Università degli Studi di Padova

Dipartimento di *Filosofia, Sociologia, Pedagogia e Psicologia Applicata*

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE

CICLO XXVI

TESI DI DOTTORATO

Implicit measurement at the service of mental health: assessment and intervention as the two sides of the same coin

Direttore della Scuola: Ch.ma Prof.ssa Francesca Peressotti

Supervisore: Ch.ma Prof.ssa Stefania Mannarini

Dottorando:

Marilisa Boffo

Abstract

Scientific research in psychology is intrinsically bound to the measurement of variables that are per nature highly complex, changeable, and most often unobservable. The design of measurement methods is mostly focused on the attempt to capture the main features of the psychological attribute of interest. The last fifteen years have seen a massive development and use of a new set of measurement instruments that go under the name of *implicit measures*, which accomplishes the primary goal of indexing psychological attributes interchangeably defined as automatic, uncontrollable, unconscious, impulsive, or implicit. The primary goal of the present work was to explore the *implicitness* feature of implicit measures and their functioning. The research covered the experimentation of several implicit measures in two different contexts within the broader domain of mental health: the automatic components of stigmatizing attitudes and behaviours *towards* people affected by a mental disease (Part 1) and the impulsive, automatic processes implied *within* people affected by a mental disease, more specifically, by an alcohol addiction disorder (Part 2).

Part 1 of this dissertation is concerned with the design of two Implicit Association Tests targeting two aspects of mental illness stigma, namely, aetiological beliefs and prejudicial attitudes. The main objectives were to verify whether these two measures could be used as assessment techniques in this particular framework and to explore the plausible existence of implicit complements of mental illness stigma.

Part 2 doubled the perspective of this research by experimenting implicit measurement techniques as means for change by adapting them to retrain the implicit processes they were initially designed to assess. The study took the form of a Randomised Clinical Trial with alcohol addict outpatients in which the combination of two training paradigms targeting maladaptive impulsive processes towards alcohol (i.e., attentional bias and approach bias) is examined.

In both studies, the measurement properties of the implicit measures developed and their meaning in relation to the theoretical to-be-measured psychological attributes have been explored within a Rasch modelling perspective, through the application of the Many-Facet Rasch Measurement (MFRM) model.

In Part 1, the MFRM model allowed disentangling the different ‘ingredients’ contributing to the emergence of the IATs effect and highlighting how implicit aetiological beliefs and evaluative associations with mental illness are multifaceted aspects. Semantic and evaluative implicit associations with mental illness resulted to be dependent on diagnostic categories and differently determined by biologic semantic associations and by a positive association primacy, respectively. Further, the MFRM evidenced the functioning of the IAT at the microscopic level.

In Part 2, analysis of data of a group of participants at pre- and post-intervention assessment sessions evidenced the first promising results of the RCT: although participants did not show a substantial change in their alcohol attentional and approach bias measures, the MFRM showed a changing process in action. Experimental conditions showed to have a differential effect in bringing in a decrease and/or a reversal of the two cognitive biases. The MFRM contributed to the exploration of the dimensional and theoretical status of the two cognitive bias implicit measures and provided informative clues about their general and domain-specific features. Further, the MFRM retrieved first evidence about a differential effect of the stimuli used in improving control processes over the impulsive reactions towards alcohol.

The intertwined elements of this work, namely, implicit measurement, mental health, and Rasch modelling, have been combined in the attempt not only to clarify the benefits of implicit methods in psychology, but also to unravel what it actually means to use implicit measures. The combination with a rigorous modelling approach indeed demonstrated both the limitations and the strength of this new family of instruments.

RIASSUNTO

La ricerca scientifica in psicologia è intrinsecamente legata alla misurazione di variabili che per natura sono mutevoli, presentano un'elevata complessità e molto spesso non sono direttamente osservabili. Lo sviluppo di metodi di misurazione è funzionale alla ricerca di un mezzo per mettere in luce le diverse sfaccettature della variabile psicologica di interesse. Gli ultimi quindici anni hanno assistito ad un enorme sviluppo e applicazione di un nuovo insieme di strumenti di misura note come *misure implicite*, le quali hanno come scopo primario quello di quantificare quelle variabili psicologiche definite come automatiche, incontrollabili, inconsce, impulsive, o implicite. L'obiettivo principale di questo lavoro è stato quello di esplorare la natura propriamente implicita di alcune di queste misure, insieme al loro funzionamento. Il progetto di ricerca ha incluso la sperimentazione di alcuni metodi di misura impliciti in due diversi contesti all'interno del più ampio ambito della salute mentale: da una parte lo studio delle componenti automatiche nei processi di stigmatizzazione nei confronti di persone affette da un qualche disturbo mentale (Parte 1); dall'altra la considerazione dei processi impulsivi e automatici in persone affette da uno specific disturbo mentale, quale la dipendenza dal alcol (Parte 2).

La Parte 1 della tesi include lo sviluppo di due Implicit Association Tests destinati alla valutazione di due aspetti inerenti lo stigma verso la malattia mentale: le credenze eziologiche e gli atteggiamenti pregiudiziali. Gli obiettivi principali hanno riguardato la verifica del possibile utilizzo di queste misure come strumenti di valutazione in questo specifico ambito, e nel contempo dell'effettiva esistenza di una controparte implicita nell'espressione dello stigma verso la malattia mentale.

Nella Parte 2 la prospettiva ha assunto un'ulteriore duplice veste attraverso la sperimentazione delle tecniche di misurazione implicita come strumenti di cambiamento, attraverso il loro adattamento alla funzione di training di quei processi impliciti inizialmente misurati. Lo studio ha preso la forma di un Trial Clinico Randomizzato (TCR) con pazienti ambulatoriali dipendenti da alcol, nel quale è valutata la somministrazione di una combinazione di due training per il trattamento dei processi cognitivi automatici disfunzionali (i.e., bias attentivo e di approccio) implicati nella dipendenza da alcol.

In entrambi gli studi sono state esplorate sia le proprietà misurative degli strumenti sviluppati, sia la loro relazione con l'ipotetica variabile psicologica misurata all'interno di una prospettiva di modellazione a tratti latenti, attraverso l'applicazione del Many-Facet Rasch Measurement model (MFRM).

I risultati ottenuti nella Parte 1 mostrano come il modello MFRM sia riuscito a separare i diversi 'ingredienti' che contribuiscono all'emergere dell'effetto IAT evidenziando come le credenze eziologiche implicite e l'atteggiamento implicito nei confronti della malattia mentale siano multi-sfaccettati. Le associazioni semantiche e valutative nei confronti della malattia mentale sembrano cambiare in funzione della categoria diagnostica e sono rispettivamente determinate da associazioni con l'area semantica biologica e da un effetto primacy di associazioni positive. Il modello MFRM ha inoltre reso evidente il funzionamento dello IAT a livello microscopico.

Nella Parte 2, l'analisi di un gruppo di partecipanti nelle sessioni di pre- e post-assessment ha dato i primi, promettenti risultanti sull'efficacia del TCR: nonostante al momento i partecipanti non abbiamo manifestato un significativo cambiamento nelle misure del bias attentivo e di approccio verso l'alcol, il modello MFRM ha dimostrato comunque che c'è effettivamente in atto un processo di cambiamento. Le condizioni sperimentali hanno prodotto un effetto discriminante nell'ottenere la diminuzione o il rovesciamento dei due bias cognitivi. Il modello ha inoltre contribuito all'esplorazione della dimensionalità e delle ipotesi teoriche alla base delle due misure implicite dei bias, dando suggerimenti rilevanti circa le loro caratteristiche dominio-generali e dominio-specifiche. Un ulteriore risultato riguarda un primo riscontro di un effetto esercitato dagli stimoli utilizzati nelle due misure nell'aumentare i processi di controllo degli impulsi nei confronti dell'alcol.

In conclusione, l'intreccio tra misurazione implicita, salute mentale, e modelli di Rasch è nato allo scopo non solo di chiarire i benefici dell'utilizzo delle misure implicite in psicologia, ma anche per svelare che cosa significa effettivamente la misurazione implicita, mostrando sia i limiti che i punti di forza di questa nuova famiglia di strumenti attraverso la combinazione con un approccio metodologico e modellistico rigoroso.

Table of Contents

Introduction

Doing research in psychological sciences is a compelling activity. You don't always see the object of your study. You cannot touch it, you cannot concretely hold it in your hands, you cannot apply a physical force to test its strength, and you cannot bear it from one place to another. You cannot take a picture of psychological attributes, such as self-esteem, personality traits, or intelligence, and save it as it appears in a precise moment, like an immutable object, crystallised in one single time frame. It is simply impossible, for the intangible, unstable, changeable and extremely complex nature of the human mind, or, in more traditional psychological terms, *psyche*.

Research in psychological science is challenging because it is fundamentally based on *inference*. You infer from a person's response to questions like "How much are you satisfied with yourself?" or "How much are you inclined to feel that you are a failure?", and so on, that this person has a high or low level of self-esteem. Ability tests or educational attainment tests are built to *infer* how much a person is able to efficiently solve problems of various nature, depending on the enquired skill and school subject they are targeted to. In both cases, the psychology researcher has designed a tool by means of which (s)he aims to index the desired psychological attribute and which is assumed to represent *concretely* some of the possible ways the attribute of interest can manifests. Even in the field of neuroscience, where the physical underpinnings of psychological cognitive processes are at the core of the scientific process, - and you might give the researcher the impression of holding something more concrete in his/her hands than the construct of self-esteem –, the main operation of finding some connections between a neuroanatomical region of the brain or a particular neural network with a certain cognitive function, let's say object recognition, is based on *inference*. There is still not a direct, visible relation between the two objects, such as the relation between a certain frequency wavelength and a certain acute or deep sound.

The core feature of doing research in psychological sciences is finding the best way to approximate as much as possible to the "lie of land". The strain to get closer and closer is raised at *n* times. And the efforts that one puts in achieving the goal have to systematically take into account that what we are looking at and what we arrive at after an often long, tiring research is a *representation* of the object we are studying. More precisely, one of the innumerable possible representations.

The reflection brought so far is not to be interpreted as a complaint about doing psychological research. Rather, it is a relativistic view accrued along the course of a growing expertise in this field. What I mean is that it is possible to do scientific research, in the strict sense of term scientific, in psychology. But it is achievable only if one accepts some constraints. And here it comes the topic of this work.

The operation of measuring something is bounded to that of finding relations between elements. By measuring two variables, A and B, the attempt is to understand the relation between the two measurement outcomes. Just like the length of a distance and the time it takes to cover it. You can get the speed with which you went through. It seems simple, isn't it? However, in psychology things aren't so definite.

In the history of psychological sciences, a lot of efforts have been put forward to reach the standards of measurement that are inherent the physical sciences. The meaning of measuring a psychological attribute and the ways to do it is probably one of the biggest afflictions when attempting the concretization of an idea about a certain psychological process or attribute. Still, any kind of research cannot disregard the operation of measuring the variables of interest. Otherwise, the process os testing hypotheses would be based on purely subjective observations and conjectures, which don't have a reference frame to compare to. The first chapter of this work gives an overview of the evolution of measurement theory and practices in psychology to delineate the framework wherein this work generated.

In particular, a specific type of measurement techniques is the keystone around which the reflections brought so far have been turned to: the *implicit measurement*. In the last fifteen years an enormous interest and body of research in various fields of psychological science has been addressing the development and use of more indirect measurement procedures to assess psychological attributes interchangeably defined as unobservable, automatic, unconscious, involuntary, uncontrollable, or *implicit*. The last decade has even been re-named the "New Age of Measurement" (Nosek, Hawkins, & Frazier, 2011) for the massive spread of measurement procedures such as the *Implicit Association Test* (Greenwald, Schwatrz, & Banaji, 1998) or the revival of older implicit measures such as the *Evaluative* and *Semantic Priming* (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Wittenbrink, Judd, &

Park, 1997). The impressive use of this measures in social psychology, clinical psychology, experimental psychology, personality psychology, and even marketing and work psychology, would impress also the most sceptical reader about the flexibility and adaptability of this new family of measures.

But what do these implicit measures actually do? How do they work? Can they be applied to index almost any unobservable psychological attributes? The first answer would be yes. But a more accurate comprehension of how an implicit psychological measure is would let you change your mind. The first chapter introduces what implicit measurement is, what it means and how it works, as well as an overview of the implicit measures sub-categories with related criticisms, the new horizons in the use of these instruments and the theoretical insights they brought in, and still are suggesting, within the understanding of human (social) cognition.

Besides the theoretical presentation of implicit measurement, the present work covered the experimentation of implicit measures in two different frameworks within the broader domain of mental health and which have in common the involvement of supposedly automatic, implicit, involuntary, uncontrollable, psychological processes: the automatic components of stigmatizing attitudes and behaviours *towards* people affected by a mental disease (Part 1) and the impulsive, automatic processes implied *within* people affected by a mental disease, more specifically, by an alcohol addiction disorder (Part 2).

Stigma towards mentally ill people is a complex phenomenon, composed of a variety of processes that forerun the manifestation of a discriminatory behaviour and that encompass stereotypes, beliefs, prejudicial attitudes, and cultural and social norms. Given the strong resistance and ambivalence in acknowledging a prejudicial attitude or behaviour towards someone else, the assessment of various facets of stigma is quite challenging and further poses several theoretical questions about their nature. Two IATs targeting two aspects of mental illness stigma, namely, aetiological beliefs and attitudes have been then designed with a two-fold objective (Chapter 2). On one hand, to verify whether they could be used as assessment techniques in this particular framework, on the other hand, to explore the plausible existence of implicit complements of mental illness stigma.

The double side of implicit measurement at the service of mental health was further doubled by the experimentation of implicit measurement techniques as means for change, by adapting them to the function of training of those implicit processes they were initially designed to assess. This ambitious goal took the form of a Randomised Clinical controlled Trial (RCT) with alcohol dependent outpatients, which is currently on going in an national public health service for addiction (SerD) (Chapter 4). The RCT was designed to evaluate the combination of two training paradigms targeting impulsive processes towards the substance of addiction, with the main objective of using the same implicit measures for the assessment of the strength of cognitive biases towards alcohol as a way to reduce them, or at least reversing them towards an aversive cue as salient as alcohol (non-alcoholics).

So far, the present work proceeded on three main fronts: measurement, implicit measurement, and mental health. The last file rouge embracing the entire work is the endorsement of a common methodological perspective, i.e. the use of a modelling approach that could allow an accurate and detailed psychometric investigation of the measures developed along the two studies to fulfil the primary guiding reflection about the meaning of measurement in psychological sciences. That was achieved by operating within a Rasch latent modelling framework. The choice of this modelling perspective was driven by the necessity of a rigorous mathematical tool to define the fundamental properties of the measures devised. Rasch models were originally developed in the attempt to combine in one unique modelling approach the precision of the measurement models in the physical science and the stochastic nature of the research objects of psychological science. The best of the two worlds.

The application of the Many-Facet Rasch Measurement model to the two IATs (Chapter 3) and to the implicit measures of alcohol-related cognitive biases (Chapter 5) has probably been the most difficult challenge, for the complexity of the variables assessed and the simultaneous feedback from the model telling you that sometimes something different from what was expected was going on.

Nonetheless, the accounts given by the MFRM results provided precious information about both the functioning of the measures and the theoretical status of what we have been measuring, which, at the very last, I think is one of the greatest conquests when doing psychological research.

Chapter 1

From measurement to *implicit* measurement in psychological science

1.1 The measurement tradition in Psychology

Doing research in the human sciences entails the formulation of hypotheses regarding any attribute or process pertaining to the human behavioural repertoire and their verification usually through an experimental approach. The latter serves the collection of empirical data, which the presumed theory is confronted with. The scientific process guiding research is then strictly and inherently connected to the procedures and methods used to *measure* the object of interest and cannot disregard methodological questions such as "does this test (or task or survey) measure what we intend to measure?", "what are we actually measuring with it?", or "does it provide a valid measure of what we are interested in?".

The definition of measurement in the social sciences has always been a controversial issue and the discussion about its fundamental theoretical roots and practical implementation is still on going (e.g., Borsboom, 2006; Markus & Borsboom, 2012; Michell, 2008).

"Scientific measurement is properly defined as the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute" (Michell, 1997, p. 358) such as for physical sciences.

This definition of measurement roots in the earlier work by Hölder (1901), whose paper presented a precise characterization of attributes' quantitative structure and its relation with the real number system, which conceptually founded scientific measurement (Michell, 2008). The definition of measurement, so conceived from the physical sciences, relies upon the speculative assumption that the to-be-measured attribute has a quantitative structure. Unless there are direct empirical proofs for it, or, as for many physical quantities, the existence of which is taken for granted (e.g., temperature and density), the evidence that the to-be-measured attribute is quantitative is entirely indirect.

Measurement always presupposes theory: the claim that an attribute is quantitative is, itself, always a theory and that claim is generally embedded within a much wider quantitative theory involving the hypothesis that specific quantitative relationships between attributes exist. It follows that when scientists *measure* something, they are testing firstly the hypothesis about the quantifiable nature of the object, which may in principle be false. As Michell (1997) states, the establishment of a quantitative science fulfils two tasks: the first one concerns with the scientific investigation whether the object has a quantitative structure; the second one involves devising procedures functional to the measurement of the magnitudes of the attribute identified to be quantitative.

In psychological sciences, the measurement of intellectual abilities, personality traits, attitudes, and so on, which are all part of the psychology matter, is based on the supposition that these attributes are inherently quantitative.

Noteworthy is the most shared and common definition of scientific measurement in the psychology realm by Stevens (1946): *Measurement is the assignment of numerals to objects or events according to some rule*. This definition of measurement is quite unlike the above-mentioned traditional concept of measurement used in the physical sciences.

Following the *quantitative imperative* (Michell, 1990), psychologists and scientists doing research in psychology mostly focused on the instrumental task of devising procedures to assign numbers to attributes, ignoring the first task of testing for the quantitative structure of the attributes of interest. After Fechner (1860) foundation of quantitative methods in psychophysical research, Spearman was influenced by the Fechnerian spirit under Wundt's mentoring and was interested in the number of intellectual abilities involved in resolving tasks and was, once again, not sensitive to the issue of whether the postulated abilities were structurally quantitative or not.

In 1932, the British Association for the Advancement of Science appointed a committee to investigate the possibility of quantitatively estimating sensory events, namely the British Ferguson Committee, which opined that psychophysical methods did not constitute scientific measurement and whose chair A. Ferguson was a physicist (Ferguson et al., 1940). In response to the committee's requirements and to solve the controversy about the scientific state of measurement practices at that time, Stevens put forth his rationale to justify the measurement practices, as have been carried out so far and in contrast to the classical definition of measurement taken form the physical sciences, through creatively translating them into representational terms. For Stevens, "measurement is possible in the first place only because there is a kind of isomorphism between (1) empirical relations among objects and events and (2) the properties of [numerical systems]"(Stevens, 1951, p. 1). From this starting point he developed his theory of the four possible types of measurement scales (nominal, ordinal, interval, and ratio) and the associated doctrine of permissible statistics. The responses to the Ferguson committee's requests were then mainly two: on one side, accepting the traditional definition of measurement as implied in the physical sciences so far, whereas, on the other side, the proposal of a new definition, with the purpose to establishing the scientific nature of psychological research.

