « Agréable » *vs.* « Désagréable »), la catégorie « Désagréable » reste saillante par rapport à la catégorie « Agréable », mais cette fois-ci, la catégorie « Insectes » est plus saillante que la catégorie « Fleurs » (en raison de l'asymétrie entre concepts négatifs et positifs). L'association « Insectes – Désagréable » est dès lors facilitée, et on observe une « attitude » défavorable vis-à-vis des insectes. La « référence à soi » constituera un dernier exemple de critère de classification non pertinent du point de vue de l'expérimentateur. Dans un IAT « Blancs » *vs.* « Noirs » et « Agréable » *vs.* « Désagréable », un Blanc ayant une bonne estime de lui-même peut réduire la difficulté de la tâche en utilisant le fait que les stimuli de personnes blanches font référence à soi (car c'est une de ses caractéristiques propres) tout comme les stimuli de la caractéristique « Agréable » si elle possède une estime de soi élevée. Pour cette personne, l'association « Blancs – Agréable » va donner lieu à des réponses plus rapides, qui risquent d'être interprétées, à tort, comme un indicateur d'attitude favorable envers les Blancs et défavorable envers les Noirs.

En fait, n'importe quelle ressemblance suscitée par des caractéristiques communes non pertinentes du point de vue du chercheur mais néanmoins utilisée par un sujet donné pour se faciliter la tâche de classification, peut provoquer des effets IAT. On peut alors réinterpréter le paradigme comme une mesure générale de similarité entre catégories (De Houwer *et al.*, 2005). Le modèle de la redondance de Fiedler *et al.*

(2003) tente d'en formuler les principes. Dans le bloc compatible, les catégories associées à la même touche de réponse partagent un maximum de caractéristiques, les catégories associées à des touches différentes n'ont que peu ou pas de caractéristiques communes : dans cette situation, il n'y a donc guère de raison de confondre les touches de réponse. Dans le bloc incompatible en revanche, les catégories associées à des touches différentes ont en commun une ou plusieurs caractéristiques, les deux touches de réponse ont par conséquent tendance à être psychologiquement confondues. Dans ce modèle, l'interprétation traditionnelle de l'effet IAT est appropriée quand, du point de vue du chercheur, la distance psychologique entre les touches de réponse est uniquement fonction des associations entre les caractéristiques pertinentes des attributs et des cibles ; elle devient inappropriée quand la distance psychologique est due à des redondances entre caractéristiques non pertinentes des attributs et des cibles.

3.5. Des alternatives

Afin de remédier à certaines des limitations venant d'être évoquées, des instruments concurrents ou plutôt complémentaires à l'IAT sont apparus. Dans ce qui suit, nous mentionnerons uniquement le GNAT (*Go/No-go Association Task* ; Nosek et Banaji, 2001), l'EAST (*Extrinsic Affective Simon Task* ; De Houwer, 2003) et le SCAT (*Single Category Association Test* ; Karpinski & Steinman, 2006), car ils nous semblent les plus prometteurs.

Le GNAT résout le problème de la mesure relative car il permet de mesurer la force d'association entre un seul concept-cible et son attribut. La tâche du sujet consiste à discriminer entre les exemplaires d'un concept-cible (*e.g.*, « Fruit ») ou d'un attribut (*e.g.*, « Agréable ») et des distracteurs. Dans le premier cas (exemplaire-cible ou exemplaire-attribut), le sujet appuie sur une touche prédéterminée (« *Go* ») alors que dans le second (distracteurs), aucune action n'est requise (« *No Go* »). Tout comme dans l'IAT, la procédure de mesure s'effectue en deux phases se distinguant par le changement d'attribut : par exemple, pour mesurer une attitude, on utilise un attribut positif dans une phase (*e.g.*, « Agréable ») et un attribut négatif dans la seconde phase (*e.g.*, « Désagréable »). Le concept-cible est donc associé soit à un attribut positif, soit à un attribut négatif. De la différence de performance entre ces deux phases, on en infère une attitude plus ou moins positive vis-à-vis du concept-cible. Bien que les qualités psychométriques du GNAT soient comparables à celles de l'IAT, les deux mesures corrèlent très faiblement

ensemble (Nosek et Banaji, 2001) ; et les seules corrélations avec une mesure explicite rapportées, à notre connaissance, dans la littérature sont très faibles à inexistantes (Nosek et Banaji, 2001). Mais il est vrai que très peu d'études ont utilisé ce paradigme jusqu'à présent (pour de rares exceptions, *cf.* Blair, Ma et Lenton, 2001 ; Mitchell, *et al.*, 2003).

L'EAST (De Houwer, 2003) élimine lui aussi le problème de la mesure relative ; il évite, en plus, la comparaison entre deux blocs reposant sur des tâches différentes. Tout comme dans l'IAT, il s'agit de classer des stimuli, mais le critère de classification est ici soit la signification sémantique (pour les concepts-attributs), soit la couleur (pour le concept-cible). Concrètement, des exemplaires-attributs sont présentés dans une couleur blanche, le sujet devant indiquer s'ils relèvent de la catégorie « Agréable » ou de la catégorie « Désagréable » en appuyant sur la touche correspondante. Le concept-cible est, quant à lui, présenté soit en bleu, soit en vert, le sujet devant discriminer la couleur en utilisant les mêmes touches de réponse que celles utilisées pour discriminer les exemplaires-attributs. Même si la signification des concepts-cibles peut être ignorée pour réussir la tâche, on suppose que le sujet sera plus rapide lorsque la valence – automatiquement activée – de la cible et sa couleur correspondent à la même touche de réponse ; et qu'il sera plus lent lorsque la valence de la cible et sa couleur relèvent de deux touches différentes. Ainsi, si les catégories « Agréable » et « Couleur bleue » sont associées à la même touche de réponse alors que les catégories « Désagréable » et « Couleur verte » sont associées à l'autre touche de réponse, un sujet qui a une attitude favorable envers les fruits devrait répondre plus rapidement lorsque ceux-ci sont présentés en bleu que lorsqu'ils sont présentés en vert, ce qui est vérifié empiriquement (De Houwer, 2003). L'EAST permet donc d'estimer des associations absolues (*i.e.*, non relatives) en une seule phase ; il présente aussi l'avantage de permettre la mesure simultanée de plusieurs attitudes puisqu'il est possible de présenter des exemplaires de plusieurs concepts-cibles, ceux-ci ne devant être discriminés que par rapport à leur couleur. Malheureusement, malgré le caractère ingénieux de la procédure, les qualités psychométriques du paradigme sont décevantes (De Houwer, 2003) et bien inférieures à celles de l'IAT, ce qui s'explique sans doute en partie par des écarts trop faibles de performances entre les deux conditions expérimentales (exemplaires-cibles d'une couleur *vs.* de l'autre couleur), et donc à des tailles d'effet insuffisantes pour assurer des différences interindividuelles stables. L'interprétation des effets mesurés dans ce paradigme étant toutefois moins équivoque que celle des effets IAT, on peut espérer que des aménagements internes puissent remédier à ces difficultés psychométriques.

Le SCAT (Karpinski & Steinman, 2006) est le dernier né de ces trois instruments alternatifs. Il permet lui aussi de mesurer la force d'association absolue entre un concept-cible et son attribut. Il ressemble beaucoup à l'IAT dans sa mise en œuvre : grâce à deux touches, le sujet doit classer à droite ou à gauche des stimuli appartenant à des catégories différentes. Mais au lieu d'être quatre, ces catégories ne sont que trois : deux catégories d'attributs opposés (*e.g.*, « Agréable » *vs.* « Désagréable ») et une seule catégorie cible (*e.g.*, « Fruit »). Ainsi, dans le premier bloc-test les sujets doivent classer par exemple les stimuli relevant des catégories « Agréable » et « Fruit » à gauche et ceux de la catégorie « Désagréable » à droite ; dans le deuxième bloc-test, ils doivent cette fois classer les exemplaires de la catégorie « Agréable » à gauche et ceux des catégories « Désagréable » et « Fruit » à droite. Dans les trois applications rapportées par Karpinski (2004), les consistances internes sont satisfaisantes et du même ordre de grandeur que celles que l'on observe habituellement avec l'IAT ; les corrélations avec des mesures directes sont sensiblement supérieures à celles impliquant l'IAT, résultat que l'auteur attribue au fait que son instrument mesure des associations absolues alors que l'IAT ne mesure que des associations relatives. Cette caractéristique fait du SCAT un paradigme prometteur pour peu que l'on puisse expliquer certains résultats étonnants, comme l'absence totale de corrélation entre un SCAT d'estime de soi et un IAT d'estime de soi (Karpinski & Steinman, 2006).

4. LES RELATIONS ENTRE MESURES DIRECTES ET MESURES INDIRECTES

L'étude de validité des mesures dérivées des différents paradigmes de mesures indirectes en général et de l'IAT en particulier a largement fait référence à la mise en relation entre mesures directes et indirectes du même construit (Nosek, 2004 ; Hofman *et al.*, 2005). Mais pour que les résultats de ces études puissent servir au processus de validation, encore faut-il que l'on ait des hypothèses précises quant à la nature des relations entre ces mesures. Or, s'il existe un relatif consensus sur le fait que les relations observées sont généralement faibles mais positives, l'interprétation de cette dissociation reste l'objet de débats : s'explique-t-elle simplement par les problèmes théoriques et méthodologiques posés par l'IAT ou bien par les biais inhérents aux mesures directes ? Renvoie-t-elle à l'existence de construits différents ? L'objectif de cette dernière partie est

de détailler les arguments fournis en faveur des différentes explications proposées.