The two divergent measurement definitions are nowadays reflected to a large extent within alternative approaches to measurement. For example, methods based on covariance matrices are typically employed on the premise that numbers, such as test raw scores, *are* measurements. Such approaches implicitly endorse Stevens' definition of measurement, which requires only that numbers be *assigned* according to some rule. The main research task, then, is generally considered to be the discovery of associations between scores, and of factors posited to underlie such associations. On the other hand, when measurement models such as the Rasch model are employed, numbers are not assigned based on a rule. Instead, specific criteria for measurement are stated, and the objective is to construct procedures or operations that provide data fulfilling the relevant criteria.

A theoretical approach is then presumed to underlie the measurement practice: one of the main breakthroughs of the past century in the psychometric thinking about measurement consists in the realization that measurement does not mean finding the right observed score to substitute for a theoretical attribute, but devising a model structure to relate an observable attribute to a theoretical attribute. An essential precondition for this realization to occur is that, either intuitively or explicitly, one already holds the philosophical idea that theoretical attributes are, in fact, distinct from a set of observations, i.e., that one rejects the operationalist thesis that theoretical attributes are synonymous with the way they are measured, as

Steven's definition of measurement put forth (Bridgman, 1927). The cognitive revolution actually promoted the rejection of operationalism, unluckily without such a striking success. The dominant idea in the psychology mainstream is that one has to find an "operationalization" (i.e., an observed score) for a construct, after which one carries out all statistical analyses under the false pretence that this observed score is actually identical to the attribute itself. This approach brings with it the tendency, which has become by now a well-established practice, of naïvely applying the properties that pertain to the sumscore (e.g., linear ordering) to the attribute object of interest. For instance, people can be linearly ordered according to the attribute they present, for the attribute sumscore has, by definition, the property of linear ordering (Borsboom, 2006).

This theoretical and pragmatic viewpoint falls under the dogmatic and uncritical assumptions of Classical Test Theory, which axiomatically fixes the link between the theoretical attribute (the true score) and the observation (the test score), and does not leave any room for review or discussion. This is in contrast with the modern test theory models, which flexibly describe the relation between theoretical attributes (i.e., latent variables) and test scores (Mallenbergh, 1994) and firstly focus their attention on the nature and forms of this relation instead of primarily and uniquely considering how well test scores correlate with other test scores (i.e., convergent validity).

This is what we try to do in this work: spelling out the structure of an attribute, its dimensionality, and the link between the structure and the "score" of the measurement instruments ad-hoc devised.

1.2 What is a measure?

In psychological research, any experimental or observational study cannot disregard the measurement of any attribute, or variable of interest, to pursue the objectives and test the hypotheses driving the research. Psychological measures are then devised to disclose individuals' internal psychological attributes. It follows that, an ideal, perfect psychological measure should provide an exact index of, or be a proxy for, the extent to which a person possesses the attribute that the measure is aimed at quantifying. Borsboom (2006) and Borsboom, Mellenbergh, and van Heerden (2004) discussed thoroughly about the status of measurement methods and practice in psychological sciences and provided a sharp normative framework of what a valid measure is. As pointed out by Borsboom et al.,

> "a test is valid for measuring an attribute if and only if (a) the attribute exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure" (2004, p. 1061).

This statement implies that when a measurement procedure is applied to a certain person, a hypothetical attribute within the person causes an observable outcome, which can then be used to make an inference about the attribute itself. The *measurement outcome* obtained by a specific *measurement procedure* is then assumed to reflect the to-be-measured attribute (De Houwer, 2006; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009a). For instance, the responses to the items (*measurement procedure*) of a questionnaire on the level of individual self-esteem are summed up to give a score (*measurement outcome*) reflecting how much people positively evaluates themselves. The *measurement procedure* defines the specific set of guidelines about which actions to take when the measurement task is on the run.

This first assumption goes hand in hand with the presupposition that the to-be-measured attribute does exist in some form and that it does affect behaviour. This second assumption refers to the ontological claims regarding measurement (Borsboom et al., 2004), the discussion of which would be afield of the present work, but which are fairly illogical when measuring. If the outcome is a measure of the attribute, the attribute must exist and must causally influence the outcome.[1] The necessary condition is then to empirically demonstrate that the to-be-measured attribute actually caused the measurement outcome, for the establishment of the validity of a measure is intrinsically connected to the nature of the psychological attributes. Measure validity and psychological attribute status go hand in hand.

The above-defined concept of validity further implies the concept of causality. If it is assumed that the attribute causes the measurement outcome, then any variation in the outcome should suggest variations in the psychological attribute. This is empirically testable via an experimental approach, in which the to-be-measured attribute is experimentally manipulated and the effects of the

[1] For a critique about views on the ontological claims regarding measurement and about the debate in philosophy of science and psychometrics see Borsboom (2005, 2006), Michell (1997), and the special issue of *Measurement* (2008, issue 6) devoted to the conceptual foundations of measurement.

manipulation are examined, for instance, by comparing the pre- and the post-manipulation measurement sessions (see Borsboom et al., 2004). This should then provide empirical evidence for the variations in the measurement outcome that are linked to some form of variation in the underlining psychological attribute. The advantages of an experimental approach to verify measurement validity can then ascertain whether an attribute causes an outcome, give a clue on how variations may occur (Wentura & Rothermund, 2007), and allow optimizing the measure in the sense of maximizing the effects of the attribute on the measurement outcome.

Nevertheless, this does not mean that a correlational approach is wrong or worthless. Despite the fact that correlational studies cannot draw conclusions about (a) causal inferences and (b) the relation between the to-be-measured attribute and the measurement outcome, they may be beneficial for clarifying the hypotheses about the nature of the psychological attribute that affects the measurement outcome. For instance, if a measurement outcome does systematically correlate as expected with other measures, it would be unlikely that these correlations are due to other hidden third factors or by chance. A correlational approach could be seen as a first step in the study of a measure validity, because of the simpler study design and the more efficient data collection, and give a first insight into the relational network of that measure (De Houwer et al., 2009a), the so called convergent validity.

In summary, the discussion about the validity of a measure refers to (1) the properties of the measurement outcome rather than to the measurement procedure itself, (2) the validity of the underlining ontological assumptions about the psychological attribute causing the outcome, and (3) the use of both a correlational and experimental approach to test it empirically. A further elucidation in the work by Borsboom et al. (2004) considers the distinction between the validity of a measure and its *overall quality*. They argued that a valid measure is not necessarily reliable or predictive of criterion variables and could even measure different attributes in different groups of respondents (Borsboom et al., 2004, p. 1070). This is because a measure can be a valid index of a psychological attribute even if this attribute is not the only source of variation in the measure. “Validity implies that the to-be-measured attribute causes variation in the measure but does not rule out the possibility that other attributes or situational factors are additional sources of variation” (De Houwer et al., 2009a, p. 350). It is then important to be sure not only that the to-be-measured attribute actually causally affected the measurement outcome, but also to verify whether there are other sources of variance, or confounders, that can alter the measure.

According to the definition of measure – "*a measurement outcome that is causally produced by the to-be-measured attribute*" (De Houwer et al., 2009a, p. 350, italics is mine) – and to the conceptual framework by Borsboom (2006) and Borsboom et al. (2004), De Houwer et al. (2009a) postulated that an ideal psychological measure should conform to two normative criteria: "[…] (a) which attributes causally produce the measurement outcome [*what criterion*] and (b) how these attributes causally produce the measurement outcome [*how criterion*]." (p. 350).

Both theoretical and empirical research are required to test the fulfilment of these criteria and to give clarity to the nature of the psychological attribute a measure is aimed at. Furthermore, both correlational and experimental empirical studies can be of great advantage to establish which attributes cause the measurement outcome and how they do it.

1.3 The New Age of Measurement: the introduction of implicit measures

In the last 15 years, psychological sciences have seen the introduction of a new family of measures of psychological attributes, aimed at signalling those mental processes occurring behind human behaviour that traditional measurement methods – self-report – fail to capture. These new measurement techniques, the so called *implicit measures*, have given birth to a new psychological research mainstream and provided a new 'ruler' for quantifying individual differences.

But where do these implicit measures come from?

Understanding what is going on in the human's mind, beyond the explicit contents of speech and manifest behaviour, has always been an exciting challenge in the historical development of disciplines dealing with the human kind. Since Plato and Aristotle discussion about consciousness and intentional behaviour, going through the philosophical debate about the limits of human introspection and understanding (e.g., Augustine), and arriving at the beginning of the 20th century with the acknowledgement of the Freudian theory about the existence of an inaccessible side of human mental experience (i.e., the *Es* or unconscious), the fact that there is something more happening in people's mind than what they say has inspired a long list of theories, conceptual frameworks, and outstanding ideas. What has further captured the reasoning machinery in centuries of research endeavours in understanding human nature, is also the acceptance, albeit difficult, that there is more happening in everyone's mind beyond what we say, pointing to the limited reach of the introspective experience.

The complete ownership of one's mind has certainly the unquestionable feature of illusionary thinking: what we experience in any moment and what is occurring in our mind while we are experiencing any mental state are not alike. The belief that the two of them are the same thing has a somewhat self-assuring function and is part of the identity and awareness of oneself, otherwise the anguishing feeling of uncontrollability of what is happening in one's own psyche would undermine the security and flexibility in the execution of any simple mental operation and daily behaviour.

Despite the belief that how we think equals what we think, the mental experience is not completely overlapped with mental processes. The actual causes of a certain action may not be completely related to what we may report when asked why we did it, since we can be as confident as incorrect in expressing our beliefs about the reasons for that performance.

In their forerunner work, Greenwald and Banaji (1995) introduced the term 'implicit social cognition' to describe cognitive processes that occur outside the conscious awareness or conscious control regarding social psychological constructs such as attitudes, stereotypes, and self-concepts. Implicit processes can be considered as the ''dark matter'' of the mind – the mental processes that operate in the absence of conscious awareness (Schacter, 1987). Although hidden from the outside view, implicit processes appear to drive much of social behaviour, particularly when responses are made quickly and spontaneously, without conscious deliberation (Greenwald & Banaji, 1995). The conceptual framework of implicit social cognition has then provided a useful explanation for why people often behave differently or in contradiction from their explicit beliefs and intentions, such as left-winged and egalitarian political ideology can indeed reveals racial stereotypes.

Greenwald and Banaji defined implicit constructs as "[…] introspectively unidentified (or inaccurately identified) traces of past experience that mediates [a relevant category of responses]" (1995, p. 5). For example, the relevant category of responses for the construct attitudes might be the evaluations of social concepts. This definition points out the substantive features of implicit (social) cognition (Nosek, Hawkins, Frazier, 2012): first, the content need not to be accessible to people's awareness or intentionally used to drive evaluative judgment and/or action, as people may be explicitly inclined to follow a plan of action, but still behave differently because of cognitions operating implicitly, or, better say, *automatically* (Moors & De Houwer, 2006). Secondly, the content of cognitions implicitly affecting behaviour cannot reach the explicitness and propositional status of deliberative reasoning, because of a variety of impediments to a fully, comprehensive thoughtfulness of processes and cognitions operating, by definition, implicitly (De Houwer, 2006).

The descriptive function of this definition of implicit cognition yields the consideration of 'implicit' as an umbrella concept for cognitive component processes which work silently and automatically, and that have unique influences on thinking and behaviour, leading to a richer network of constructs and theory that can provide specific predictions about resulting behaviour (Moors & De Houwer, 2006).

The introduction of the implicit social cognition framework to extend the understanding of behaviour antecedents and driving processes has then grounded a large body of theory-driven empirical and theoretical research on what is implicit and how implicit/automatic constructs and mental processes relate to their explicit/controlled counterpart and, eventually, to the behavioural outcome. The surge of this new approach to the study of psychological processes grounded the consequential development of a new family of measurement instruments in the psychological sciences – the *implicit measures* –, of which the *Implicit Association Test* (IAT – Greenwald, Schwartz, & Banaji, 1998) has been the forefather[2].

One might argue why self-report/explicit measures couldn't be useful in the measurement task in implicit cognition research. It is important to stress that the invention of implicit measures occurred especially because self-reported social

---

[2] Sequential priming tasks (e.g., Devine, 1989; Fazio et al., 1986; 1995) had been already available for several years before the development of the IAT, but in different fields from the implicit social cognition. These tasks were mainly used in attitude formation studies to test the connection between the strength of attitudes and their ability to be automatically activated (Fazio et al., 1986). The design of sequential priming tasks was informed by cognitive theories of automatic and controlled information processing (Posner & Snyder, 1975; Shiffrin & Schneider, 1977) and priming techniques (Neely, 1977), and used to indirectly measure individuals attitudes without asking to report them, feature that later constituted a central theme of implicit social cognition.

cognitions were not as successful predicting some behaviours that they theoretically should be predicting (e.g., differential behaviour with Black or White targets unrelated to self-reported racial attitudes). This does not mean that self-report measures are not accurate, but that sometimes their accuracy and ability to reliably detect the to-be-measured attribute do fail and can be based on information distinct from the actual causes of behaviour. A variety of factors limit the value of introspectively derived explicit measurement (Wilson & Brekke, 1994), as people might have: limits in their motivation to report what they actually think; limits in their opportunity to report the mental content because, for instance, the conditions of measurement might constrain what is reported; limits in their ability to translate mental contents into a report; as well as limits in their awareness, the mental content might simply be inaccessible to introspection (Nosek, Hawkins, & Frazier, 2011). Eventually, in socially sensitive domains, such as racial stereotype or out-group attitudes, for example, responses on self-report measures are often distorted by social desirability and self-presentational concerns.

The introduction of implicit measures targeted these problems, since they are not direct, deliberate, controlled, and intentional self-assessments. An implicit measure assesses mental content without requiring awareness of and/or deliberate control over the relation between the response and the measured content. The response to an implicit measure is used to infer the mental content rather than pointing directly to it, sketching out the *indirectness* as the signature feature of this family of new *measurement procedures*.

In the last fifteen years the application of the IAT and its derivates, and the development of second-generation implicit measures (§1.3.2) has proliferated exponentially and has collected a massive body of results across different content domains in psychology, promoting the identification of a 'New Age of Measurement' in psychological research (Nosek et al., 2011). In their seminal paper, Greenwald and Banaji (1995) ended their review with a call for the refinement of individual difference measures of implicit social cognition. They predicted that "when such measures do become available, there should follow the rapid development of a new industry of research on implicit cognitive aspects of personality and social behaviour" (p. 20). Their prediction has not been disappointed. With the development of the IAT (Greenwald et al., 1998) implicit social cognition research seemed to reach the mountain peak, reaching the point that, nowadays, implicit social cognition has become almost synonymous with research using implicit measurement procedures (Payne & Gawronski, 2010).

Before giving an overview of the most common and used implicit measures, it is useful to give a definition of what an implicit measure is, what it does, and, under the broader perspective of measurement theory, what does it measures.

1.3.1 What is an *implicit* measure?

A central feature of implicit measures is that they are designed to capture psychological attributes, such as attitudes, stereotypes, beliefs, self-esteem, without asking the participant for a self-evaluation of these attributes. For most implicit measures, the construct of interest is *inferred* through a within-subject experimental design, in which the measurement procedure is to compare the individual behavioural performance (e.g., response latency, categorization errors) between some key-conditions (i.e., different primes/stimuli or response configurations). This procedure is essentially different from explicit, self-report measurement in which the attribute of interest is assessed directly and the response is assumed to reflect it.

A controversial issue in the application and study of implicit measurement is concerned with what the term *implicit* means. The label 'implicit' is indeed applied to a family of measures and processes that have in common the fact that they are not direct, deliberate, controlled, and intentional self-assessment. Some researchers have used the term *implicit* to describe a particular characteristic of measurement procedures considered to be a proxy for psychological attributes without requiring participants to verbally report the desired information (e.g., Fazio & Olson, 2003; Nosek & Greenwald, 2009). Yet, other researchers have used the term *implicit* to describe the constructs assessed by a particular class of measurement procedures, namely constructs assessed by tasks that do not require conscious introspection, and therefore might reflect psychological attributes that are introspectively inaccessible (e.g., Banaji, 2001).

The conceptual confusion in the use of the term implicit has produced a substantial amount of theoretical and conceptual literature aimed at a normative taxonomy of what makes a measure *implicit* (e.g., De Houwer, 2006; De Houwer & Moors, 2007, 2010, 2012; De Houwer et al., 2009a, 2009b; Gawronski & De Houwer, in press; Moors & De Houwer, 2006; Moors, Spruyt, & De Houwer, 2010). The result of this debate is best displayed in the conceptualization of the feature implicit as *automatic*, as both terms have been used to describe the conditions under which psychological attributes and processes operate (e.g., De Houwer et al., 2009a; De Houwer & Moors, 2007; Moors & De Houwer 2006). More specifically, a measurement outcome may be described as *implicit* if the impact of the to-be-measured psychological attribute on participants' responses is unconscious, efficient, unintentional, resource-independent, and/or uncontrollable. Conversely, a measurement outcome may be described as *explicit* if the impact of the to-be-measured psychological attribute on participants' responses is intentional, resource-dependent, conscious, and/or controllable (cf. Bargh, 1994; Moors & De Houwer, 2006; Moors, Spruyt, & De Houwer, 2010). In their normative conceptualization of implicit measures, De Houwer et al. (2009b) stated,"[…] the implicitness of a measure is determined only by the automaticity features of those processes by which the to-be-measured attribute causes the measurement outcome" (p. 378). Hence, an implicit measure can be defined as

> "[a] *measurement outcome that reflects the to-be-measured construct by virtue of processes that are uncontrolled, unintentional, goal-independent, purely-stimulus-driven, autonomous, unconscious, efficient, or fast*" (De Houwer & Moors, 2007, p.).

The advantage of equating the term implicit to automatic is that it encompasses all the previous, fuzzy, and informal definitions of the concept of implicit. Indeed, "automaticity is not an all-or-none property of mental processes but refers to a set of features that do not necessarily co-occur within each automatic process" (De Houwer & Moors, 2012, p. 183). Given that most automaticity features are defined in terms of mental constructs such as (proximal – distal) goals, awareness, required processing resources, and time (Bargh, 1994; Moors & De Houwer, 2006), it should be important to specify which automaticity feature a process or attribute is assumed to possess (De Houwer, 2006; De Houwer et al., 2009a). For instance, stereotype activation is known to be automatic as it occurs in the absence of conscious activation of awareness and processing resources, but it is not automatic for its association with certain motives and goals (Bargh, 1992).