Premièrement, la dissociation observée entre mesures directes et IAT peut s'expliquer par les problèmes méthodologiques et théoriques posés par l'IAT. Nous avons déjà fait référence au fait qu'il ne produit que des mesures d'associations relatives. Par exemple, un IAT d'estime de soi mesurerait en fait la force d'association entre les concepts « Moi – Agréable » comparativement à la force d'association entre les concepts « Pas moi – Agréable » et non la force d'association « Moi – Agréable » en elle-même (Karpinski & Steinman, 2006). Or, même si certains items des mesures directes d'estime de soi font référence à autrui, elles ne sont pas entièrement construites sur ce modèle, ce qui pourrait participer à la faible relation observée entre mesures directe et IAT. De plus, le fait que la tâche IAT se décompose en deux blocs indépendants introduit un certain nombre de biais de mesure, qui, alliés à ceux cités plus haut, contribuent sans doute aussi au manque de convergence entre mesures directes et IAT. Ces explications ne sont toutefois pas entièrement convaincantes puisque avec les nouveaux paradigmes censés corriger certains des défauts de l'IAT (GNAT, EAST, SCAT), les corrélations entre mesures directes et indirectes restent très inférieures à ce que l'on pourrait attendre. Il faut donc chercher des explications complémentaires.

Deuxièmement, la dissociation observée entre mesures directes et IAT pourrait être imputée à certains biais propres aux mesures directes : l'auto-présentation et les limites des capacités introspectives des sujets. Alors que les mesures directes sont connues pour leur sensibilité aux stratégies de présentation adoptées par les sujets (Kop et Chassard, 2005), l'IAT n'en serait pas affecté (Greenwald *et al.*, 2002). Néanmoins, les résultats des études portant sur le rôle modérateur de l'auto-présentation dans la relation entre mesures directe et IAT sont contradictoires. Alors que Nosek (2004) trouve effectivement que plus l'auto-présentation est élevée, plus la relation entre mesures directes et IAT est faible et que Banse (2004) montre que l'auto-présentation modère la relation entre mesures directes et IAT, Nosek et Banaji (2002) et Egloff et Schmukle (2003) ne répliquent pas ces résultats. Il semble en outre peu probable que l'auto-présentation puisse expliquer entièrement les dissociations observées. Si elle devait rendre compte, par exemple, de l'absence de corrélation (Greenwald *et al.*, 1998) entre un IAT et une mesure directe censés appréhender l'attitude envers les fleurs, il faudrait alors faire l'hypothèse que ce type d'attitude est particulièrement sensible à l'auto-présentation ! Enfin, les limites des capacités introspectives des sujets, plus difficiles à appréhender dans des études empiriques, ont fait l'objet

de moins de discussions mais continuent à être invoquées comme une cause possible des dissociations entre mesures directes et indirectes. Toutefois, comme le reconnaissent Greenwald *et al.* (2002), il est souvent quasi impossible de faire la distinction entre ces deux facteurs explicatifs potentiels : les domaines pour lesquels les sujets n'ont que peu de raisons à chercher à dissimuler leur attitude sont aussi généralement ceux qui sont les plus facilement accessibles à l'introspection.

Troisièmement, la dissociation entre mesures directes et IAT peut s'expliquer par des éléments théoriques relevant de la cognition implicite, grâce notamment à certaines réflexions récentes relatives aux attitudes. Deux grandes classes de modèles peuvent être distinguées selon que l'on considère que la mesure directe et la mesure indirecte appréhendent chacune un construit différent ou selon que l'on suppose qu'elles sont deux manières différentes d'appréhender le même construit. Nous commencerons par évoquer l'hypothèse de l'existence de deux construits différents. En postulant l'existence de deux construits, les partisans de cette approche supposent que chacun des construits possède sa propre représentation en mémoire (Greenwald et Banaji, 1995 ; Wilson, Lindsley et Schooler, 2000). Par exemple, le modèle des attitudes duelles (*model of dual attitudes*) développé par Wilson *et al.* (2000), distingue entre le construit d'« attitude explicite » (Ae) et celui d'« attitude implicite » (Ai) et repose sur les cinq hypothèses suivantes : a) Une Ae et une Ai envers un même objet peuvent coexister en mémoire ; b) Lorsque des attitudes duelles coexistent, Ai est activée automatiquement alors que Ae requiert davantage de ressources cognitives et de motivations pour être récupérée. Si les individus sont capables de récupérer Ae, celle-ci prend alors le pas sur Ai, de sorte que les sujets expriment Ae. Au contraire, si les individus n'ont ni les ressources ni la motivation nécessaires pour récupérer Ae, ils expriment Ai ; c) Même lorsque Ae est récupérée, Ai influence les réponses implicites, c'est-à-dire les réponses que les individus ne peuvent contrôler (*e.g.*, comportements non-verbaux) ou qu'ils ne considèrent pas comme l'expression de leur attitude, et donc n'essaient pas de contrôler ; d) Les Ae changent relativement facilement, alors que les Ai, comme de vieilles habitudes, changent plus lentement ; e) Les attitudes duelles sont distinctes de l'ambivalence ainsi que des attitudes à composantes affective et cognitive contradictoires. En conséquence, plutôt que de faire l'expérience d'un état conflictuel subjectif, les individus expriment l'attitude la plus accessible (Wilson *et al.*, 2000, p. 104).

Wilson *et al.* (2000) distinguent quatre types d'attitudes duelles selon que l'on ait ou non conscience de posséder une attitude implicite Ai et selon que l'on ait besoin de capacités cognitives et de motivations particulières

pour la remplacer par une attitude explicite Ae (cf. tableau II). Si l'on s'en tient à ce modèle, l'IAT permettrait plus particulièrement de mesurer des attitudes implicites que les individus seraient motivés à remplacer par leurs attitudes explicites, et ce qu'ils en aient conscience ou non (*cf.* cas de refoulement et de remplacement motivé). Mais que les individus aient conscience ou non de leurs attitudes implicites, qu'ils aient besoin ou non de ressources cognitives et de motivations particulières pour les remplacer, le modèle de Wilson *et al.* (2000) permet d'expliquer la dissociation entre mesures directe et indirecte par l'existence de deux construits de nature différente : un construit implicite qui se manifesterait systématiquement dans les mesures de type indirect, et un construit explicite qui s'exprimerait préférentiellement dans les mesures de type direct. Dans cette perspective, il est essentiel de mieux connaître les facteurs pouvant modérer la relation entre attitudes explicites et attitudes implicites. Nosek (2004) s'est ainsi lancé dans une entreprise de grande ampleur visant à repérer les modérateurs pertinents pour 57 objets d'attitude différents. Les corrélations entre mesures directes et indirectes (IAT) varient entre -0.05 et 0.70 selon l'objet d'attitude et une part non négligeable de la variabilité de ces corrélations (39 %) peut effectivement être expliquée par les quatre modéra-

Tableau II. Quatre types d'attitudes duelles (d'après Wilson *et al*., 2000, p. 105)

Table II. Four types of dual attitudes (cf. Wilson *et al* ., 2000, p. 105)

	Refoulement	Systèmes indépendants	Remplacement motivé	Remplacement automatique
Ai est-elle accessible à la conscience ?	NON	NON	OUI	Sous certaines conditions
A-t-on besoin de ressources cognitives et de motivations particulières pour remplacerAi par Ae ?	OUI	NON	OUI	NON

Note : Ai = Attitude implicite ; Ae = Attitude explicite

teurs retenus dans l'étude : auto-présentation (*i.e.*, motivation à altérer ses réponses à des fins personnelles ou sociales) ; intensité de l'attitude ; dimensionnalité de l'attitude (*i.e.*, degré de bipolarité) ; différence perçue entre son attitude personnelle et l'attitude moyenne. Ces résultats, prometteurs, méritent néanmoins d'être affinés et répliqués.

Une seconde classe de modèles théoriques considère au contraire qu'il n'existe qu'un seul construit, supposé se différencier par le type de processus de traitement de l'information en jeu lors de sa mesure (Greenwald et Banaji, 1995 ; Dambrun et Guimond, 2003 ; Fazio et Olson, 2003). Ainsi, une dissociation entre mesures directes et IAT peut s'expliquer par le fait qu'une mesure indirecte appréhenderait un construit activé automatiquement sous l'effet de processus spontanés de traitement de l'information alors qu'une mesure directe appréhenderait ce même construit après intervention de processus délibérés (*i.e.*, contrôlés). Le modèle de Fazio et Towles-Schwen (1999, « MODE model of attitude-behavior processes ») censé rendre compte des relations entre attitude et comportement est, dans ce contexte, particulièrement pertinent. En effet, ces auteurs distinguent entre des processus de type spontané qui ne nécessiteraient ni effort conscient, ni intention, ni contrôle de la part des individus et des processus délibérés qui demanderaient un travail cognitif important (inspection de l'information disponible, analyse des caractéristiques positives et négatives, des coûts et des bénéfices…), ce qui suppose de pouvoir disposer à la fois de la motivation et de l'opportunité (*i.e.*, ressources cognitives et temps) nécessaires pour que ce type de processus puisse se mettre en place (Fazio et Towles-Schwen, 1999 ; Koole, Dijksterhuis et Van Knippenberg, 2001). Les processus spontanés et délibérés seraient relativement indépendants les uns des autres, mais ils interagiraient afin de déterminer conjointement le comportement, la prédominance d'un type de processus sur l'autre serait alors fonction de la motivation et de l'opportunité. Ainsi, lorsque la motivation et l'opportunité sont faibles, ce sont les processus spontanés qui détermineraient en grande partie le comportement. Au contraire, si la motivation et l'opportunité sont élevées, les processus de type délibéré prendraient le pas sur les processus spontanés (Fazio et Towles-Schwen, 1999). Banse et al. (2001) ont obtenu des résultats compatibles avec ce modèle théorique dans une étude d'attitude envers l'homosexualité. En particulier, les sujets ayant des attitudes négatives spontanées (mesurées à l'aide d'un IAT) envers l'homosexualité et ayant peu de motivation à contrôler leurs réactions se caractérisent par une attitude homophobe prononcée à une mesure directe. En revanche, chez des sujets ayant un même niveau d'attitudes négatives spontanées, mais chez lesquels la motivation à contrôler leurs réactions est accentuée, cette homophobie dans les mesures directes est fortement réduite.