According to the broader definition of measure earlier presented (§1.1), which is meant to reveal internal psychological attributes of individuals, an ideal psychological measure should provide an exact index of the extent to which an individual possesses the psychological attribute that the measure was designed to capture (De Houwer, et al., 2009a). This requires the measure to satisfy the *what* and *how* normative criteria (De Houwer et al., 2009a). According to the concept of implicit presented so far, a third normative criterion should be met before a measure can be identified as an implicit measure: The to-be-measured attribute should cause the measurement outcome in an automatic manner. This *implicitness* criterion implies the specification of which automaticity feature is under consideration and the provision of evidence about the automatic nature of the measure.

The terms *implicit* and *explicit* describe the process by which a psychological attribute influences measurement outcomes rather than the measurement procedure itself or the underlying psychological attribute. Moreover, whereas the classification of measurement outcomes as implicit or explicit depends on the processes that underlie a given measurement procedure, measurement procedures may be classified as *direct* or *indirect* on the basis of their objective structural properties (De Houwer, 2006; De Houwer & Moors, 2010). A measurement procedure can then be deemed as *direct* when the measurement outcome is based on participants' subjective evaluation of the to-be-measured attribute, for instance when participants' level of self-esteem is inferred from their self-reported evaluation of themselves. Conversely, a measurement procedure qualifies as *indirect* when the measurement outcome is not based on a self-assessment, for instance when participants' attitudes are inferred from their reaction time performance in a speeded categorization task, or when it is based on the self-assessment of attributes other than the to-be-measured attribute, such as in the case of evaluative priming tasks in which participants' attitudes are extrapolated from their self-reported liking of a neutral object that is quickly presented after the target prime (Gawronski & De Houwer, in press).

The use of two bipolar concepts of *implicit/explicit* and *indirect/direct* finds its roots historically in implicit memory research (De Houwer et al., 2009b)– which has also influenced the implicit (social) cognition framework from the very beginning (Greenwlad & Banaji, 1995). Calling a measurement procedure implicit may force the incorrect conclusion that the processes underlining the measure are also implicit. This led implicit memory researchers to adopt the term *implicit* to refer to a particular type of memory (i.e., the unconscious or unintentional impact of past events on current events) and the term *indirect* to refer to a particular type of memory task (i.e., a task that does not require participants to consciously or intentionally take into account past events) (see Butler & Berry, 2001; Richardson-Klavehn & Bjork, 1988).

A second reason for the use of the term *implicit* only when referring to a measurement outcome is that it seems fairly illogical to use the adjective implicit when referring to a measurement procedure. There is nothing *implicit* about a measurement procedure because it is simply an objective set of guidelines about what to do. Conversely, a measurement outcome can be meaningfully implicit, according to the processes by which the to-be-measured construct is translated into the measure (De Houwer & Moors, 2007).

1.3.2 Overview of most common *indirect* measures

The use of implicit measures in the psychological sciences has a longer history than the last 15 years, with roots in the mid-1980s when researchers adopted Evaluative Priming (EP) tasks from cognitive psychology to study the automatic activation of attitudes (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986) and stereotypes (Gaertner & McLaughlin, 1983), or even earlier when they designed the *Thematic Apperception Task* (TAT – Morgan & Murray, 1935) to evaluate implicit motives which are not directly available to introspection (for a discussion on the implicitness of TAT measure see McClelland, Koestner, & Weinberger, 1989; De Houwer et al., 2009a). Another pioneer of implicit measures has been the *Stroop Task* (Stroop, 1935), which has become a well-established instrument in cognitive psychology and psychopathology for the evaluation of inhibitory control cognitive function.

These studies provided the basement for the design of Greenwald et al.'s (1998) IAT and its derived paradigms, such as the *Brief Implicit Association Test* (BIAT – Sriram & Greenwald, 2009), the *Single Category Implicit Association Test* (SCIAT – Karpinski & Steinman, 2006) or the *Single Target Implicit Association Test* (STIAT – Wigboldus,, Holland, & van Knippenberg, 2005), and of other indirect measures of implicit associations in memory, such as the *Go/No-go Association Task* (GNAT – Nosek & Banaji, 2005) and the *Approach Avoidance Task* (AAT – Rinck & Becker, 2007).

These more recent implicit measures are set themselves apart from their ancestors for the assessment of mental associations between concepts (e.g., math-male, Blacks-negative, Self-positive) rather than being directly concerned with mental concepts (e.g., achievement, anxiety, motives).

Over the past decade, the toolbox of available measurement instruments has grown substantially through the development of new paradigms, the refinement of existing tasks, and the adoption of these for the assessment of a variety of social constructs (e.g., attitudes, stereotypes, beliefs, identities, self-esteem). In 2011, Nosek et al. conducted a citation analysis of 20 articles that introduced a new implicit measurement procedure to estimate each measure's impact and use. They found that: (a) the IAT accounted for more than 40% of the total citations and about 50% of citations in 2010, (b) EP was the second most cited with 20% of total citations and about 12% of citations in 2010, (c) a cluster of AMP, GNAT, STIAT, Semantic Priming Task (SP – Wittenbrink, Judd, & Park, 1997), and Extrinsic Affective Simon Task (EAST – De Houwer, 2003) with each of them presenting between 4-6% of the citations, and (d) a recent cascade of new methods suggested that growth of implicit measurement is still on-going (Nosek et al., 2012).

In Table 1.1 currently available and most common task paradigms used in implicit cognition are presented. The various indirect measures can be grouped in three macro-categories, according to the structural and conceptual features they have in common (Nosek et al., 2011): priming tasks, categorization tasks, and approach-avoid tasks.

1.3.2.1 Priming Tasks

Sequential priming tasks have been the first measures of individual differences in implicit (social) cognition (Fazio et al., 1986) and are by now one of the most widely used methods in attitudes research.

"Priming involves presenting some stimulus with the aim of activating a particular idea, category, or feeling and then measuring the effects of the prime on performance in some other task" (Cameron, Brown-Iannuzzi, & Payne, 2012, p. 330).

Table 1.1 Most relevant implicit measurement procedures used in implicit cognition research: type of task, applications, target/attribute elements, and range of reliability estimates. (Table adapted from Gawronski & De Houwer, in press).

Paradigm	Type of task	Reference	Applications	Target	Attribute	Reliability
Affect Misattribution Procedure	Priming	Payne et al. (2005)	Evaluative, semantic	Individual	Pairs	.70 - .90
Approach-Avoidance Task	Approach-Avoid	Chen & Bargh (1999)	Evaluative	Individual	Individual	.00 - .90[a]
Brief IAT	Categorization	Sriram & Greenwald (2009)	Evaluative, semantic	Pairs	Pairs	.55 - .95
Evaluative Priming Task	Priming	Fazio et al. (1986)	Evaluative	Individual	Individual	.00 - .55
Extrinsic Affective Simon Task	Categorization	De Houwer (2003)	Evaluative, semantic	Individual	Individual	.15 - .65
Go/No-Go Association task	Categorization	Nosek & Banaji (2005)	Evaluative, semantic	Individual	Pairs	.45 - .75
Identification Extrinsic Affective Simon Task	Categorization	De Houwer & De Bruycker (2007)	Evaluative, semantic	Individual	Pairs	.60 - .70
Implicit Association Procedure	Approach-Avoid	Schnabel et al. (2006)	Self-related	Individual	Pairs	.75- .85
Implicit Association Test	Categorization	Greenwald et al. (1998)	Evaluative, semantic	Pairs	Pairs	.70 - .90[c]
Implicit Relational Assessment Procedure	Categorization	Barnes-Holmes et al. (2010)	Evaluative, semantic	Individual	Individual	.20 - .80
Recoding Free Implicit Association Test	Categorization	Rothermund et al. (2009)	Evaluative, semantic	Pairs	Pairs	.55 - .65
Semantic Priming (Lexical Decision Task)	Priming	Wittenbrink et al. (1997)	Semantic	Individual	Individual	n/a
Single Attribute IAT	Categorization	Penke et al. (2006)	Evaluative, semantic	Pairs	Individual	.70 - .80
Single Block IAT	Categorization	Teige-Mocigemba et al. (2008)	Evaluative, semantic	Pairs	Pairs	.60 - .90
Single Category IAT	Categorization	Karpinski & Hilton (2006)	Evaluative, semantic	Individual	Pairs	.70 - .90
Stimulus Response Compatibility Task	Approach-Avoid	Mogg et al. (2003)	Evaluative, semantic	Individual	Individual	n/a
Sorting Paired Features Task	Categorization	Bar-Anan et al. (2009)	Evaluative, semantic	Individual	Individual	.40 - .70

[a] Reliability estimates differ depending on whether approach-avoidance responses involve valence-relevant or valence-irrelevant categorizations, with valence-irrelevant categorizations showing lower reliability estimates (.00-.35) compared to valence-relevant categorizations (.70-.90).
[b] Reliability estimates differ depending on whether the scores involve within-participant comparisons of preferences for different objects or between-participant comparisons of evaluations of the same object, with between-participant comparisons showing lower reliability estimates (.30-.75) compared to within-participant comparisons (~.80).
[c] Reliability estimates tend to be lower (.40 - .60) for second and subsequent IATs if more than one IAT is administered in the same session.

The assumption that the human mind is organized as networks of associations, promoted the development of priming techniques that allow measuring what associations are automatically activated in response to a given stimulus.

The *Evaluative Priming* (EP) was the first task developed for the assessment of automatically activated evaluations when encountering a social object (i.e., stereotype). In this kind of tasks, participants are very briefly presented with a prime stimulus (e.g., a Black face) followed by a positive or a negative target word. In the typical version of the task, participants are then asked to quickly determine whether the target word is positive or negative by pressing one of two response keys (*evaluative decision task* – Gawronski & De Houwer, in press). Whenever the prime stimulus triggers faster responses to positive words (compared to a neutral baseline prime) the prime stimulus is assumed to be automatically positively evaluated. Yet, if faster responses are given to negative words after the prime stimulus presentation (compared to the neutral baseline prime), the prime is then associated with a negative valence (e.g., Wittenbrick, 2007). If the prime elicits the same response as the target, responding is then facilitated with faster responses and less errors, if the prime triggers a different response, it conflicts with the response elicited by the target creating a conflict that needs to be resolved with slower responses and more errors.

The EP can be used to assess evaluative responses to any object that can be presented as a prime in a sequential priming task, both supraliminally (i.e., above the conscious awareness threshold) and subliminally (i.e., below the conscious awareness threshold). One of the advantages of EP is the possibility to separating priming scores for different associations – differently from the standard IAT, for instance, where associations are mixed together – and obtain four single indices for positive and negative associations to the targets (e.g., Black people/positive, White people/positive, Black people/negative, White People/negative), respectively. These indices are computed by comparing response latencies, for instance, to positive words following White versus neutral primes.

Conceptually speaking, this kind of task is similar to the classic Stroop color interference paradigm (Stroop, 1935), in that in both tasks a task-irrelevant feature (i.e., the word content in the Stroop task) or a task-irrelevant stimulus (i.e., the prime in the priming task) is automatically processed, interfering with responding in incongruent cases (Wentura & Degner, 2010).

Procedurally similar to the EP is the *Lexicon Decision Task*, or *Semantic Priming* (SP – Wittenbrink et al., 1997), already available in the early 1970s (Meyer & Schvaneveldt, 1971). What distinguishes the SP from the EP is that (a) the target stimuli are meaningful words and meaningless letter strings and (b) the task is to specify as quickly as possible whether the letter string is a meaningful word or a meaningless non-word (*lexical decision task* – Gawronski & De Houwer, in press). Similarly to the EP, whenever the presentation of a given prime stimulus facilitates quick responses to a meaningful target word (compared to a baseline prime), the prime stimulus is then assumed to be associated with the semantic meaning of the target word. Rather than being concerned with the valence association to target objects (e.g., self and positive), the SP is primarily concerned with semantic associations between a target object and a semantic concept (e.g., self and extraverted).

A variant of the SP includes only meaningful words as target stimuli and requires categorizing them according to their semantic meaning rather than to their valence (*semantic decision task*). For example, Banaji and Hardin (1996) presented prime words referring to stereotypically male or female occupations (e.g., nurse, doctor), which were followed by male or female pronouns (e.g., he, she). Participants' task was to classify the pronouns as male or female as quickly as possible. Results showed that participants were faster in responding to the male and female pronouns on stereotype-compatible trials (e.g., nurse-she, doctor-he) than stereotype-incompatible trials (e.g., nurse-he, doctor-she) (Gawronski & De Houwer, in press; Wentura & Degner, 2010).

Another recent priming measure is the *Affect Misattribution Task* (AMT – Payne, Cheng, Govorun, & Stewart, 2005), in which participants are briefly presented with a prime stimulus that is followed by a neutral Chinese pictograph. The task requires indicating whether they consider the Chinese ideograph as visually more pleasant or visually less pleasant than the average Chinese ideograph. What consistently emerged is that neutral Chinese ideographs tend to be evaluated more favourably when participants have been primed with a positive stimulus than with a negative stimulus. The positive evaluation of the ideographs is not a function of the match or mismatch between prime and target, as for the EP and SP; rather, it results from a misattributing affect triggered by the prime on the neutral target (Payne et al., 2005; Gawronski & Ye, 2014). Priming is then measured as the influence of the prime valence on the frequency of positive judgments, shifting from measuring reaction times to the measurement of accuracy and consequently favouring the measure reliability (Cameron et al., 2012).

Similarly to Fazio et al.'s EP, the AMP can be used to evaluate any kind of stimuli presented as primes in the task. The combination of the procedural guidelines of sequential priming tasks (e.g., compatible and incompatible trials are intermixed rather than blocked) and the higher reliability and effect sizes, compared to the EP and SP (see Table 1.1), makes the AMP a promising indirect measure in implicit social cognition (Gawronski & De Houwer, in press). Indeed, the extension of this task to the investigation of semantic associations, for instance between gender and stereotypical occupations, has already started (e.g., Gawronski & Ye, 2014).

A relevant feature common in sequential priming tasks is that they may be highly sensitive to the target stimuli used, because individual characteristics of the items than specific categories are mostly susceptible to influence the priming effects. Whenever the stimuli belongingness category is made explicit, then the task reliability increases likewise category-driven implicit measures, such as the IAT (Olson & Fazio, 2003).

Sequential priming tasks have been mainly used in attitude and automatic activation of stereotypes research, but extensions to other domains have been advanced, such as prejudice (including race, gender, and groups), consumer preferences, political preferences, personality traits (including self-esteem and self-concept), impulsive behaviour (e.g., eating, drinking, and smoking), clinical psychology (e.g., studies with clinical populations such as depressed individuals), close relationships (for a meta-analysis, see Cameron et al., 2012).

1.3.2.2 Categorization Tasks

When one thinks about implicit social cognition, the first thought that comes up in mind is not what implicit cognition means; rather the first thought goes to the most representative *measurement procedure* in this field: the *Implicit Association Test* (Greenwald et al., 1998). The term implicit social cognition, intended originally in a much broader sense (Greenwald & Banaji, 1995), has by now become a synonymous of research using indirect measures such as the IAT and similar tasks, which are part of the largest group of implicit measures, namely, the categorization tasks.

Whereas sequential priming tasks assess individual's *automatic* responses to selected stimuli presented as primes, categorization tasks identify the specific feature(s) for the processing of a set of stimuli. For example, presenting the picture of a Black man's face in an EP can trigger associations with Black people, men, or with other specific characteristics of the picture, which can influence the responses to the target. In a categorization task, the same Black face can be presented as a stimulus that is to be identified in term of race, gender, or other pre-specified category. This type of tasks is then more sensitive to the pre-defined categories.

The most prominent exemplar and forerunner of categorization tasks is the IAT (Greenwald et al., 1998), which has been and still is the most used indirect measure in social and personality psychology (Nosek et al., 2011) for the assessment of mental representations of associated concepts. The IAT basically consists of two binary categorization tasks combined together so that the sorting task is compatible or incompatible with the to-be-measured psychological constructs. It implies the classification of textual and/or visual stimuli as quickly and accurately as possible into four categories differently labelled, by pressing either a left (e.g., E) or a right (e.g., I) key on a keyboard.

The IAT comprises seven blocks of trials (see Table 1.2): three single practice blocks of categorization of stimuli pertaining to either two target or two attribute categories, and four test blocks (two practice test blocks and two critical test blocks), which involve the simultaneous double categorization of stimuli pertaining to the target and attribute categories combined together on two response mappings presented on the top left and right sides of the screen.

For instance, the race IAT traditionally used in racial prejudice studies requires the categorization of people's faces according to their race (*Whites* or *Blacks)* and *positive* and *negative* words according to their valence in the first two practice blocks. In one of the two double categorization tasks (blocks 3 and 4), the two single classification tasks are combined in such a way that participants have to respond to *Whites* faces and *positive* words, presented in alternating order, with one key (E), and to *Blacks* faces and *negative* words with another key (I). In the other double categorization task (blocks 6 and 7), the target and attribute categories pairing is reversed.

Table 1.2 Task structure of an IAT for the assessment of preference for Whites over Blacks.

Block	Task[a]	Stimuli	N° of trials	Left categories labels	Right categories labels
1	Target Practice	Pictures	20	Whites	Blacks
2	Attribute practice	Words	20	Positive	Negative
3	Compatible practice	Pictures + Words	20	Whites + Positive	Blacks + Negative
4	Compatible Test	Pictures + Words	40	Whites + Positive	Blacks + Negative
5	Reversed target practice	Pictures	20	Blacks	Whites
6	Incompatible Practice	Pictures + Words	20	Blacks + Positive	Whites + Positive
7	Incompatible Test	Pictures + Words	40	Blacks + Positive	Whites + Negative

[a] *The blocks order is counterbalanced across participants, with reversed target practice and incompatible combined tasks (practice and test) completed first in the Incompatible-Compatible order.*

The basic idea underlying the IAT paradigm is that quick and accurate responses are facilitated when the key mapping in the task is congruent to a person's automatic association (e.g., *Whites-Positive* versus *Black*s-*Negative*), but impaired when the key mapping is association-incongruent (e.g., *Blacks-Positive* versus *Whites-Negative*). According to this consideration, the mean difference in participants' response latency in the test blocks, divided by their inclusive response latency standard deviation (for details about the IAT scoring procedure, see Greenwald, Nosek, & Banaji, 2003) is typically interpreted as the measurement outcome indicating the preference for White over Black people or the other way round.

The IAT is a really flexible reaction-time paradigm widely used in a variety of domains in psychology and that can be used to assess almost any type of association between pairs of concepts. For example, by using evaluative attribute dimensions (e.g., positive vs. negative) the IAT can be used to assess the relative preference between pairs of objects or categories. The evaluative attribute dimension may also be replaced with a semantic dimension to assess semantic associations between concepts (e.g., stereotypical associations between female and male people and Literature versus Math). The same flexibility applies to the use of target categories, which may include any pair of objects or categories that can be reasonably contrasted (e.g., male vs. female).

The high flexibility and the easy experimental implementation in any research domain promoted the extensive application of the IAT, and its later derivates, to assess, among others, prejudice, stereotypes, attitudes toward consumer products, self-concept, self-esteem, adult attachment, and any reasonable semantic association between social and non-social objects (for a review, see Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005).