5. CONCLUSION

L'apparition de l'IAT en 1998 a suscité un engouement dépassant largement le cercle restreint de la psychologie scientifique : articles de presse, émissions télévisées, sites Internet (https://implicit.harvard.edu/implicit/france/) ont donné à cet outil une notoriété inattendue. Il faut sans doute y voir une certaine fascination de la psychologie populaire pour la possibilité d'accéder au plus profond de l'âme humaine. Se présentant comme un instrument permettant de révéler des sentiments que l'on tente habituellement de dissimuler, voire des sentiments auxquels on n'a pas accès, l'IAT a tout naturellement fait écho à ce penchant. La communauté scientifique, elle non plus, n'a pas échappé à cet effet de mode, si l'on en croit le nombre considérable de publications consacrées à ce paradigme depuis cette date. Toutefois, au fur et à mesure des années, la fascination initiale a peu à peu cédé la place à des approches plus critiques, davantage conformes à l'esprit scientifique. Et l'on s'aperçoit, aujourd'hui, que le monde de l'IAT est un monde de paradoxes.

Paradoxe de sa construction, tout d'abord. L'étude de la cognition sociale implicite nécessitait en effet de nouveaux outils et c'est bien dans ce cadre de référence théorique que s'inscrit la naissance de l'IAT. Mais au lieu de construire un instrument s'inspirant des théories de ce courant, les concepteurs de l'IAT ont privilégié une démarche empirique s'appuyant essentiellement sur un critère de maximisation de la consistance interne qui n'est qu'une condition nécessaire pour produire des mesures offrant des scores fidèles et valides et en aucun cas une condition suffisante.

Paradoxe de son fonctionnement ensuite. L'analyse des processus en œuvre dans l'IAT conduit à différentes interprétations de la mesure obtenue, interprétations parfois très éloignées de celle que souhaiteraient ses concepteurs (une force d'association entre concepts reflétant, selon la nature des concepts pris en compte, une attitude, un stéréotype, une représentation de soi…). Comment toutefois expliquer que les mesures IAT offrent par ailleurs des gages de validité compatibles avec l'interprétation naïve, comme par exemple, des fidélités comparables à celles des scores des mesures directes et des corrélations avec des indicateurs comportementaux spontanés ou des indicateurs physiologiques ? Qui plus est, pourquoi des outils considérés comme étant « processuellement » plus purs que l'IAT (et donc moins équivoques quant à l'interprétation de la mesure résultante) donnent des résultats si décevants, là même où l'IAT affiche ses résultats les plus convaincants ?

Paradoxe des études de validation des mesures IAT enfin, qui, à quelques exceptions près, s'appuient sur la mise en relation de mesures directes

avec des mesures indirectes, alors que les premières sont accusées de maux que les secondes sont justement censées corriger. La signification donnée aux résultats obtenus dans cette perspective ne peut donner lieu qu'à des controverses difficiles à trancher, sauf à considérer comme Nosek et Smyth (2004), que les construits tels que mesurés par l'IAT sont théoriquement distincts de ceux mesurés par questionnaires. Par rapport aux scores recueillis à l'aide de questionnaires, il deviendrait alors plus intéressant d'explorer, à l'aide de plans de recherche différents, la valeur originale des scores produits par la famille de mesures dont l'IAT fait partie.

Il faut avoir la modestie de reconnaître qu'aujourd'hui, alors que l'on continue d'avancer de manière remarquable (*cf.* Klauer et Mierke, 2005 ; Conrey, Sherman, Gawronski, Hugenberg et Groom, 2005), on ne maîtrise toujours pas complètement ce que mesurent les applications de l'IAT. L'article de Greenwald *et al.* (1998) reste néanmoins un révélateur : il a permis de décomplexer les chercheurs face à la possibilité de mesurer de manière fiable des construits inaccessibles aux mesures directes traditionnelles. Les imperfections de cet instrument de mesure relevées depuis ont eu en outre le mérite de stimuler la créativité des chercheurs dans différentes directions : étude des processus cognitifs en jeu dans des tâches de classification faisant intervenir des associations entre concepts, développement de modèles théoriques intégrant les construits implicites, mise au point de nouveaux paradigmes théoriquement mieux fondés.

NOTES

1. Même si l'on parle souvent de « mesure » ou d'« instrument » à propos de l'IAT, il serait plus correct de faire référence à un paradigme permettant de générer des instruments de mesure. Dans la suite du texte, afin de ne pas trop alourdir l'exposé, nous sacrifierons toutefois à l'usage en utilisant ces expressions.

2. Conventionnellement, d = .20 : taille d'effet faible, d = .50 : taille d'effet moyenne et d = .80 : taille d'effet importante (Cohen, 1988).

3. Il n'existe pas de consensus concernant la façon d'évaluer la consistance interne des scores produits par une adaptation particulière de l'IAT. Cette question est en outre très rarement débattue. Dans tous les articles l'évaluant, il est cependant une constante consistant à la calculer à partir d'un certain nombre de "sous-effets IAT" obtenus à partir d'autant de sous-ensembles de temps de réponse. Tout comme l'effet IAT principal, ils sont considérés comme reflétant la plus ou moins grande tendance à associer tel ou tel concept cible avec tel ou tel concept attribut. La consistance interne

de ces sous-scores représente alors la consistance de cette tendance telle que révélée lors de la tâche IAT. Certains auteurs calculent deux sous-effets IATs et les corrèlent (*i.e.*, méthode split-half, *ex.*, Marsh, Johnson et Scott-Sheldon, 2001 ; Greenwald *et al.*, 2003), d'autres utilisent l'alpha de Cronbach soit à partir de deux sous-effet IAT (*ex.*, Banse *et al.*, 2001), soit à partir de quatre (*ex.*, Asendorpf *et al.*, 2002 ; Gawronski, 2002), soit encore à partir d'autant de sous-effets qu'il y a d'essais (*i.e.*, essai 1 du deuxième bloc moins essai 1 du premier bloc, essai 2 du deuxième bloc moins essai 2 du premier bloc, etc. ; *ex.*, Bosson *et al.*, 2000 ; Egloff et Schmuckle, 2004).

4. Lorsque le bloc compatible précède le bloc incompatible, les associations du premier bloc correspondent à des associations connues nécessitant peu d'apprentissage (*e.g.*, « Fleurs – Agréable »). Dans le bloc incompatible, il faut « désapprendre » ces associations et en apprendre de nouvelles, moins évidentes (*e.g.*, « Fleurs – Désagréable »). La différence de vitesse entre les deux blocs est maximale. Or, lorsqu'on inverse les deux blocs, ce sont cette fois les associations du bloc incompatible qui sont nouvelles et nécessitent un apprentissage (*e.g.*, « Fleurs – Désagréable »). À cela s'ajoute le fait que ces associations devront de surcroît être inhibées lors du second bloc. D'où le fait que dans cet ordre d'apparition des blocs, la différence de vitesse entre les deux blocs est minimale. C'est pourquoi, si l'on s'intéresse à l'effet IAT au niveau du groupe, il est généralement recommandé de contrebalancer les deux ordres de présentation (Greenwald *et al.*, 1998). Par contre, il serait moins pertinent de le faire si l'on s'intéressait aux différences interindividuelles. Dans ce cas le fait de doubler le nombre d'essais d'entraînement entre les deux blocs tests éliminerait ou tout du moins réduirait l'effet de l'ordre de présentation des blocs pour un grand nombre d'IAT d'attitude (Nosek *et al.*, 2005a).

5. Même si Fiedler *et al.* (2005) évoquent plus volontiers le concept d'attitudes, leur critique peut être étendue aux autres types de concepts appréhendés par l'IAT (i.e. stéréotypes, estime de soi, concept de soi).

BIBLIOGRAPHIE

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

Ashburn,-Nardo, L., Voils, C.I., & Monteith, M.J. (2001). Implicit associations as

the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81,* 789-799.

Banaji, M.R. (2001). Implicit attitudes can be measured. In H.L. Roediger, III, J.S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*. Washington: American Psychological Association.

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

Banse, R. (2004, July). *Indirect measures of aggressiveness.* Paper presented at the 12th European Conference on Personality. Groningen: The Netherlands.

Banse, R., Seise, J. & Zerbes, N. (2001). Implicit attitudes towards homosexuality : Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48,* 145-160.

Blair, I.V., Ma, J.E., & Lenton, A.P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81,* 828-841.

Blaison, C. & Gana, K. (2004, July). *Using the Implicit Association Test (IAT) to measure affectivity-trait versus state.* Paper presented at the 12th European Conference on Personality. Groningen: The Netherlands.

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

Brendl, C.M., Markman, A.B & 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,* 760-773.

Chassard, D., & Kop, J.-L. (2003). Des processus automatiques d'évaluation à la mesure des différences individuelles : l'essor des mesures indirectes. In A. Vom Hofe, H. Charvin, J.-L. Bernaud, D. & Guédon (Eds.), *Psychologie différentielle : recherches et réflexions.* Rennes, Presses Universitaires de Rennes.