Despite of the easiness of application and the strong internal consistency and test-retest reliability estimates (see Table 1.1), which makes it almost comparable to traditional, well-established explicit measures, the IAT evidenced some critical procedural weaknesses that put it at risk for being influenced by extraneous factors other then the to-be-measured attribute, and/or by method confounds (Teige-Mocigemba, Klauer, & Sherman, 2010). One of the most discussed criticisms of the IAT is its comparative nature, as the IAT gives an insight into the relative associations among four categories, albeit not allowing the distinction in single indices for each category (e.g., absolute preference for Blacks *and* for Whites) (Nosek, Greenwald, & Banaji, 2005). The first to address this problem were Nosek and Banaji (2001), who designed a task for the assessment of single target categories: the GNAT. The task requires a *go* response to target stimuli (e.g., by pressing the space bar) and a *no-go* response to distracter stimuli (i.e., no button press). In one block of the task, the targets include stimuli related to the target concept (e.g., Black faces) and stimuli related to one pole of the attribute dimension (e.g., positive words); the distracters are typically stimuli for the other pole of the attribute dimension (e.g., negative words). In a second block, the classification task is reversed (e.g., *go* for Black faces and negative words, and *no-go* for positive words). Similar approaches to the assessment of evaluative and semantic associations with single target objects have been progressively developed, such as the SCIAT (Karpinski & Steinman, 2006) or STIAT (Wigboldus et al., 2005), the *Single Attribute IAT* (SAIAT – Penke, Eichstaedt, & Asendorpf, 2006), and the novel *Sorting Paired Feature* Task (SPFT – Bar-Anan, Nosek, & Vianello, 2009).

A second source of criticism, which encompasses also the recent single target versions of the IAT, is related to the procedural norm of presenting compatible and incompatible trials in separate, consecutive blocks, which can distort measurement scores through various sources of systematic error variance (Teige-Mocigemba et al., 2010). This practice has been related to the possibility that people might pair the items not only according to the presumed 'compatibility effects' triggered by the associated concepts (e.g., flowers and positive). Other possible sources of 'compatibility effect' can impact on the sorting task, in so far as people pair the items along any salient dimension available at the time and subjectively recoding the task using this salient heuristic. Such 'salience asymmetries' could then create 'compatibility effects' on the tasks that are unrelated to the associations of interest. This issue has been addressed by several new methods that present compatible and incompatible trials randomly in a single block rather than blocked, such as De Houwer's (2003) EAST, the *Single-Block IAT* (SBIAT – Teige-Mocigemba, Klauer, & Rothermund, 2008), the *Recoding-Free IAT* (RFIAT – Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009), and the *BIAT* (Sriram & Greenwald, 2009). The latter has been particularly designed as an attempt to answering to the two main criticisms at once, by reducing the task length, switching between only combined categorization blocks repeated two times each, and enabling the evaluation of a single target object paired alternatively with one of the two attribute categories. Specifically, the BIAT procedure requires the specification of a focal concept in each block as well as a single attribute, instead of two. For example, in the previously described race IAT, although Whites, Blacks, Positive, and Negative stimuli all appear, participants would press one key when White and Positive words appear and another key for "anything else" (i.e, the other two categories stimuli). Subsequently, participants would press one key when Black and Positive stimuli appear and another key when "anything else" appears.

1.3.2.3 Approach-avoid Tasks

Another group of indirect measures falls under the label of approach-avoidance tasks, which adhere to the general assumption that positive stimuli can facilitate approach reactions and inhibit avoidance reactions, whereas negative stimuli facilitate avoidance reactions and inhibit approach reactions. Tasks involving an approach-avoid response pattern incorporate movement toward or away from presented stimuli to detect whether concepts automatically elicit approach or avoidance tendencies.

For instance, the *Implicit Association Procedure* (IAP – Schnabel, Banse, & Asendorpf, 2006) has a similar structure as the IAT but instead of pressing response keys to categorize stimuli to the left or right, participants pull a joystick toward the self (approach) or push it away from the self (avoid). This task was designed to measure self-related associations. For instance, self-associations with shyness are reflected by faster responses when the positively evaluated target concept is mapped to pulling the joystick toward oneself than when it is mapped to the pushing away response (avoid), suggesting that associations can be measured with physical actions of pushing and pulling in relation to the self (Nosek et al., 2011). Similarly, the *Stimulus Response Compatibility Task* (SRCT – Mogg, Bradley, Field, & De Houwer, 2003) requires using arrow keys to move an image of a person toward or away from a stimulus, such as a cigarette or a spider. In a further variant of an approach-avoid task (Rinck & Becker, 2007), a zooming features was added to the task, so that the pushing response makes the stimulus increasingly smaller – giving the perceptual sensation of avoidance – and the pulling response makes the picture increasingly bigger – giving the perceptual sensation of approach.

In some types of approach–avoidance task, the participant's response is based on the picture contents (e.g., push spider pictures and pull neutral pictures), which could be deemed as more direct responses and possibly not satisfying the *implicit* feature. In other types, responding is based on an aspect of the stimulus not related to its content (e.g., push landscape-oriented pictures and pull portrait-oriented pictures). The task would then appear to be more indirect and therefore possibly more implicit when the participant's response is based on the more unobtrusive content-irrelevant aspect, and this procedural difference might influence measurement outcomes (Roefs et al., 2011).

Approach-avoid tasks are presumed to reveal embodied implicit responses to objects in the environment, which are associated to a positive valence. Thus, because approach is related to positive valence and avoidance to a negative valence, performance on these tasks is theorized to reflect affective associations. Approach-avoid tasks have been applied to domains in which approach-avoidance tendencies have implications for social functioning, such as tendencies to approach or avoid (a) drugs and alcohol (e.g., Cousijn, Goudriaan, & Wiers, 2011; Mogg, et al., 2003; Wiers, Rinck, Dictus, & Van den Wildenberg, 2009), (b) food (e.g., de Jong & Veenstra, 2007), and (c) spiders in cases of specific spider phobia (e.g., Rinck & Becker, 2007).

The exponential appearance of recent alternatives to the IAT and of new indirect measures on implicit cognitions does highlight how the field is still on its way to find the better procedure for the selected outcome, and how no measure is perfect. Yet, the effort on the refinement of existing techniques and on the conception of new, outstanding and feasible solutions to address the evidenced criticisms, speak about the rapid ascension of implicit social cognition and implicit measurement in psychology research practice. Further, the identification of weaknesses and theoretical and conceptual issues in the use of implicit measures has progressively promoted the shift in the research focus. Recent directions in the use of implicit measures, such as innovative tools to modify the to-be-measured attribute, and the reflection on the mechanisms and cognitive processes underlying the performance of implicit measures and production of behaviour that escapes the conscious control and awareness, are paving the road towards what has been called the "New Age of Mechanisms" (Nosek et al., 2011, p. 152).

1.4 New directions in implicit (social) cognition and implicit measurement

Parallel to the development of "second generation" measures, aimed at solving the reliability problems of priming techniques and advancing on the structural problems of the first run of IAT measures, the field of implicit social cognition has been recently reshaped by two major theoretical and applicative advances.

The first major theoretical advance, which has contributed to wide spreading the conceptual and applied framework of implicit cognition to domains other than social psychology, involved the emergence of generalized, domain-independent dual-process models of cognition. The second major advance involved the use of implicit measures as a tool to explore the malleability and change of the to-be-measured attributes, extending the research on the ontological status of implicit psychological constructs and broadening the applicability of implicit measurement into an applied and experimental framework. Together, these advancements have had a significant impact on how researchers turned out to interpret the indirect measures scores and stimulated the shift of the questions focus on the underlying mechanisms of implicit measures as well as their ability to predict behaviour.

1.4.1 Dual-process theories of cognition

The idea that human behaviour is driven by more than one force or process has encouraged the development of several dual-process theories of human cognition, which depict behaviour as resulting by the interplay of separate mental processes.

The focal assumption of dual-process theories is that the mental processes underlying social phenomena can be divided into two distinct categories depending on whether they operate in an automatic or non-automatic fashion, a distinction somewhat (partially) resembling the status of implicit and explicit measures. Early dual-process theories emerged already in the 1980s and were mainly domain-specific, in the sense that their applications were circumscribed to specific content domains in social psychology, such as attitude-behaviour relations (Fazio, 1990) and prejudice and stereotype (Devine, 1989). The common feature of these domain-specific models of social cognition was the separation of social-cognitive processes into effortless, automatic processes versus effortful, controlled processes.

With the beginning of the new millennium, the focus of dual-process theorizing shifted toward the development of integrative theories that aimed at identifying general principles that are independent of particular content domains. A first, essential step towards a theoretical reunification was made by Smith and DeCoster ‘s influential review article (2000) where they conceptually integrated the various domain-specific theories into a single dual-process framework. The central argument in support of their account was that all the multiple dualisms advanced by domain-specific models do not reflect the contents or the occurring conditions of cognitive operations, but refer to the operations themselves of two basic processes and/or systems characterizing any kind of human cognition: associative versus rule-based processes. This distinction has called forth the development of generalized dual-process theories, including models that distinguish between reflective and impulsive processes (Strack & Deutch, 2004), System 1 versus System 2 processing

(Kahneman, 2003), and associative and propositional processes (Gawrosnki & Bodenhausen, 2006). Furthermore, the refinement of dual-process models set up the theoretical reinterpretation of direct and indirect measures as reflecting the outcomes of two qualitatively distinct processes (Rydell & McConnell, 2006) and suggested a conceptual framework for anticipating how and when implicit and explicit processes predict behaviour independently or interactively.

In common across the variety of generalized dual-process models of cognition is the description of qualitatively distinct processes or systems, one of which is "reflective, rule-based, propositional, systematic, deliberate, controlled, conscious, or explicit, and another that is impulsive, associative, heuristic, spontaneous, automatic, unconscious, or implicit" (Nosek et al., 2011, p. 156). These theories differ from the domain-specific dual-process accounts in that they are more concerned with the broader, basic architecture of information processing, to provide a general account of how human mind works (Gawronski, Sherman, & Trope, in press).

The most dominant and influential exemplar of *dual-system*[3] theory is the Impulsive-Reflective Model (IRM) by Strack and Deutsch (2004), which distinguishes between associative and propositional processes and representations, which are active simultaneously and operate interactively. The core principle of the IRM is that social cognition and behaviour are a function of an impulsive system and a reflective system.

The impulsive system is represented by a network of associations between concepts, events, and stimuli, which associative links differ in strength according to the frequency of combined occurrence in the network. The input information coming from the environment is always processed by the impulsive system – which resides outside the awareness and control of the individual – but the influence of the perceived stimuli is determined to a great extent by the pre-activation and weights of the connections in the part of the network in which the information is stored (Strack & Deutsch, 2004). When we encounter a stimulus triggering a pattern of strongly

---

[3] According to the mental processes assumed to explain phenomena, dual-process theories can be distinguished in *dual-process theories*, which emphasize functionally distinct mental processes, *dual-representation theories*, which relate different bahavioral outcomes to different mental representations, and *dual-system theories*, in which both processes and representations are elements of distinct processing systems that affect behavioural outcomes (Gawronski et al., in press).

associated concepts, related behavioural schemata are activated, which elicit a certain action towards the stimulus.

An example can help clarifying: when we see an advertisement of our favourite food, like a delicious apple-pie, certain associations activate in the impulsive system. The smell of the pie, its tempting aspect, the appetitive urgency to eat the cake, and also the feelings we have after eating it, may be strongly activated because of the great value the connections have. Associative processes involve then the activation of associations in memory, which is guided by the principles of similarity and spatio-temporal contiguity. Associations are bounded to *behavioural schemata*, which have been slowly and implicitly learnt in the past through repetition of certain action pattern in relation to a particular stimulus and have become part of those procedural memories, defined as 'habits', that serve the production of behaviour denoted by some degree of automaticity (Strack & Deutsch, 2004).

Let's continue with the example: once we see the apple pie advertisement, the *apple-pie* and connected elements activate. The connection weight increases, which implies that the associations are strengthened and the threshold for a future activation is lowered. If we go for some grocery shopping at the supermarket, we might encounter the bakery stand with some fresh-made, still warm apple-pies. This time the stimulus has a greater impact because of the previous pre-activation, which lowered the activation threshold, leading to a higher likelihood that we end buying a couple of apple-pies. The more we eat the cake, the more the effects become stronger and the more apple-pie-related associations activate when encountering an apple-pie-related stimulus.

Learning in the impulsive system is often paired with an affective component, a motivational orientation, which leads to positive or negative feelings and arousal preparing for action. Moreover, the impulsive system is fairly inflexible because the previously formed associations change only gradually and slowly. That implies that it operates in an automatic manner and once an association is activated it is hard to stop it and to change the course of action. The impulsive system generates prompt responses to stimuli without considering rightness or possible consequences of the action, for it cannot generate new action plans never done before.

The control and modulation of the associations and automatic schemas stored in the impulsive system is managed by its complementary, the reflective system, which is non-automatic and non stimulus-driven, and serves different regulatory goals and processes in relation to the encountered stimuli in light of the current circumstances and constraints. This system is driven by propositional processes conceptualized as the validation of the information implied by activated associations in the impulsive system, and dependent on syllogistic principles of logical consistency (Gawronski & Strack, 2004). The reflective system is busy in forming a meta-representation of what is activated in the impulsive system, to generate judgments, decisions, and intentions, which result in verbal and nonverbal behaviour. Taken together, all these operations are resources expensive.

The ability of the reflective system to build symbolic representations, based on intentions, processing capacities, and situational constraints, can let a certain stimulus have a meaning in the reflective system in opposition to that one held in the impulsive system, where it simply depends on previous related associations and on the organism need state (e.g., hunger, thirst, etc.). The reflective system can then generate explicit, propositional judgments and decisions as well as correcting judgments to make them more accurate or socially desirable. Although this ensures great flexibility, the system works slowly, depends on intentions, and can be easily disrupted by other processes. For instance, we may be in dietary and know that cakes aren't part of our planned diet, thus assigning an aversive, negative value to the apple-pie and explicitly state the unwillingness to buy and eat sweets. But, on the other hand, we have an implicit positive opinion about the apple-pie because of the numerous previous times we enjoy eating it, and give up to resist the temptation because in the impulsive system the apple-pie is associated with positive elements, such as the feeling of pleasure of eating it. Yet, we can feel dubious and stop in front of the bakery stand to think about the appropriateness of buying an apple-pie according to out current dietary plans. The reflective system influences behaviour through judgmental processes that result in a decision about the desirability and feasibility of a particular action: eating or not the cake while on dietary. The impulses suppression and conflict solution are among the regulatory operations of the reflective system.

Imagine we weren't in the grocery alone, but with our two children, who, excited by going to the supermarket, strive for sweets and repeatedly ask us to buy this and that, jumping from one stand to the other. Now, our cognitive resources are busy in dealing with our children aroused by the fun of going to the supermarket, and at the same time with remembering our shopping list. Once we find ourselves in front of the bakery stand, the time and pressure constraints do not leave room for reflection on whether or not buying the cake. It is indeed highly probable that we just take a couple of apple-pies while keeping under control the two kids and buying the bread, and surprisingly find them in the shopping bag once at home. The motivational, affective and arousal associations related to the apple-pie and the behavioural schemata learnt and automatically retrieved before we decided to start a diet, can disrupt reflection depending on motivation, intentions, deprivation and cognitive resources.

When using the IRM as a theoretical framework to understand implicit measures, the existing dissociation between implicit and explicit measures can be understood as reflecting the differential input from the two processing systems (Deutsch & Strack, 2006). Implicit measures have shown to be better off predicting highly automatic behaviours, such as facial expressions, gestures, or spontaneous approach and avoidance behaviours, whereas explicit evaluative judgments are better at predicting controlled behaviours. According to their conceptual design, implicit measures are to primarily assess associations and action tendencies pertaining to the impulsive system, whereas explicit measures mirror the propositional and judgmental processes in the reflective system. The latter will construe its judgments based on what is stored in the first one, and both types of measures will correspond. However, if actual goals and awareness of unwanted influences suggest a judgmental correction, a low correspondence between the two types of measures will occur (Deutsch & Strack, 2006). When behaviour occurs in situations where reflection is more likely, it will be determined by judgments and decisions, including potential corrective processes, and hence explicit measures will perform better. When the conditions favour more impulsive behaviour or the irruption of automatic behavioural schemata out of the balancing control of symbolic processing, implicit measures can be better predictors.

1.4.2 Implicit measures as a *medium* for change

In the recent years, the transition from the 'Age of Measurement' to the 'Age of Mechanisms' (Nosek et al., 2011) in implicit cognition has been occurring, and hasn't reached a stop yet. The development of implicit measures, such as the IAT and priming tasks, propelled the accumulation of evidence and theoretical insights for the value of implicit measurement in understanding human behaviour and in disentangling the different facets contributing to human behaviour itself.

Implicit cognition field is now transiting from the first stage of discovering and emergence of a corollary of paradigms and instruments for the assessment of what is difficult, for one reason or another, to reach through traditional measurement instruments, towards "second-generation" questions, such as: how do implicit cognitions form, change, and predict behaviour?, how do implicit and explicit processes interact?, what taxonomies of implicit or automatic content and processes are useful for theory and explanation?, what is the role played by implicit cognitions in generating behaviour?

Following the theoretical insights of the above-mentioned dual-process theories of implicit cognition, a new line of research has grown up recently. The explanation of the processing differences between implicit and explicit cognition stimulated the interest in the formation of implicit and explicit attitudes, and in general of mental associations (e.g., Gawrosnki & Sritharan, 2010), and, most of all, encouraged the scrutiny of the possibility of manipulating and changing implicit associations. This perspective challenged the early models emphasizing the steady and stable nature of automatically activated mental contents, which were considered to be slow to form, relatively insensitive to situational features, and slow to change (e.g., Schneider & Shiffrin, 1977; Smith & DeCoster, 2000). The practical application of implicit measures in domains beyond the boundaries of social psychology has then proven the flexibility and the value of implicit measures as crosscutting *media* to open a window on human behaviour and its antecedents.

A particularly productive domain of research on the malleability and change of implicit cognitions concerns research in psychopathology and experimental clinical psychology, in which the study of dysfunctional behaviours and maladaptive cognitive processes is of key importance for the human health and well-being and where the strive for the development of new ways to target negative behaviours and thoughts is continuously burning.

In the last ten years the concepts of implicit cognition has been implemented in the psychopathology framework in the endeavour of finding new routes toward a better understanding of the impulsive, implicit, largely unconscious and involuntary processes underlying harmful behaviours and psychopathological conditions.

The hypothesis on the basis of the use of implicit cognition concepts in clinical psychology research envisions the fact that many forms of psychological disorders are characterized by a lack of intentional control, by irrational features and by emotional deregulation. Implicit processes may be important in the aetiology and maintenance of psychological disorders (Wiers, Teachman, & De Houwer, 2007), as they may help to understand why people persist in producing dysfunctional behaviours, despite knowing that they should refrain from these action, such as continuing abusing drugs, or avoiding spiders or social situations.