Chee, M., Sriram, N., Soon, C.H., & Lee, K.M. (2000). Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *Neuroreport: For Rapid Communication of Neuroscience Research, 11,*135-140.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. 2e ed.* Hillsdale (NJ): Erlbaum.

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

Cunningham, W.A., Preacher, K.J. & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 121,* 163-170.

Dambrun, M. & Guimond, S. (2003). Les mesures implicites et explicites des préjugés et leur relation : développements récents et perspectives théoriques. *Les Cahiers Internationaux de Psychologie Sociale, 57,* 52-73.

Dasgupta, N. & Greenwald, A.G. (2001). On the malleability of automatic attitudes: combatting automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 85,* 800-814.

Dasgupta, N., Greenwald, A.G. & Banaji, M.R. (2003). The first ontological challenge to the IAT: attitude or mere familiarity. *Psychological Inquiry, 14,* 238-243.

Dasgupta, N., McGhee, D.E., Greenwald, A.G., & Banaji, M.R. (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. (2002). The Implicit Association Test as a tool for studying dysfunctional associations in psychopathology: strength and limitations. *Journal of Behavior Therapy and Experimental Psychiatry, 33,* 115-133.

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

De Houwer, J., Geldof, T, & De Bruycker, E. (2005) The Implicit Assocation Test as a general measure of similarity. *Canadian Journal of Experimental Psychology*, 59, 228-239.

De Jong, P.J. (2002). Implicit self-esteem and social anxiety: differential self-favouring effects in high and low anxious individuals. *Behaviour Research and Therapy, 40,* 501-508.

De Jong, P.J., Van Den Hout, M.A., Rietbroek, H &, Huijding, J. (2003). Dissociation between implicit and explicit attitudes toward phobic stimuli. *Cognition and Emotion, 17,* 521-545.

Devos, T., & Banaji, M.R. (2003). Implicit self and identity. *Annals of the New York Academy of Sciences, 1001,* 177-211.

Egloff, B. & Schmukle, S.C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology, 83,* 1441-1455.

Egloff, B. & Schmukle, S.C. (2003). Does social desirability moderate the relationship between implicit and explicit anxiety measures? *Personality and Individual Differences, 34,* 1-10.

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

Fazio, R.H, & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54,* 297-327.

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.

Fazio, R.H. & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology.* New York, Guilford Press.

Fazio, R.H., Williams, C.J. & Powell, M.C. (2000). Measuring associative strength: Category-item associations and their activation from memory. *Political Psychology, 21,* 7-25.

Fiedler, K., Messner, C. & Blümke, M. (2003). Unresolved problems with the "I", the "A" and the "T": Logical and psychometric critique of the implicit association test (IAT). Manuscrit non publié.

Fiedler, K., Messner, C. & Blümke, M. (2005). *Some psychometric problems with the "I", the"A" ans the"T" of the implicit association test.* Soumis à publication.

Gawronski, B. (2002). What does the implicit association test measure ? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology, 49,* 171-180.

Gemar, M.C., Segal, Z.V., Sagrati, S. & Kennedy, S.J. (2001). Mood-induced changes on the Implicit Association test in recovered depressed patients. *Journal of Abnormal Psychology, 110,* 282-289.

Greenwald, A.G, & Banaji, M.R. (1995). Implicit social cognition : Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4-27.

Greenwald, A.G., Banaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A. & Mellott, D.S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3-25.

Greenwald, A.G., & Farnham, S.D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Jour-*

nal of Personality and Social Psychology, 79, 1022-1038.

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

Greenwald, A.G. & Nosek, B.A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48,* 85-93.

Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: an improved algorithm. *Journal of Personality and Social Psychology, 85,* 197-216.

Hofman, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between IAT and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369-1385.

Hummert, M.L., Gartska, T.A., O'Brien, L.T., Greenwald, A.G., & Mellott, D.S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognition. *Psychology and Aging, 17,* 482-495.

Jacoby, L.L., & Kelley, C.M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Behavior Bulletin, 13,* 314-336.

Jacoby, L.L., Toth, J.P., & Yonelinas, A.P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122,* 139-154.

Jelenec, P., & Steffens, M.C. (2002). Implicit attitude toward elderly women and men. *Current Research in Social Psychology, 7,* 275-292.

Karpinski, A. & Hilton, J.L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81,* 774-788.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91,* 16-32.

Kihlstrom, J.F., Barnhardt, T.M., & Tataryn, D.J. (1992). The psychological unconscious: Found, lost, and regained. *American Psychologist, 47,* 788-791.