As a consequence, implicit measures are thought to provide a tool for assessing specific underlying cognitive processes, such as attentional processes (e.g., *Visual-Probe Task* – MacLeod, Mathews, & Tata, 1986; *Emotional Stroop Task* – Williams, Mathews, & MacLeod, 1996), appraisals or interpretations of ambiguous situations or memory associations (e.g., IAT, GNAT, EAST, EP and SM).

To the extent that implicit measures reflect uncontrollable, unaware, fast mechanisms, they could provide important information that increases that from explicit measures. This is important in psychopathology research where self-presentation strategies and limited introspection into mental contents are often a concern. Further, by considering the various facets of implicitness of a measure (De Houwer, 2006; Moors & De Houwer, 2006), measures targeting attributes of which a person is unaware and/or based on rapid processing (such as those involving speeded response times) are more likely to capture automatic effects of psychological attributes and go beyond their explicit and conscious counterpart (Roefs et al., 2011). Therefore, implicit measures are assumed to reflect associations between disorder-relevant targets (e.g., the self in depression or alcohol in alcohol addiction) and particular focal attributes (e.g., negative or positive). As a result, they have the potential to reveal aspects of dysfunctional beliefs that explicit measures cannot reveal and to predict behaviours that explicit measures do not predict.

In 2011, a relevant, integrative review by Roefs et al. provided an extensive resume of the application of implicit measures in psychopathological research across 12 categories of disorder. The review of experimental, cross-sectional and incremental predictive validity studies unravelled mixed patterns of disorder-related associations that were partly disorder-congruent and partly unexpected, suggesting the need for revising the existing theories in some cases (e.g., the positive implicit self-esteem in depressive disorders and social-phobia, and negative associations with craved-substances in food, alcohol and drug-related disorders) and for deepening the experimental research in others (e.g., obsessive-compulsive disorder, pain disorder, panic disorder). Further, implicit measures explained variance in a range of behavioral measures in addition to that explained by explicit measures (e.g., panic symptoms, mirror avoidance, self-reported alcohol uses, food choice), pointing to their complementary value. Noteworthy is the pattern of experimental results, which often found the expected effect of the manipulation on the implicit measures used, and which showed to be consistent with the view that the processes indexed by implicit measures play a role in the targeted disorder, pointing to the predictive validity of implicit measures as add-on tools in psychopathology research.

The connubial of implicit cognition, implicit learning theory and experimental clinical psychology made a step further. Given the discovery on the malleability and changing potentials of implicit mental associations, researchers enquired about the possibility to extend this issue also to implicit processes involved in the development and maintenance of psychological disorders.

At the very beginning, the interest was mostly directed towards the causal role played by implicit mental associations and cognitive biases, by directly altering them using computerized training procedures, to reveal the consequences for clinically relevant symptoms. These computerized training procedures, the Cognitive Bias Modification (CBM), were developed by adapting the original assessment implicit measures to the suits of re-training and manipulation techniques of the to-be-measured constructs. The evidence about the efficacy of CBM paradigms on influencing symptoms severity has immediately led to an enormous growth of interest in the potential therapeutic value of CBM. The literature on the design and use of CBM in emotional disorders and addiction disorders has increased fourfold in the past five years (MacLeod, 2012).

A recent review and commentary on CBM by Colin MacLeod stated, "Cognitive bias modification (CBM) techniques have proven capable of systematically training change in the patterns of selective attention and selective interpretation known to characterize various forms of psychopathology" (2012, p.115). The author also called for the necessity of "[…] large-scale randomized controlled trials, to compare the efficacy of CBM with that of alternative approaches, and to identify how best to integrate CBM techniques into multimodal treatment packages" (2012, p.115).

The application of CBM paradigms, which include the training of attentional bias, approach bias, evaluative bias, and interpretative bias in the information processing of environmental inputs, has expanded to experimental interventions in pain disorders, anxiety, depression, addiction, eating disorders, and dysmorphophobia; obtaining promising results (MacLeod, 2012). The stage of proof-of-concept has been positively passed.

1.5 Conclusion

In the last ten years, implicit cognitive processes, measured with a variety of different strategies, have garnered a great deal of support in basic cognitive research across multiple areas of psychology, including social psychology and experimental clinical psychology. Further, these processes consistently have been found to predict or correlate with a variety of behaviors, above and beyond more explicit processes. Implicit processes can be differently conceptualized and operationalized as implicit memory associations, attentional biases, approach-avoidance action tendencies, or ambiguous information interpretation bias; and can measured with a wide variety of measurement procedures. So far, implicit measures in interventions has proven helpful in increasing the understanding of the non-reflective, associative, impulsive and 'automatic' side of intervention effects and, last but not least, showed to be flexible enough to also provide a framework for intervention strategy and design in mental health.

PART 1

IMPLICIT MEASURES IN THE ASSESSMENT OF MENTAL ILLNESS STIGMA

Chapter 2

Implicit measurement of mental illness stigma: causal beliefs and attitudes.

Development of two Implicit Association Tests

2.1 Mental illness stigma: an introduction

The stigma associated with mental illness is a burden on mentally ill people and a relevant clinical and public health issue that can worsen the course of a mental disorder. People with mental illness are often confronted with a double problem. First, they suffer from a wide range of negative effects and impairments related to the disorder itself and have to cope with its symptoms, which can make it difficult for someone with a mental illness to work, live independently, or achieve a satisfactory quality of life. Second, mentally ill people have to deal with the misunderstandings of society about mental illness and the social stigmatization of their illness. For instance, those who manage their problems well enough and are able to maintain a job are often confronted with difficulties in finding or keeping a job position because their colleagues and/or bosses do discriminate them.

Being diagnosed with a mental illness results thus not only in troubles arising from the disorder itself but also in overt disadvantages due to the society's reactions. The effects of stigma add to those emanating from the mental illness itself, with deleterious consequences for the individual (for a review see, Hinshaw, 2007).

People diagnosed with a mental illness often face public discrimination (*public stigma*) resulting in social, educational, vocational, and even health care setbacks. As a further complication, some people with mental illness may accept the common prejudices and stereotypes about the mental disorder, internalise these negative views and apply them to themselves (*self-stigma*), resulting in lowered self-confidence and negative outcomes, such as low quality of life, limited employment opportunities and occupational withdraw, avoidance of and failing in help-seeking actions, and treatment discontinuation (e.g., Corrigan, 2004; Corrigan, Tsang, Shi, Lam, & Larson, 2010; Hinshaw, 2006; Link, Struening, Neese-Todd, Asmussen, & Phelan, 2001; Livingston & Boyd, 2011). Moreover, most findings reveal mental health professionals' attitudes to be comparable to those found in the general population, which means that mental health professionals themselves are a significant source of public stigma. That is potentially one of the reasons why people with mental illness are discouraged from seeking help (Schulze, 2007), with negative relevant consequences for a responsive care planning and patients' treatment intervention.

2.1.1 What is stigma?

In 2013 occurred the 50th anniversary of Erving Goffman's seminal work *Stigma: Notes on the management of spoiled identity*, which constituted a milestone in stigma conceptualization, understanding, and related research not only in the sociological realm – where it was generated – but also in medicine, health sciences, and psychology. Inspired by the pioneering work of Goffman (1963), the 2001 World Health Report defines stigma as "a mark of shame, disgrace or disapproval which results in an individual being rejected, discriminated against, and excluded from participating in a number of different areas of society" (WHO, 2001).

The term *stigma* originates from ancient Greek to denote a physical mark applied to social outcasts (e.g., criminals, slaves, etc.) to indicate socially devaluated people. The term stigma has since been used to describe an individual attribute associated with undesirable characteristics and is defined and enacted through social interaction (Goffman, 1963). "Stigma is typically a social process, experienced or anticipated, characterized by exclusion, rejection, blame or devaluation that results from experience or reasonable anticipation of an adverse social judgment about a person or group" (Weiss & Ramakrishna, 2004, p. 536). The definition of stigma has generated a long-lasting debate about its featuring components (for a discussion see Feldman & Crandall, 2004; Hinshaw, 2007; Link & Phelan, 2001). Currently, the perhaps most thorough definition of stigma describes it as a pervasive and global

> "devaluation of certain individuals on the basis of some characteristic they possess, related to membership in a group that is disfavoured, devalued, or disgraced by the general society" (Hinshaw, 2007, p. 23).

Stigma refers then to forms of social rejection and is therefore a social construct. Certain negative attributes may be particularly salient within a social community (Link, 2001), because deviant from the shared social norms and values, and the link between labels and undesirable attributes can be relatively strong or weak. It follows that certain conditions, such as mental illness, may be more socially stereotyped compared to medical conditions such as diabetes (Link, Yang, Phelan, & Collins, 2004). The phenomenon of stigma may more or less prominent depending on the social and cultural context (Abdullah & Brown, 2011).

Applied to mental illness, stigma then refers to the social judgment, degradation, or devaluation of individuals because they have mental illness symptoms or have been labelled as having a mental illness.

From a social-cognitive perspective, stigma involves a triad of cognitive, emotional, and behavioural components: (a) stereotypes, which are knowledge structures known to the majority of a social group and aimed at quickly and efficiently categorizing people and creating expectations about them, by working as 'umbrella' concepts; (b) prejudice, which involves the agreement and endorsement of negative stereotypes followed by a negative emotional reaction towards the stereotyped group; and (c) discrimination, which is the behavioural consequence of prejudice and entails differential treatment of one group respect to another, for instance by curtailing rights and life opportunities of those who are attached the 'mark of shame' or favouring those who do not have mental illnesses and are not part of the stigmatized group (Corrigan 2004, 2007; Hinshaw & Stier, 2008; Stier & Hinshaw, 2007).

The stigmatization process starts with some cues indicating that a person may have a mental illness: psychiatric symptoms, lack of social skills, unusual physical appearance, and labels are the four major cues the public uses as an indication of mental illness (Corrigan, 2004). These cues are likely to activate negative stereotypes that can lead to prejudice, which can bring about discriminating behaviours (Corrigan, 2007).

This social–cognitive approach can be applied to both *public stigma* and *self-stigma.* For both types of stigma, the first two parts of the process (cues and stereotypes) are the same, since they arise from general socialization processes in the public. When related to *public stigma*, prejudice involves a person without a mental illness endorsing a stereotype towards those who do have mental illnesses (e.g., believing that people with mental illnesses are dangerous), while *self-stigma* involves a person with a mental illness internalizing stereotypes about mental illness and believing in those stereotypes as they apply to herself (e.g., believing "I am dangerous because I have a mental illness"). As applied to *public stigma*, discrimination occurs when a person or social policy system devalues people with mental illnesses and, because of prejudicse toward them, engages in unfair behaviours that negatively impact them (e.g., a care provider believes that people with mental illnesses are dangerous and stick to unmotivated reclusion or coerced hospitalization because a patient has a mental illness). Within *self-stigma*, discrimination occurs when people devalues themselves for having a mental illness and engage in detrimental or self-damaging behaviours (e.g., not going out because I'm dangerous because of my mental illness and I'm afraid of hurting other people) (Corrigan, 2007).

The two typologies of stigma points to the diverse environments where stigma can occur, namely, social, interpersonal, and personal environments. Besides *public stigma* and *self-stigma*, a recent model summarized previous stigma conceptualizations by identifying other two interrelated manifestations of stigma: *stigma by association*, which, likewise Goffman's (1963) courtesy stigma, entails social and psychological reactions to people affiliated with a stigmatized person (e.g., family and friends) as well as people's reactions to being associated with a stigmatized person; and *structural stigma*, which mirrors a stigmatized status at the organizational, institutional, and ideological levels of society (Pryor & Reeder, 2011). The four manifestations are interrelated one to each other; however, *public stigma* is at the core of the other three forms, given the breadth of public consensus about a devalued social attribute, and roots the formation and inception of the other stigma manifestations (Bos, Pryor, Reeder, & Stutterheim, 2013).

2.1.2 Antecedents of mental illness stigma: aetiological beliefs

The stigma introduction given so far presented public stigma as the source of all forms of stigmatizing manifestations towards a *different* social group, to be considered as deviant from the common shared social, moral and, above all, cultural norms (Abdullah & Brown, 2011). But what are the prodromes of a stigmatizing attitude and/or behaviour? What is the underlying structure of mental illness stigma? What leads the condition of being diagnosed with a mental illness to cause social rejection?

In 1984, a general model of stigma components introduced six dimensions that could help explain and characterise what is stigmatising (Jones, Farina, Hastdorf, Markus, Miller, & Scott, 1984): *concealability* (can it be kept secret?), *course* (is it stable?), *disruptiveness* (does it hurt relationships?), *origin* (what caused it?), *aesthetics* (is it unpleasant to the sense?), and *peril* (is it dangerous?). This list of dimensions is not exhaustive of the phenomena, since additional dimensions may be relevant. Further, the six dimensions can be applied with different degrees of significance to the different recipients of stigma. However, for the purpose of the present work, some of these dimensions can be highly valuable in predicting what is stigmatizing about mental illness. As a matter of fact, Feldman and Crandall (2007) empirically tested the theoretical status of Jones et al. (1984)'s stigma dimensions. Results of their study indicated that people generally desire more social distance when the mental disorder is perceived to be the individual's own fault (*origin*), when the mental illness is perceived to cause the individual to be dangerous to others (*peril*), and when the mental illness is perceived as uncommon or rare. These three predictors accounted for about 60% of the variance in stigma.

Peril points to the fact that the more people believe that mental illness is associated with dangerous or aggressive behaviour; the more they discriminate (Feldman & Crandall, 2007). Perceived dangerousness and unpredictability elicits fear and avoidance.

The dimension *origin* is strictly bounded to *peril* and refers to the causes of the condition, with biologic and genetic explanatory accounts on one side and psychological and environmental explanations on the other side (hereinafter the terms biogenetic and psychosocial are used). This dimension has been hypothesized to be a kind of antecedent of negative stereotypes and has been used as a promotional medium to overcome stigma in a number of public health programs aimed at combat discrimination (e.g., National Alliance for Mental Illness, 2008, 2009). These campaigns have been emphasizing biogenetic causal models of mental disorders by sponsoring a "mental illness is an illness like any other" account and explicitly portraying mental disorders as medical conditions (e.g., Corrigan, 2000;

Read, Haslam, Sayce, & Davies, 2006; Schomerus et al., 2012), which are to be treated with medical treatments.

The promotion of biogenetic aetiological beliefs about mental illness has been deemed as a promising approach to reduce stigma, for they are connected to the perception of *onset* and *offset controllability* for the stigmatized condition (see Weiner, 1995).[4] The hope for medicalization of mental illness to alleviate stigma manifestations rests on the assumption that endorsing biogenetic causes of mental illness may reduce ascriptions of responsibility and guilt to the affected person, since such causes are beyond the individual control, and may reverse the perception that people with a mental disorder is to blame for their troubles, with less rejection in the social environment. High levels of attributed personal responsibility for the onset of the deviant condition evoke blame and stigmatizing behaviour, as would be the case with a smoker who gets lung cancer, whereas low levels of personal responsibility yield feelings of sympathy and greater tendencies to provide help, which would be more likely with a woman who receives a diagnosis of breast cancer.

Unfortunately, a large number of studies in the last 15 years consistently evidenced biogenetic explanations to be associated with higher levels of stigma, mostly in terms of perceived dangerousness and uncontrollability (e.g., Angermeyer, Holzinger, Carta, & Schomerus, 2011; Angermeyer & Matschinger, 2005; Kvaale, Gottdiener, & Haslam, 2013; Kvaale, Haslam, & Gottdiener, 2013; Lee et al., in press; Mehta & Farina, 1997; Read & Harré, 2001). This was not found for psychosocial causal beliefs (Read & Harré, 2001), which appear to have remained stable during the last 20 years (Schomerus et al., 2012). On the other hand, the medical approach to mental illness seemed useful to enhance the endorsement of professional medical treatment and promote an increase in mental illness literacy, but not on the detriment of psychiatric and psychological interventions, such as psychotherapy (Angermeyer et al., 2011; Schomerus et al., 2012).

Although there is a growing body of research suggesting that the belief that mental illnesses have biogenetic causes is associated with greater stigma, even now the predominant view is that mental illness stigma is less likely when people

---

[4] The link between aetiological beliefs and disorder onset controllability is explained within the framework of attribution theory, which holds that causal attribution of one' behaviours lead to characteristic emotional, attitudinal, and behavioural responses towards the person in question (e.g., Weiner, 1995).

perceive the illness as outside of the affected person's control (Feldman & Crandall, 2007; Hinshaw, 2006). This view actually contrasts with the research results. A recent review on the dissemination of mental illness biogenetic causes evidenced some critics which should caution the popularization of such explanations insofar as they may have unintended side effects that could exacerbate stigma (e.g., Haslam, 2011). More precisely, it has been argued that casting mental disorders as biogenetically caused diseases may induce an "essentialist thinking", which involves that members of a stereotyped group share a fixed, crystalized and unchangeable negative essence and are set apart from other social groups by appealing to a fundamentally 'neurobiological otherness' (Kvaale, Gottdiener et al., 2013; Kvaale, Haslam et al., 2013). It follows that unpredictability and danger stereotypes, as a result of biogenetic models, are put at the core of psychological problems features. Essentialist thinking can then have a gradual harmful effect, for it increases the endorsement of these stereotypes and the belief that they are intrinsic to the mentally ill people (e.g., Boysen, 2011; Kvaale, Haslam et al., 2013). For instance, it emerged that biogenetic causal beliefs about mental illness are linked to a higher perception of dangerousness and prognostic pessimism – according to psychological essentialism theories (e.g., Haslam, 2011) – but to lower blaming responses towards affected people (Kvaale, Gottdiener et al., 2013; Kvaale, Haslam et al., 2013) – according to attribution theory (e.g., Weiner, 1995). Surprisingly, a null effect was found for social distance (Kvaale, Haslam et al., 2013). Beyond the different theoretical accounts of stigma components, this heterogeneity of findings points to the different operationalization of mental illness stigma (e.g., social distance, dangerousness perception, blame, etc.) and to the myriad of stigma measures, which mainly rely on the use of explicit measures, calling for a standardization of measurement and for an exploration of instruments targeting more subtle processes in mental illness stigmatization.

2.2 Implicit measurement of mental illness evaluative and semantic automatic associations

A growing body of research evidenced the risk for biases due to individual's self-presentation and/or limited introspective ability, when assessing stigma (e.g., Stier & Hinshaw, 2007). Certain attitudes are discriminatory, such as those towards mentally ill people, and can increase the likelihood that social desirability concerns impact on self-reported views. According to a dual-process account of explicit and implicit processes (see Chapter 1), implicit attitudes may be held regardless of whether an individual believes these to be true or false (Gawronski & Bodenhausen, 2006). Explicit processes are characterised by evaluative reasoning that assesses the validity of automatic associations stored in memory associative networks, determining whether these are true or false. Therefore, divergent scores on implicit and explicit attitude measures may result from an individual's assessment of the validity of automatic attitudes and their rejection if they are deemed to be inappropriate or wrong (Sritharan & Gawronski, 2010). Stigmatization processes involves then both automatic implicit responses as well as controlled deliberate responses (Pryor et al., 2004).

To date, work on mental illness stigma has relied primarily on self-report measures (for a review of explicit measures of mental illness stigma facets, see Brohan, Slade, Clement, & Thornicroft, 2010; Link, Yang, Phelan, & Collins, 2004), but there has been a recent growing interest in the investigation into the role of implicit processes in the expression of bias towards mentally ill people. The evaluations that people arrive at after thoughtful deliberation may diverge from their initial, immediate evaluative impulses (Gawronski & Bodenhausen, 2006; Greenwald & Nosek, 2009), underlining the necessity of assessing implicit as well as explicit processes. Indirect measures such as *the Implicit Association Test* (IAT – Greenwald, Schwartz, & Banaji, 1998) and its derivates, can measure automatic evaluative and semantic associations between two concepts and thus index implicit attitudes and stereotypes towards mental disorders. Recently, a bunch of studies have examined the value of including indirect assessments of implicit stigmatizing attitudes and stereotypes and the role of stigma dual processes in both healthy and diagnosed samples, providing promising results about the differential functioning of negative attitudes when measured at the two levels (for a summary of studies, see Table 2.1).

Teachman, Wilson, and Komarovskaya (2006) showed that the general public, and even those diagnosed with a mental disease, were both implicitly and explicitly biased against other mentally ill people, compared to physically ill people (see also Rüsch, Corrigan, Todd, & Bodenhausen, 2011; Rüsch, Todd, Bodenhausen, & Corrigan, 2010a,b). Peris, Teachman, and Nosek (2008) also demonstrated the value of implicit stigma assessments, finding that people who had received mental health training reported more positive implicit and explicit evaluations regarding mental illness. Nonetheless, more negative implicit attitudes toward mentally ill people predicted more overdiagnosis of clinical case vignettes than explicit attitudes. Interestingly, an experimental study on the possibility of reducing stigma towards mental illness evidenced a significant decrease of explicit negative bias towards schizophrenia (compared to depression) after an educational intervention. However, this did not happen for implicit negative attitudes (Lincoln, Arens, Berger, & Reif, 2008). A related study found that completing an indirect measure of implicit stigma prior to an explicit measure induced less explicit negative bias towards mentally ill people compared to the reversed order (Menatti, Smyth, Nosek, & Teachman, 2012).

Automatic self-attributions of guilt/blame resulted to be related to lower quality of life (Rüsch, Corrigan, Todd, & Bodenhausen, 2010), perceived discrimination legitimacy (Rüsch, Todd, Bodenhausen, Olschewski, & Corrigan, 2010), and self-punishment attitudes ("I get what I deserve") (Rüsch et al., 2010b) among mentally ill people, pointing to the potential adverse outcomes of internalized stigma (i.e., self-stigma) for treatment seeking and mental well-being.

Only one study addressed the indirect assessment of implicit attitudes, stereotypes (i.e., stability and controllability), and causal beliefs about mental illness – specifically about depression – compared to physical illness among psychology undergrads (Monteith & Pettith, 2011). Depression was implicitly more negatively evaluated. Implicit associations regarding its temporary nature and underlying psychological causes were found, in comparison to physical illness. No difference emerged for controllability associations. The effect of aetiological beliefs on mental illness negative evaluations emerged in another study, where explicit mental illness biogenetic causal beliefs were associated to greater implicit self-guilt and explicit fear of mental illness amongst clinically diagnosed individuals (Rüsch et al., 2010a).

The literature sketch provided so far encourages the experimentation of indirect measures for the assessment of more covertly expressed features of stigma, which can open an additional window on stigma, by focusing on more automatic facets of prejudiced and discriminatory attitudes towards mental illness.

Table 2.1 Summary of studies exploring implicit mental illness evaluative and semantic associations.

Reference	Sample	Design	Implicit Measure	Results
Teachman et al., 2006	Study 1: undergraduates (N=119). Study 2: clinical (N=35) and healthy control (N=36) groups.	Cross-sectional; between-groups	Three IATs: mental illness vs physical illness on badness, helplessness, and blameworthiness (replaced by me/not me in study 2).	Implicit negative bias against mental illness in all IATs and across samples, except for the self attribution IAT (no effect and difference)
Lincoln et al., 2008	Medicine (N=60) and psychology (N=61) undergraduates.	Experimental; Within-subjects	Three IATs: schizophrenia vs depression on responsibility, treatability, and dangerousness.	Strong implicit negative bias against schizophrenia in t1 across groups. After educational intervention, change in explicit attitudes but not in implicit stereotypes.
Peris et al., 2008	Mental health professional (N=682), undergraduates (N=204), general public working with clinical populations (N=112), other health/social services (N=541).	Cross-sectional; between-groups	Implicit attitude IAT: mentally ill people vs welfare recipients.	Higher implicit and explicit positive bias for mental health professionals. Explicit bias predicted clinical prognosis and implicit bias predicted clinical overdiagnosis.
Rüsch, Corrigan et al., 2010	Clinical group (N=85).	Cross-sectional; correlational	Two Brief IATs: mental illness vs physical disability and me/not me (self-esteem) on valence. The BIAT scores product was taken as an index of implicit self-stigma.	Lower implicit self-stigma uniquely predicted higher quality of life after controlling for diagnosis, depressive symptoms, and demographics.

Table 2.1. *(continued).*

Reference	Sample	Design	Implicit Measure	Results
Rüsch, Todd, Bodenhausen, & Corrigan, 2010a	Clinical (N=85) and general public control (N=50) groups.	Cross-sectional; between-groups	Two Brief IATs: mental illness vs physical disability and me and not me on guilty.	Mental illness biogenetic causal beliefs were associated to greater implicit self guilt and explicit fear amongst the clinical group
Rüsch, Todd, Bodenhausen, & Corrigan, 2010b	Clinical (N=85) and general public control (N=50) groups.	Cross-sectional; between-groups	Two Brief IATs: mental illness vs physical disability and me and not me on guilty.	Stronger "just world" beliefs ("I get what I deserve") were related to implicit self-guilt amongst the clinical group. Stronger Protestant ethic was associated to stronger guilt bias against mental illness amongst the general public.
Rüsch, Todd, Bodenhausen, Olschewski et al., 2010	Clinical group (N=75).	Longitudinal	Brief IAT: mental illness vs physical disability on shame (shameful/proud).	Stronger shame automatic associations with mental illness at baseline predicted higher perceived legitimacy for discrimination at 6-month follow-up, after controlling for diagnosis and comorbidity.
Monteith & Pettith, 2011	Undergraduates (N=162).	Cross-sectional; correlational	Four IATs: depression vs physical illness on valence, stability, controllability, and aetiology.	Stronger negative implicit attitudes towards depression; stronger explicit and implicit psychological accounts and implicit less stability of depression. No differences for controllability.
Rüsch et al., 2011	Clinical (N=85) and general public control (N=50) groups.	Cross-sectional: between-groups	Semantic Priming (Lexical Decision Task) for negative stereotypes against mental illness ("crazy" and "sane" as primes).	Less negative automatic stereotyping against mental illness in the clinical group. Among the general public, stronger negative automatic stereotyping was related to more shame for a hypothetical mental illness and more negative reactions towards mentally ill people.

Table 2.1. *(continued).*

Reference	Sample	Design	Implicit Measure	Results
Menatti et al., 2012	Study 1: general public (N=610). Study 2: general public (N=1161).	Experimental; between-groups	Brief IAT: mentally ill people vs physically ill people on valence.	Efficacy of an implicit awareness intervention (performing an implicit measure about mental illness attitudes before completing an explicit measure of stigma, independently of feedback on implicit stigma). The stronger the implicit negative bias, the lower the explicit negative bias.
Wang et al., 2012	Chinese undergraduates (N=56).	Cross-sectional; correlational	Three Single-Target IATs: mental illness on positive and negative words for cognitive evaluation, affect, and behavioural reactions.	Women presented overall (the three ST-IATs were combined into one general ST-IAT) more negative associations to mental illness.
Brener et al., 2013	Social services and care provision for mentally ill people (N=74).	Cross-sectional; correlational	Single Category-IAT: mental illness on valence.	Positive correlations between explicit and implicit attitudes towards mental illness. Positive implicit attitudes uniquely predicted helping behaviour and marginally predicted positive emotions towards mentally ill people.

The design and use of indirect measurement procedures is then prone to the adoption of a multi-method approach for the analysis of current levels of mental illness stigma, with considerable implications for the development and testing of interventions aimed at reducing such stigma.

First, an accurate assessment of pre-existing levels of stigma provides a standard of comparison between pre- and post-intervention levels of stigma. Second, the limited introspection and conscious awareness of the processes underlying stigmatization of individuals with mental illness can make people unable or unwilling to report their reasons for doing so. For their lowered susceptibility to social desirability factors, implicit measures can then spell out potential mechanisms that underlie such stigma. This information can thus be used to design more effective and efficient interventions and for an accurate assessment of the efficacy of such interventions (Stier & Hinshaw, 2007).

A final remark goes to the measurement status of implicit measures of mental illness stigma, which hasn't been covered yet. Given the affective, cognitive, and behavioural components of stigma, the degree of automaticity of the to-be-measured facet should be explicitly clarified (e.g., Moors & De Houwer, 2006), when measuring *implicit* stigma. By addressing the methodological investigation of the newly devised measures for implicit stigma, such as the IAT for mental illness semantic associations/stereotypes, a researcher should come to know what actually (s)he is measuring and how the measurement outcome is produced by the to-be-measured construct (see Chapter 1).

2.3 The present study: objectives

The present study focuses on the psychometric investigation of two implicit measures of mental illness stigma, which were administered within a broader research project on explicit and implicit mental illness stigma (e.g., Boffo & Mannarini, 2012; Mannarini & Boffo, 2013a,b). The two implicit measures of mental illness-related automatic semantic and evaluative associations were designed to be an index of people's automatic associations of mental illness to psychosocial or biogenetic attributions and of implicit evaluations of mental illness, compared to physical illness. The IAT was the elected indirect measurement procedure for two reasons: 1) the relative nature of the mental illness concept with respect to the physical illness, both in terms of salience and valence, prompted the use of a relative measure of implicit associations towards the target concept (e.g., Teachman et al., 2006); 2) among the corollary of indirect measures of implicit cognition, the IAT is the most widely used and tested, and one of the most reliable (see Table 1.1 in Chapter 1), suggesting it as a starting point for the study of the measurement validity of two implicit measures of mental illness stigma.

The choice of developing an implicit measure for causal beliefs and attitudes towards mental illness followed the literature pattern of mixed results on the relationship between mental illness causal loci and negative evaluations (mostly in terms of social distance), with some evidence supporting a direct connection between the two of them, some evidence supporting a more indirect connection, and some neglecting it. Hence, it seemed reasonable to start focusing on causal beliefs and attitudes as the two endpoints of the implicit stigmatizing process path. The first step was then to design an indirect measure of implicit causal attributions (psychological versus biologic) of mental illness and an indirect measure of implicit attitudes towards mental illness. In both measures, mental illness was compared physical illness, which is akin to entail negative evaluations as well (e.g., Monteith & Pettith, 2011; Rusch et al., 2010a,b; Rusch, Corrigan et al., 2010; Teachman et al., 2006).

The measurement validity of the two IATs has been then tested via the application of a latent trait modelling approach (see Chapter 3) to determine whether the to-be-measured attribute does exist, how it affects the measure, and to check for the effect of any confounding individual variable that can alter the measure, such as prior contact with mentally ill people and/or personal experience with mental illness/psychological problems.

The main aim was then to establish the possibility of actually measuring mental illness implicit causal beliefs and attitudes with an IAT paradigm.

2.4 Methods

2.4.1 Participants and procedure

The study involved 360 undergraduate students of the University of Padova (mean age = 23.82, *SD* = 3.146) who freely participated in the study. The research did not required the approval of the institutional Ethics Review Board, for it was conducted during research methods classes and was part of class demonstrations. Participation in the study conformed to the Declaration of Helsinki norms. Participants' informed consent request for confidential data treatment, free participation in and eventual withdrawal from the study, was part of the questionnaire package.

The 83.8% of the participants was female, the 72.7% did not have a job, the 58.9% was religious, and the 91.9% had some direct previous experience with mentally ill people. Of this, the 18.13% reported the presence of mentally ill people in the family (e.g., parent, sibling), 52.87% reported the presence of friends, relatives, or partner with psychological problems, and 29% had a contact with mentally ill people not in the personal circle (e.g., patients, neighbourhood, colleagues). The 33% of participants referred to have requested professional psychological help for themselves.

Participants were approached during a regular class and asked to take part in a research on "What do you think about mental illness?". They completed a first battery of questionnaires and clinical case vignettes and arranged an appointment for an experimental session in the lab, to complete the second part of the research. [5]

In the lab, participants completed the two indirect measures of implicit beliefs and attitudes towards mental illness and were briefly debriefed about the study objectives.

2.4.2 Implicit measures of mental illness stigma

The two IATs were administered on a 15-inch personal computer in a controlled setting (quiet lab room, no distractors) with Inquisit 2.0 software (http://www.millisecond.com). Each task block was preceded by a short instruction

---

[5] The explicit measures included in the study are not reported here because they are not part of the present study's main objectives. These included various demographic variables, the *Mental Disorders Causal Beliefs* and *Mental Disorders Therapeutic Relationship* questionnaires (Mannarini & Boffo, 2013a,b), and an unlabeled vignette describing a person affected by a particular mental disorder, about which respondents had to evaluate causal beliefs, recommend treatment, social distance, and perceived dangerousness. This part of the protocol was part of a separate, broader study and it didn't impact upon the mental illness related implicit measures, since these were administered at least two weeks later the first explicit measurement session.

of the following task, reminding the exact key assignment and to be ready to correct in case of mistake (a red cross appearing in the centre of the screen).

The presentation order of the critical blocks for combined double categorization for each task (compatible and incompatible blocks) was counterbalanced across participants (e.g., Greenwald et al., 1998; Nosek, Greenwald, & Banaji, 2005). The two IATs were administered in fixed order to all of participants (causal beliefs IAT first). It was hypothesized that performing the causal beliefs IAT first would have activated the linked memory evaluative associations towards mental illness, following the automatic associations network account theorized by dual-process models of cognitions (e.g., Strack & Deutsch, 2004; Gawronski & Bodenhausen, 2006). Therefore, the two IAT were expected to be positively correlated to each other.

2.4.2.1 Mental illness causal beliefs Implicit Association Test

The IAT measures association strengths between pairs of target concepts and an attributive dimension and consists of a stimuli categorization task into super-ordinate categories. It is a relative measure, so *Mental Illness* target category was compared with semantic associations of *Physical Illness* on the bipolar dimension of psychosocial (*Psychological*) versus biogenetic (*Biologic*) causal explanations of mental illness. *Mental Illness* was contrasted to *Physical Illness* for two main reasons: first, mental illness is a negative concept given that it reflects illness, so physical illness seemed the most obvious comparison term for its salience, which nonetheless is not as stigmatized as the mental illness. A second reason relied on the "mental illness is an illness like any other" (Read et al., 2006) approach, which has been promoting the conception of mental disorders as medical conditions to be treated with medical interventions, leading to the consideration of physical illness as the most effective contrasting category.

As previously described (see Chapter 1), the logic behind the IAT is that stimuli are classified more quickly during one critical block, when the target and attribute category pairing (e.g., *Mental illness-Psychological*) matches respondents' automatic associations between the two concepts, versus the other block, where the target and attribute category pairing is mismatched (e.g., *Mental illness-Biologic*) (see Table 2.2). Therefore, an individual who presents a stronger automatic association of psychosocial features and elements with mental illness, is expected to respond more quickly when *Mental Illness* and *Psychological* categories are paired and contrasted to *Physical Illness* and *Biologic* pairing (compatible block), when compared to the reversed pairing (incompatible block) (for an example of trial, see *Figure 2.1*).

The selection of stimuli for the two target categories was carried out with the advise of two clinicians, following criteria of prevalence and representativeness of diagnostic categories (DSM-IV-TR – APA, 2000) and familiarity ratings in the Italian daily language for both mental and physical illness stimuli (CoLFIS – Bertinetto et al., 2005). For the *Mental Disease* category the five selected word stimuli were the following: *depression* (depression), *schizophrenia* (schizophrenia), *psychopathy* (psicopatia), *paranoia* (paranoia), and *hysteria* (isteria). For the *Physical Disease* category the five words were *tumour* (tumore), *heart attack* (infarto), *pneumonia* (polmonite), *flu* (influenza), and *diabetes* (diabete).

Figure 2.1 Example of a compatible trial (image on top) and incompatible trial (image on the bottom) in the mental illness causal beliefs Implicit Association Test.

For the bipolar causal beliefs semantic dimension, a set of pictorial stimuli was created (see Appendix A). The selection of pictures instead of words complied with the need of more effective exemplars to make the sorting task easier and more efficient, due to the difficulty in conveying aspects of biological and psychological semantic areas by means of words without ambiguity or misinterpretation. Similarity in clearness and pictorial features, low degree of ambiguity, and easiness of categorization guided the stimuli selection. For the *Psychological* category, the six pictures depicted several interpersonal relationships and environments: a *mother/child relation (psico1)*, *grandparents/grandchildren relation (psico2)*, a *work meeting (psico3)*, *two friends arguing (psico4)*, a *romantic couple relation (psico5)*, and a *family relation (psico6)*. For the *Biologic* category, the six images pictures depicted several objects pertaining to the area of natural sciences, biology, genetics, and chemistry: the image of a *cell* under the microscope (*bio1*), a *filament of DNA (bio2)*, a coloured image of an *atom structure (bio3)*, *test tubes* for clinical analysis (*bio4*), a *microscope (bio5)*, and two *chromosomes (bio6)*.

To enable the visual understanding and subsequent categorization of the pictures, an instruction screen at the beginning of the IAT informed the participants that *Psychological* referred to the relations between individuals and the environment, whereas *Biologic* referred to anything related to the organic and biological aspects of life.

2.4.2.2 Mental illness attitude Implicit Association Test

A second IAT paradigm was developed to measure evaluative associations with *Mental Illness* compared to the *Physical Illness*. The two critical sorting conditions in the IAT reflect negative (*Mental illness-Negative*) and positive (*Mental illness-Positive*) automatic evaluations of mental illness when compared to the contrasting category of *Physical Illness* (see Table 2.2). Given the more frequent and stronger stigmatizing attitudes and behaviours towards mentally ill people, the pairing *Mental Illness-Negative* contrasted to *Physical Illness-Positive* was hypothesized to be participant' associations-congruent (compatible block; for an example of trial, see *Figure 2.2*).