Kihlstrom, J.F., Mulvaney, S., Tobias, B.A, & Tobis, I.P. (2000). The emotional unconscious. In E. Eich, J.F. Kihlstrom, G.H. Bower, J.P. Forgas & P.M. Niedenthal (Eds.), *Cognition and emotion.* Oxford: Oxford University Press.

Klauer, K.C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the IAT effect. *Personality and Social Psychology Bulletin, 31,* 208-217.

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.

Kop, J-L. & Chassard, D. (2005). La falsification des réponses dans l'évaluation de la personnalité : une solution du côté des mesures indirectes ? *Psychologie du Travail et des Organisations, 11,* 15-23.

Marsh, K.L., Johnson, B.T., & Scott-Sheldon, L.A. (2001). Heart versus reason in condom use: implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift für Experimentelle Psychologie, 48,* 161-175.

McFarland, S.G., & Crouch, Z. (2002). A cognitive skill confound on the IAT. *Social Cognition, 20,* 483-510.

Mierke, J., & Klauer, K.C. (2001). Implicit association measurement with IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie, 48,* 107-192.

Mierke, J., & Klauer, K.C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85,* 1180-1192.

Mitchell, C.J., Anderson, N.E., & Lovibond, P.F. (2003). Measuring evaluative

conditionning using the Implicit Association Test. *Learning and Motivation, 34,* 203-217.

Mitchell, J.P., Nosek, B.A. & Banaji, M.R. (2003). Contextual variations in implicit variations. *Journal of Experimental Psychology: General, 132,* 455-469.

Musch, J., & Klauer, K.C. (Eds.) (2003). *The psychology of evaluation: Affective processes in cognition and emotion.* Mahwah (NJ, US): Lawrence Erlbaum Associates.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Nosek, B.A. (2004). Moderators of the relationship between implicit and explicit attitudes. *Dissertation Abstracts International: Section-B: The Sciences and Engineering, 63,* 4965.

Nosek, B.A., & Banaji, M.R. (2001). The go/no go association task. *Social Cognition, 19,* 625-664.

Nosek, B.A, & Banaji, M.R. (2002). [Polish language] (At least) two factors moderate the relationship between implicit and explicit attitudes. In R.K. Ohme & M. Jarymowicz (Eds.), *Natura automatyzmon.* Warszawa: Wip Pan & SWPS.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002a). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83,* 44-59.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002b). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics, 6,* 101-115.

Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2005a). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

Nosek, B.A., Greenwald, A.G., & Banaji, M.R. (2005b). The Implicit Association Test at age 7: A methodological and conceptual review. In J.A. Bargh (Edit.), *Automatic Processes in Social Thinking and Behavior.* Psychology Press.

Nosek, B.A., & Hansen, J.J. (2004). *The associations in our heads belong to us: Measuring the multifaceted attitude construct in implicit social cognition.* Manuscrit non publié.

Nosek, B.A., & Smyth, F.L. (2005). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. Sous presse.

Olson, M.A., & Fazio, R.H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: personalizing the IAT. *Journal of Personality and Social Psychology, 86,* 653-667.

Palfai, T.P., & Ostafin, B.D. (2003). Alcohol related motivational tendencies in hazardous drinkers: assessing implicit response tendencies using the modified-IAT. *Behaviour Research and Therapy, 41,* 1149-1162.

Peeters, G. (1992). Evaluative meanings of adjectives in vitro and in context: Some theoretical implications and practical consequences of positive-negative asymmetry and behavioral-adaptive concepts of evaluation. *Psychologica Belgica, 32,* 211-231.

Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, E.S., Gatenby, J.C., Gore, J.C., & Banaji, M.R. (2000). Performance on indirect measures of race evaluation predicts amygdalia activation. *Journal of Cognitive Neuroscience, 12,* 729-738.

Piontkowski, U., Blanz, M., Rohmann, A., Schmermund, A., & Florack, A. (2001). The impact of multiple fit on implicit associations between categories: the Implicit Association Test as a research instrument in the study of social categorization. In F. Columbus (Ed.), *Advances in psychology research, t. VI.* Hauppauge (NY): Nova Science Publishers.

Poehlman, T.A., Uhlmann, E., Greenwald, A.G, & Banaji, M.R. (2005). *Understanding and Using the Implicit Association Test: III. Meta-analysis of Predictive Validity.* Manuscrit non publié.

Pratto, F., & John, O.P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380-391.

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

Steffens, M.C. (2004). Is the Implicit Association Test Immune to Faking? *Experimental Psychology, 51,* 165-179.

Steffens, M.C., & Buchner, A. (2003). Implicit Association Test: separating transituationally stable and variable components of attitudes toward gay men. *Experimental Psychology, 50,* 33-48.

Swanson, J.E., Rudman, L.A., & Greenwald, A.G. (2001). Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cognition and Emotion, 15,* 207-230.

Teachman, B.A., Gregg, A.P. & Woody, S.R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110,* 226-235.

Wilson, T.D., Lindsley, S. & Schooler, T.Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101-126.