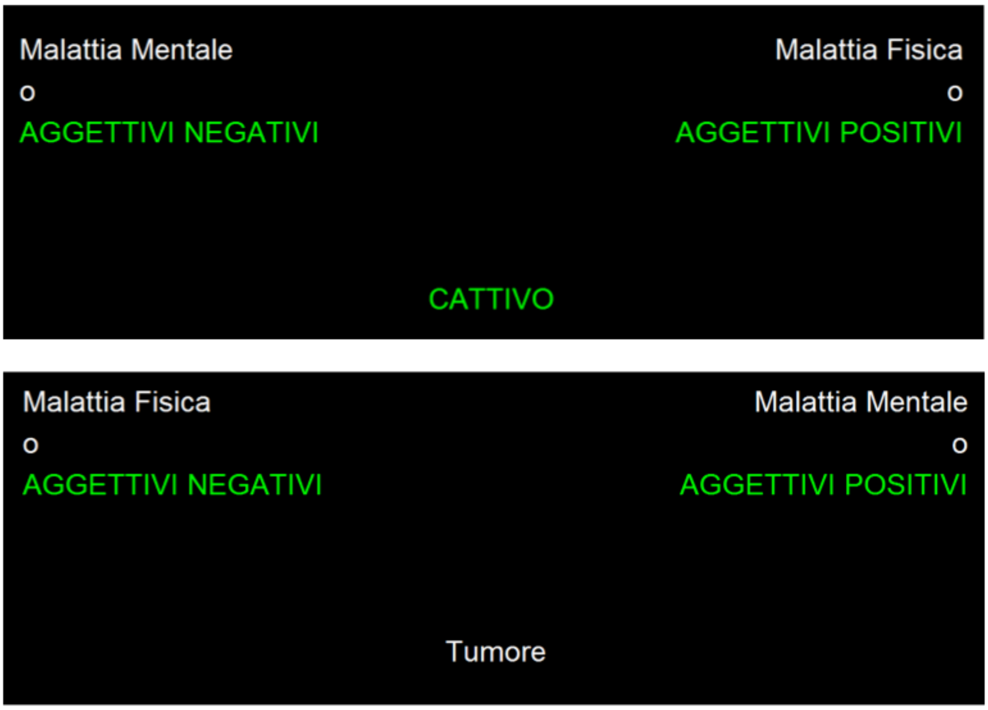


Figure 2.2 Example of a compatible trial (image on top) and incompatible trial (image on the bottom) in the mental illness attitude Implicit Association Test.

Stimuli for the *Mental Illness* and *Physical Illness* target categories are the same of the IAT for mental illness-related causal semantic associations, whereas for the evaluative categories *Positive* and *Negative* a new set of stimuli, five for each category, was created, following the same criteria of similar familiarity in the Italian daily language (CoLFIS – Bertinetto et al., 2005). The five word stimuli for the *Positive* category were the following: *beautiful* (bello), *good* (buono), *joyful* (gioioso), *safe* (sicuro), *wonderful* (splendido). For the *Negative* category the five words were *ugly* (brutto), *bad* (cattivo), *horrible* (orribile), *dangerous* (pericoloso), *sad* (triste).

2.5 Tasks scoring and preliminary data analyses

The scoring of the two IATs followed the guidelines for the D-score improved algorithm for an IAT with built-in penalty by Greenwald, Nosek, and Banaji (2003). Practice trials and latencies greater that 10000ms were removed from the dataset. Subjects with latencies lower than 300ms in more than 10% of the trials and/or with an error rate greater of 30% in one of the two critical blocks were excluded from the

Table 2.2 Task structure for the mental illness causal beliefs and attitude IAT: blocks and categories pairings.

		Categories		
Task	**Blocks**	**Left**	**Right**	**No. Trials**
IAT Causal Beliefs	Target Practice	Mental Illness	Physical Illness	20
	Attribute Practice	Psychological	Biologic	20
	Compatible Practice	Mental Illness/ Psychological	Physical Illness/ Biologic	20
	Compatible Test	Mental Illness/ Psychological	Physical Illness/ Biologic	40
	Reversed target practice	Physical Illness	Mental Illness	20
	Incompatible Practice	Physical Illness/ Psychological	Mental Illness/ Biologic	20
	Incompatible Test	Physical Illness/ Psychological	Mental Illness/ Biologic	40
				180
IAT Attitude	Target Practice	Mental Illness	Physical Illness	20
	Attribute Practice	Negative	Positive	20
	Compatible Practice	Mental Illness/ Negative	Physical Illness/ Positive	20
	Compatible Test	Mental Illness/ Negative	Physical Illness/ Positive	40
	Reversed target practice	Physical Illness	Mental Illness	20
	Incompatible Practice	Physical Illness/ Negative	Mental Illness/ Positive	20
	Incompatible Test	Physical Illness/ Negative	Mental Illness/ Positive	40
				180

dataset because of too many random responses and errors, respectively. Latencies lower than 300 ms were then recoded to 300.

Mean value for responses ($\overline{RT}$) in the compatible and incompatible practice (*pc* and *pi*) and test (*tc* and *ti*) blocks were computed and subtracted, divided by the inclusive standard deviation of practice blocks (SD_p) and test blocks (SD_t). The two partial D-scores were then averaged, to obtain a measure of the IAT effect:

$$D = \frac{\left(\frac{\overline{RT_{pi}} - \overline{RT_{pc}}}{SD_p}\right) + \left(\frac{\overline{RT_{ti}} - \overline{RT_{tc}}}{SD_t}\right)}{2}$$

The D-score was computed so that positive values indicate faster response in the compatible block (e.g., *Mental Illness/Psychological* versus *Physical Illness/Biologic*) compared to the incompatible block (e.g., *Physical*

Illness/Psychological versus *Mental Illness/Biologic*), suggesting a stronger automatic association of mental illness with the psychosocial semantic area for the causal beliefs IAT and stronger negative automatic evaluations of mental illness, compared to the physical illness, for the attitude IAT.

The correlation between the two implicit measures was then computed to test the hypothesis about the positive relationship between the two implicit facets of stigma.

A subsequent latent trait analysis of the two implicit measures was conducted separately on the two IATs by applying the *Many-Facet Rasch Measurement* model (MFRM – Linacre, 1989), which is presented in the following chapter.

Chapter 3

Mental illness aetiological beliefs and attitude:
A Many-Facet Rasch analysis of two Implicit Association Tests

3.1 Measurement theory and Rasch models: what's the state of the affairs?

Psychometrics is concerned with formulating *measurement* models for psychological attitudes, abilities, and personality traits. Valid measurement practice is possible when enough is known about the to-be-measured attribute so as to justify its operationalization into a measurement instrument, which, by definition, can represent a more or less successful attempt to provide a good estimate of the attribute in question. Two reasons ground this claim: first, the development of an instrument is based on a theory or an idea about the structure of the attribute of interest; it follows that measurement procedures do not coincide with the attribute. Second, as already mentioned earlier in this work (see Chapter 1), assigning numbers to objects with respect to an attribute does not produce measurement values, as was intended by the *operationalist* definition of measurement (Stevens, 1946). Counting the number of correct responses to an attainment measure is simply counting. It does not produce a measurement of the attribute (Michell, 1999). The qualitative reactions that people have to items are transformed in 'item scores', which have to undergo a psychometric analysis to support the hypotheses about the attribute theory and its operationalization into those measurement prescriptions.

Measurement theory provides a solid framework where the procedure of quantifying attributes can be put at the test. Two measurement perspectives are contemplated: on one hand, there is the physicist perspective represented by Additive Conjoint Measurement (ACM – Luce & Tuckey, 1964), which is typical of physical measurement and seen as the ideal practice for psychological measurement (Michell, 2008), yet hardly applicable to the complexity of psychological attributes for its rigid assumption of precision (Sijtsma, 2012). The main goal of measurement in ACM is to represent an empirical system consisting of qualitative relations and operations by finding a numerical system consisting of numerical relations and operations that has the same structure.

On the other hand, there is the statistical perspective of modern psychometrics, in particular Item Response Theory, which is more prone to take into account the complex shadows of psychological attributes by assuming a stochastic structure for the item response processes and thus overlooking the deterministic structure assumed by the physicist perspective (Borsboom & Mellenbergh, 2004).

In particular, among the corollary of models contained in the general recipient called IRT, Rasch models are considered an IRT version of the ACM and strive to the realization of ACM's ambitions. In a Rasch model, a monotone transformation of the dependent variable (the item response probability) is an additive function of two independent variables, namely person ability and item difficulty. This is precisely the way that ACM pictures the situation (Luce & Tukey, 1964). Because the model is structurally equivalent to ACM, but also incorporates probabilities to deal with measurement imprecision in the model – it bears the possibility of giving the best of two worlds.

The Rasch models were developed to specify a model that allows for the determination of person abilities that do not depend on the specific items used and reversely, that allows for the determination of item difficulties that do not depend on who responds to the items. This property is called *specific-objectivity*.

The use of Rasch modelling provides then a powerful tool for the psychometric analysis of measurement procedures devised to quantify an attribute of interest, of which the quantifiable nature needs to be demonstrated and the theoretical status updated by the feedback provided by the psychometric model. Obviously it is not assumed that the Rasch model is the *answer* for all questions, for many latent variable models do exist. However, it can bear a compelling framework in which the investigation of the measurement properties of a measure can add another tile in the comprehension of the enquired phenomenon and help unravel a complex data structure.

3.2 A Many-Facet Rasch analysis of implicit measures: rationale

Since its development, the Rasch modelling approach has a long tradition in the development and psychometric analysis of psychological, educational, and medical assessment tools, and it has been used as a template that operationalises in a very flexible form the formal axioms of ACM, which underpin measurement and against which data collected from self-report measures may be tested for measurement validity (e.g., Karabatsos, 2001; Kyngdon, 2011). Since the model defines measurement, data are fitted to the model to see if they meet the model's expectations. This is opposite to the practice in statistical modelling where models are developed to best represent the data. Fitting data to the Rasch model offers then an elegant approach to address several methodological key aspects generally associated with scale development and construct validation, as well as providing a log-odds transformation of the ordinal raw score.

Given the inner assessment features of implicit measures, including the IAT, the adoption of a Rasch modelling perspective seemed to be a possible answer to the question whether it is possible to reach a deeper comprehension of the IAT measure and to run a first attempt to establish its measurement validity. The application of the Rasch model to implicit measures of automatic associations has seen some first endeavours in recent research (Anselmi, Vianello, Voci, & Robusto, 2013; Anselmi, Vianello, & Robusto, 2011; Robusto, Cristante, & Vianello, 2008; Vianello & Robusto, 2010). The main idea underlying the application of a latent trait modelling perspective envisions the stimuli categorization task as a variant of the item responding performance required by traditional self-report measures. According to this conceptualization, IAT stimuli can thus be considered just like questionnaire items, to which respondents should reply according to the supposed underlying psychological process(es) and/or construct(s).

Within this perspective, the methodological investigation of IAT stimuli in terms of measurement validity and reliability was then directly faced in a fashion that resembles the test development approach applied to traditional assessment measures and addressed within a latent trait modelling framework, by applying the Many-Facet Rasch Measurement model (MFRM – Linacre, 1989). There are several advantages for using a Rasch model in the investigations of implicit associations and implicit measures:

a) All Rasch models conform to the properties of stochastic independence, specific objectivity, linearity, and measurement unit (for a discussion, see Bond & Fox, 2007);

b) The MFRM allows modelling, besides the traditional subject and item parameters, other variables, or *facets*, that might interfere and affect the outcome of a rating process (traditional self-report measures) or of a task (stimuli categorization);
c) All facets are located on the same latent continuous trait, allowing comparisons between their elements;
d) All of the model parameters, or *facets*, lie on a common latent dimension of categorization accuracy;
e) As a consequence of the specific objectivity, the measures obtained by the model are sample-, stimulus-, condition-, and all other facet-free and can be compared with any other;
f) Specific goodness-of-fit statistics assess the fit of the data to the model and are highly informative about the results interpretation of each single item, participants, association condition, task block, or any other relevant variable in the model;
g) The MFRM allows interaction analyses among different facet parameter estimates, to detect any differential functioning of any facet parameter estimate in relation to the other variables entered in the model. The last feature is of great importance in so far as it provides a powerful tool to examine systematic patterns of deviations from the model expectations in the data, and to identify possible factors causing this patterns (e.g., the procedure of counterbalancing the trial blocks in speeded reaction-time tasks).

Last but not least, in the specific case of implicit measures, the estimation of the parameters and the calculation of the associated measurement errors provide a simple and direct means of determining the significance of the differences between the experimental conditions, which will then represent individual and group-level measures of implicit associations (Vianello & Robusto, 2010).

3.3 The present study: objectives

In the present study, the psychometric analysis of two implicit measures of automatic semantic and evaluative associations with mental illness was addressed. The main objectives entailed the analysis of the contribution of specific associations to the overall implicit measures by decomposing the general IAT effects into its specific components supplied by stimuli and categories. The guiding approach considered investigating how the speed of categorization of individual stimuli changed according to the associative condition they were presented in. The analytical procedure guiding the present study is different from sorting the IAT trials into subsets and computing separate IAT effects for the two targets, which is not recommended in analyzing IAT data (Nosek, Greenwald, & Banaji, 2005). Instead, the differential contribution of individual stimuli to the overall IAT effect was assessed while keeping the relative nature of the measure.

Being a Rasch model, the MFRM provides the researcher with a rigorous measurement system, in so far as the speed of categorization of the stimuli is expressed by interval measures characterized by a common measurement unit, which, if the data fit the model, maintains the same size over the entire continuum. It follows that the measurement and comparison of different elements is more precise. Moreover, the MFRM allows investigating whether the speed of categorization of individual stimuli differs in the two traditional associative conditions (i.e., compatible and incompatible sorting conditions) and in relation to potential external variables that can have an impact on the to-be-measured construct (e.g., previous experiences with mental illness). For instance, one of the most criticized method variable affecting the IAT effect and participants' performance has been the order of presentation of combined critical blocks presentation (compatible-incompatible versus incompatible-compatible), which is normally counterbalanced across participants (i.e., *compatibility effect* – Klauer & Mierke, 2005; Teige-Mocigemba, Klauer, & Sherman, 2010).

A Rasch measurement perspective endowed a great potential for understanding the meaning of implicit measures. Let us consider the object of the present research, namely implicit associations of mental illness with the psychosocial or biogenetic semantic realms, or with negative or positive automatic evaluations. If the stimuli that mostly contribute to the measure are related to the biogenetic realm, rather than to the psychosocial domain, then the implicit association should be interpreted as rooted on the automatic activation of biogenetic concepts, or representations, that are part of the associative network of 'mental illness', and more strongly bound to the concept of mental illness, relative to the 'physical illness' associative network (which is supposed to mostly activate bio-genetic concepts rather than psycho-social ones). The strength of an implicit association is related to the ability of one concept to activate another. According to a neural network theory, this instigates a spread of activation between two connected links. This association makes it easier to process subsequent similar elements (e.g., Greenwald et al., 2002; Strack & Deutsch, 2004). Hence, the main hypothesis refers to the ability of mental illness concept in triggering the linked associations stored in the network.

The same line of thought applies to the evaluative automatic associations to mental illness: if the stimuli that contribute most to the implicit measure are the positive ones, then the implicit negative association to mental illness could be triggered by an implicit 'preference' for physical illness rather than a real anti-mental illness attitude (i.e., phenomenon called *positive association primacy*; Anselmi et al., 2011; Popa-Roch, & Delmas, 2010; van Ravenzwaaij, van der Maas, & Wagenmakers, 2011). This needs yet to be verified.

Disentangling the contribution of individual stimuli can then be useful to provide a detailed picture of the implicit associations that mostly underlie the measure and that might differ across individuals, and allows a more precise definition of the enquired construct. Furthermore, the modelling properties of the MFRM offer a flexible tool for the investigation of the effect of other independent variables on the implicit measure. For instance, it may help answering the question concerning the impact of both method variables related to the measurement procedure, such as the order of presentation of the critical combined blocks, and of variables related to the individual differences in the population, such as the presence of previous experiences with mental illness, which might moderate the automatic associations aroused by the activation of this concept.

The application of the MFRM analysis to the IAT data was operationalized as follows:

1. Verification of a common latent trait wherein the IAT stimuli parameter estimates express their location on the latent dimension on a common measurement unit, which described their speed of categorization;
2. Estimation of the local measures of the stimuli in the two IAT critical blocks, considered as two partial sub-dimensions, and test for the different contribution of each stimulus to the IAT effect;

3. Estimation of other facets parameters for the two critical blocks (i.e., blocks order, previous contact and personal experience with mental illness);
4. Analysis of the effect of task blocks counterbalancing on the stimuli functioning across the critical blocks;
5. Analysis of the two-way interactions of each stimulus latency estimate in relation to each facet, to test any differential functioning of the IAT stimuli in relation to possible external confounders, beyond their difficulty and the ability of participants.

3.3.1 The model

The MFRM (Linacre, 1989) derives from the *Simple Logistic Model* (SLM – Rasch 1960), which is the traditional and most basic Rasch model for the transformation of ordinal observations into interval measures. The SLM is meant for dichotomous data and expresses, according to a logistic distribution, the probability of a response *x* to a test, which can be correct (1) or incorrect (0), as a function of the ability β of respondent *v* and difficulty δ of the item *i,* expressed on the *logit* scale ($\beta_v - \delta_i$) (Rasch 1960), as formalized in the following mathematical form:

$$P(X_{vi} = x_{vi} | \beta_v, \delta_i) = \frac{\exp[\, x_{vi}(\beta_v - \delta_i)]}{1 + \exp(\beta_v - \delta_i)} \tag{1}$$

The more (or less) able the individual is and the easier (or more difficult) the item is, the more (or less) probable it will be that a correct response will be obtained.

By using Equation 1 it is possible to compute the probability of a correct response and of an incorrect response as follows:

$$P(X_{vi} = 1 | \beta_v, \delta_i) = \frac{\exp(\beta_v - \delta_i)}{1 + \exp(\beta_v - \delta_i)} \tag{2}$$

and

$$P(X_{vi} = 0 | \beta_v, \delta_i) = \frac{1}{1 + \exp(\beta_v - \delta_i)} \tag{3}$$

By computing the logarithm of the ratio of Equation 2 and 3, then it obtains

$$ln\frac{P(X_{vi}=1|\beta_v,\delta_i)}{P(X_{vi}=0|\beta_v,\delta_i)}=$$

$$=ln\frac{\frac{\exp(\beta_v-\delta_i)}{1}/[1+\exp(\beta_v-\delta_i)]}{1/[1+\exp(\beta_v-\delta_i)]}=$$

$$=\beta_v-\delta_i \qquad (4).$$

It is evident from the equations presented so far that the individual's ability and item difficulty can be considered as two facets that interact with each other to produce the response to an item, and that can be modelled to operate independently, so that their parameter estimates can be combined additively on a latent variable. However, in the measurement contexts complex situations are more the rule than the exception, and other aspects may interfere with the person and item's attributes, such as specific situational, social, and personality attributes. As is evident in Eq. 4 it is then possible to introduce other facets that lie on the same latent trait.

Within the context of Rasch modelling mono-dimensionality and mathematical properties (for a review see, Bond & Fox, 2007), Linacre (1989) developed an extension of the SLM, namely, the MFRM, which extends the analysis to more complex situations by including other sources of systematic variability (facets), in addition to respondents' ability and item (or stimulus) difficulty, accounting for the likelihood of a response.

Consequently, in the present study additional facets besides respondents' categorization ability (i.e., speed of categorization) (facet 1) and stimuli easiness of categorization (facet 2), were entered in the model equation, to account for other variables that may affect the performance of the task, namely, the critical block of stimuli presentation (facet 3), and some variables accounting for individual differences previously shown to influence attitudes towards mental illness, such as of the presence any previous contact with mentally ill people (facet 4) and personal experience with psychological suffering and/or mental problems (facet 5). An additional parameter accounting for the response latency rating scale $k = \{1, \ldots, m\}$ provided by the response latency distribution discretisation, was embedded in the model.

In the present study, two MFRM model estimations were carried on: a first one on the data matrix with the median values of IAT stimuli discretized over the pooled critical blocks, and a second one on the data matrix with the median values discretized over the compatible and incompatible block separately. The two MFRM model equations are then formally expressed as follows:

$$ln\left(\frac{P(X_{vibk})}{P(X_{vib(k-1)})}\right) = \beta_v - \delta_i - \lambda_b - \tau_k \quad (5)$$

$$ln\left(\frac{P(X_{vibcdk})}{P(X_{vibcd(k-1)})}\right) = \beta_v - \delta_i - \lambda_b - \gamma_c - \eta_d - \tau_k \quad (6).$$

Equation (5) specifies the probability that a respondent *v* would respond to stimulus *i* in the task order setting *b* with response speed *k* rather than *k – 1;* β_v is the person *v*'s ability (categorization speed) parameter, δ_i is the stimulus *i* difficulty (ease of categorization), parameter, λ_b identifies the different order of presentation of the critical blocks, and τ_k is the parameter for the step up to category *k* rather than *k – 1* of the response latency rating scale.

Equation (6) specifies the probability that a respondent *v* would respond to the stimulus *i* in the critical block *c*, given his/her previous experience *d* with mental illness, with a speed *k* rather than *k – 1*; β_v is the person *v*'s ability (categorization speed) parameter, δ_i is the stimulus *i* difficulty (ease of categorization), parameter, λ_b identifies the different order of presentation of the critical block *b*, γ_c describes the presence *c* of any previous experience with mental illness, η_d describes any experience with personal psychological problems and/or disease, and τ_k is the parameter for the step up to category *k* rather than *k – 1* of the response latency rating scale.

The Rasch model parameter are additive, hence satisfying one of the requisites for interval measures, and are based on the transformation of scores into a *logit* scale, i.e., the logarithmic transformation of the probability of giving a particular response, given certain conditions (e.g., participants' ability, stimuli recognisability, difficulty of the critical block, previous experiences with mentally ill people, and personal experience with mental illness). In Equations 5 and 6, the *logit* of a certain response *k* can be seen as the dependent variable, whereas the various factors act as independent variables that influence (or control) the response.

All parameter estimates were positively scaled in the analyses, so that positive values indicate fast responses, whereas negative measures indicate slow responses.

To evaluate the goodness-of-fit of the parameter estimates, the MFRM presents two fit statistics which show how much the data for each parameter adhere to the model requirements: the *mean square Outfit* and *mean square Infit* statistics. These statistics are calculated for each participant, each item, and any other facet parameter, and express the relationship between observed and model-derived expected scores, ranging from zero to infinity. Statistics equal to or near 1 indicate perfect correspondence between observed and expected values; statistics above 1 indicate the presence of greater variance than that modelled (*underfit*); and statistics below 1 indicate the existence of lower variance in the data than that predicted by the model (*overfit*). A range of .50–2 indicates a satisfactory fit of the observed data to the model requirements (Bond & Fox, 2007; Linacre, 2009). *Infit* and *Outfit* statistics are both derived from the squared standardized residuals for each item/participant interaction (for details, see Myford & Wolfe, 2003). The *Outfit* statistic is the average of the squares of the standardized residuals and is unweighted, meaning that it is more sensitive to outlier observations. For the *Infit* statistic the residuals are information-weighted by their individual variance, thus relatively more affected by inlying response patters. In other words, the *Outfit* statistic places greater emphasis on the residuals associated with responses that are farther from the measure of a given element, whereas the *Infit* statistic gives greater emphasis to those responses that are nearest to the measure of a given element (Bond & Fox, 2007).

A Chi-square statistic – the *Fixed (all same)* χ^2 – is also provided for each facet, and tests the hypothesis that the elements of a facet have the same *logit* in relation to the measurement error (SE). In other words, the Chi-square statistic helps to reject the null hypothesis that there is no group-level difference in the different elements composing a facet. For instance, a *Fixed (all same)* χ^2 with an associated probability value lower than .05 points to the presence of group-level implicit associations between the task targets and attributes, which are differently paired in the compatible and incompatible blocks.

When evaluating the model fit and the estimated measures, other indices can be informative, such as the *separation ratio* (G) and the *separation reliability* (R).

The *separation ratio* (*G*) represents a measure of the difference between the scores obtained by the elements of the facet in relation to their precision (Linacre, 2009a; Myford & Wolfe, 2003). It is expressed as the relationship between the "true" standard deviation (i.e, the standard deviation of the estimates corrected for measurement error: $adjSD = SD - RMSE^2$) and the average of the standard error of the elements (*RMSE*): $G = adjSD / RMSE$ *.* The G index is extremely important in the analysis of the critical blocks utilized in experimental procedures. If only two conditions are included in the analysis, the G is a measure of the mean automatic association effect among participants (Vianello & Robusto, 2010). The *G* of the facet conditions can be interpreted as a measure of the sensitivity of the instrument, and therefore, it is relevant, for example, in a study in which groups are hypothesized to be strongly polarized and the expected value should be elevated. However, in the case of implicit measures, the separation of the participant facet is not as important as in traditional intelligence and attitude tests, where it represents a measure of the resulting discrimination, because implicit measures of associations are based on a comparison (bias/interaction analysis) between the performance in one condition (e.g., *mental illness/psychological*) and that in another condition (e.g., *mental illness/biologic*). In implicit techniques, the general level of performance (speed of response) is not of direct interest. It is theoretically possible to obtain a good measure of implicit association even without discriminating between participants in terms of their ability in completing the tasks. *G* for the participants' facet simply gives an idea of how difficult the procedure is, and, all things being equal, it is preferable to obtain a measure that is just as difficult for all the participants; therefore, low indexes of separation between participants are expected (Vianello & Robusto, 2010). Conversely, when considering the stimulus facet, *G* provides useful information concerning the degree to which the stimuli represent the trait examined.

The *separation reliability* (R) index indicates how well the elements of a facet are separated to reliably represent the facet and ranges from 0 to 1. It reflects an estimation of the relationship between true scores and true variance: $R = trueSD^2 / observedSD^2 = G^2 / (1 + G^2)$, where observed *SD* is the standard deviation of the estimates (not corrected for measurement error). If $R < .5$, the value of *G* (separation) is probably due to measurement error. The expected value is high if homogeneity is expected between the facets and low if separation is expected. In the case of experimental procedures for the assessment of automatic associations, the reliability (*R*) of the items gives us a measure of their equivalence (or interchangeability). Thus, it is desirable to obtain low reliability indexes for the facet item (Vianello & robusto, 2010).

Once estimated each facet measures, it is possible to compare different parameter estimates (i.e., *logits*) by standardizing their difference, which approximates to the Student's *t* distribution, $t = \frac{\delta_1 - \delta_2}{\sqrt{SE_1^2 + SE_2^2}}$, with degrees of freedom (*dfs*) equal to the sum of the respective *dfs.*

After estimating the model parameters, the MFRM gives the possibility to carry out bias/interaction analyses, i.e, the analysis of the interactions between elements of different facets (for details, see Linacre, 2010). A bias can be due to any kind of interaction, such as differential stimuli functioning, differential person functioning or differential functioning of any other facet, and is estimated from the residuals left over after estimating the parameters in the main analysis (Linacre, 2010), and tested for statistical significance by means of *t* statistic. This feature allows identifying possible factors causing any systematic deviation from the model expectations in the data, such as the associative condition the stimuli are presented in (i.e., the compatible or incompatible blocks). In particular, the *differential stimulus functioning* (DSF) analyses the interaction between the elements of the facet *stimuli* and elements of other facets. For instance, the *bias index* involves introducing an interaction parameter into the model between the facets (e.g., for the stimuli × critical blocks interaction). The *logit* of a stimulus *i* in a critical block is computed by adding a bias measure to the overall speed of categorization of the same stimulus (δ_i) if the response to the latter is faster in the critical block than overall and by subtracting it if the response to the stimulus is slower. The two biased stimulus measured are then subtracted. To test for the interaction significance, such difference are terms are transformed into *t* points used to run pairwise contrasts (Linacre, 2009) between the two biased stimulus measures in the two critical blocks, divided by their joint *SE* ($SE_{ij} = \sqrt{\left(SE_i^2 + SE_j^2\right)}$). The *dfs* of the *t* value for the difference between the *logits* of two elements is the number of "free" observations for each element ($df = N_i - 1 + N_j - 1$).

The use of *t* values is a straightforward way to add significance to the MFRM measures interactions, for they are standardized, usually reliable, easily interpretable, and normally distributed (if *df* > 30). They can be computed both for a single element (dividing the Rasch measure by its *SE*) and for a difference of *logits* (using the joint *SE*). Furthermore, a Cohen's *d* can also be computed ($d = \frac{2t}{\sqrt{df}}$; Rosenthal & Rosnow, 1991), to have a quantification of the contribution of that stimulus to the general IAT effect.

In the present study, the two-way interaction analyses of each stimulus measure by the task critical block in which stimuli were presented, previous contact with mentally ill people, and any personal experience with psychological problems, were conducted.

FACETS software (version 3.66.0) was used for the analyses (Linacre, 2009).

3.3.2 Data pre-processing

Prior to the MFRM analyses, each IAT dataset was pre-processed as previously described in Chapter 2, following Greenwald, Nosek, and Banaji (2003)'s recommendations. Only critical trials data were used for the analyses. Latencies greater than 10000ms were discarded from the dataset. No participants presented response latencies lower than 300ms in 10% of the trials. Twenty-nine and three participants were excluded from the causal beliefs IAT and the attitude IAT datasets, respectively because of an error rate greater than 30% in at least one of the two critical blocks. The resulting usable data are then composed of 331 participants for the causal beliefs IAT and 357 for the attitude IAT.

The preparation of IATs data for the MFRM analyses was done according to the following steps:

1. For each IAT stimulus *i* three median values were computed: a median value of response latencies to stimulus *i* in the pooled critical blocks, a median value of response latencies to stimulus *i* in the compatible critical block, and a median value of response latencies to stimulus *i* in the incompatible critical block. The median descriptive statistics was chosen

as a measure of the latency central tendency of participants' responses to the stimuli, due to the lower sensitivity to the distribution tails than the mean statistic.

2. For each of the three median latencies distributions (pooled, compatible, and incompatible) the distribution of the stimuli latency median values was successively discretised in four categories according to the quartiles (25^{th}, 50^{th}, 75^{th}) computed on the pooled dataset, the compatible dataset, and the incompatible dataset, to index *very fast* (1), *fast* (2), *slow* (3), and *very slow* (4) response latencies (Blanton & Jaccard, 2006). This discretization procedure complies with the Rasch modelling requirement of entering only discrete variables in the model.
3. Two matrices were created: a first P × S matrix, where P identifies participant *n* and S identifies the latency score of stimulus *i* in the pooled critical tasks; and a second matrix P × S × C, where P identifies participant *n* and S identifies the latency score of stimulus *i* computed on the critical block *j*.
4. A binary variable coding for the two orders (compatible-incompatible and incompatible-compatible) was added to each matrix to test for its confounding effect on the categorization task.
5. Two additional categorical variables coding for the participants' previous contact with mentally ill people (none, scarce, moderate, and high) and for any experience of personal psychological problems (yes or no), were added to each compatible and incompatible dataset.

3.4 Results

3.4.1 Preliminary analyses

In the IAT measure for automatic semantic associations with mental illness, the participants displayed a significant propensity to automatically associate mental illness with psychological aspects ($\overline{D}$ = .9862, *SD* = .36; one-sample *t*-test: $t_{(330)}$ = 49.701, *p* = .001), whereas the attitude implicit measure did not reveal any evident association of mental illness with neither negative nor positive automatic evaluations ($\overline{D}$ = -.0008 SD = .555; one-sample *t*-test: $t_{(355)}$ = -.030, *p* = .976). The two implicit measures resulted positively correlated (Spearman ρ = .246, p < .001, 95% C.I. = [.134, 352]; $t_{(329)}$ = 4.6035, p < .001).

3.4.2 MFRM analysis of the mental illness causal beliefs IAT

When analysing participants' latency scores computed across the two critical blocks, fit indices were excellent for the 22 stimuli (.79 ≤ Infit/Outfit ≤ 1.32), indicating that they are measuring a common latent trait. Only sixteen respondents (4.83%) presented a misfit to the model requirements (i.e., Infit and Outfit greater than 2 or lower than .5). Differences in the general respondents' speed were satisfactory (β ranged from -4.22 to 4.22, $\bar{\beta}$ = -.04, $\overline{SE}$ = .30, SD = 1.22; $\chi^2_{(330)}$ = 3329, p < .01), reproducible (R = .93) and three times and a half greater than the imprecision of their estimates (G = 3.61).

Across the critical blocks, the compatible-incompatible blocks (CI) order condition was more difficult (λ_{CI} = .10) than the incompatible-compatible (IC; λ_{IC} = -.10; $\chi^2_{(1)}$ = 45.8, p < .01, G = 6.69, R = .98; 1.00 ≤ Infit/Outfit ≤ 1.02), meaning that, at the group level, respondents slowed down when the pairing Mental Illness/Psychological versus Mental Disease/Biologic appeared first. The distance between the locations of the two orderings on the latent trait (.20) represents the size in *logit* of the counterbalanced ordering effect on the general performance of the task. A bias/interaction analysis of indices × blocks ordering evidenced a DSF only for two stimuli, the images psycho2 (*grandparents/kids relation*; $t_{(327)}$ = 2.18, p = .03) and psycho4 (*two friends arguing*; $t_{(327)}$ = 2.67, p = .008), which were sorted relatively quicker in the IC ordering.

Table 3.1 provides overall (i.e., across the two task conditions) and local (i.e., in each task condition block) *logits* of each stimulus (δ_i). Stimuli local measures were computed by estimating them separately for the two critical blocks, compatible and incompatible (see §3.3.1, Equation 6).

The stimuli were generally categorized with different speed rates (δ range = [-.70, 1.15]; $\chi^2_{(21)}$ = 1781.7, p < .01; G = 9.26, R = .99). The most recognizable stimulus, in terms of speed of categorization, was *flu* (*Influenza*) for the Physical Disease target category. The least recognizable stimuli were the images bio2 (*DNA filament*) and bio6 (*chromosomes*), which took a bit more time to be recognized during the whole task, relative to the others.

Table 3.1 Speed of categorization of stimuli in the mental illness causal beliefs IAT: median latency values and δ estimates across the critical blocks and in each block. standardized contrast values (*t*) and Cohen's *d* effect size between the two critical blocks.

	Across Blocks					Compatible			Incompatible						
Stimulus	**Median**	**δ**	***SE***	**Infit**	**Outfit**	**Median**	**δ**	***SE***	**Median**	**δ**	***SE***	***t***	***df***	***p***	**Cohen's *d***
Tumour	816.75	.40	.07	.88	.85	620	.12	.08	1160	.48	.09	-2.990	656	.005	**-.233**
bio2	684.5	-.70	.07	1.11	1.14	575.5	-.81	.10	928	-.45	.08	-2.811	656	.008	**-.220**
Depression	874.5	.77	.07	.88	.82	656	.45	.08	1178	.78	.10	-2.577	656	.015	**-.201**
Psychopathy	834.25	.44	.07	.98	1.12	644	.26	.08	1128.5	.55	.09	-2.408	656	.022	**-.188**
bio4	691.25	-.68	.07	1.05	1.01	564.25	-.67	.09	929.5	-.48	.08	-1.578	656	.115	-.123
bio1	706	-.47	.07	1.25	1.32	575.5	-.52	.09	969.5	-.35	.08	-1.412	656	.147	-.110
bio6	701	-.70	.07	1.02	1.03	567.5	-.77	.10	898.5	-.66	.08	-.859	656	.276	-.067
Diabetes	888	.88	.07	.91	.92	681	.71	.08	1162	.80	.10	-.703	656	.311	-.055
Pneumonia	877.25	.87	.07	.97	.98	678	.72	.08	1226.5	.80	.10	-.625	656	.328	-.049
Schizophrenia	805	.32	.07	.79	.79	639.5	.27	.08	116.75	.30	.09	-.249	656	.387	-.019
Paranoia	859	.58	.07	.93	.89	662	.48	.08	1137	.48	.09	.000	656	.399	.000
bio5	704.5	-.58	.07	1.16	1.17	581	-.50	.09	928	-.54	.08	.332	656	.377	.026
psycho4	701.75	-.50	.07	1.22	1.20	603.75	-.32	.09	936.25	-.37	.08	.415	656	.366	.032
bio3	699.75	-.53	.07	1.14	1.15	580.75	-.50	.09	925.5	-.58	.08	.664	656	.320	.052
Heart Attack	855.75	.70	.07	.85	.84	657.5	.61	.08	1135.5	.49	.09	.997	656	.243	.078
Hysteria	861.25	.69	.07	.91	.87	667.5	.67	.08	1158	.52	.09	1.246	656	.184	.097
psycho3	720.75	-.37	.07	1.08	1.15	600	-.17	.08	966.5	-.33	.08	1.414	656	.147	.110
psycho5	699.5	-.63	.07	1.12	1.14	606.5	-.24	.09	927.75	-.41	.08	1.412	656	.147	.110
psycho2	688	-.62	.07	1.14	1.15	586	-.41	.09	911.25	-.64	.08	1.910	656	.064	.149
Flu	920	1.15	.08	.86	.89	719	1.11	.07	1212	.80	.10	2.540	656	.016	**.198**
psycho6	717	-.53	.07	.91	.95	591	-.27	.09	909.75	-.61	.08	2.824	656	.008	**.220**
psycho1	703.5	-.48	.07	1.04	1.04	596.5	-.24	.09	921	-.59	.08	2.907	656	.006	**.227**

Note: The *t* values test the hypothesis that difference between the local δ is equal to zero.
Bio1: cell; Bio2: DNA filament; Bio3: atom structure; Bio4: test tubes; Bio5: microscope; Bio6: chromosomes. Psycho1: mother/child relation; Psycho2: Grandparents/kids relation; Psycho3: job meeting; Psycho4: friends arguing; Psycho5: romantic relationship; Psycho6: family.

When considering the local δ estimates, it is possible to break up the IAT effect into the individual component supplied by each stimulus. The overall speed of categorization of some stimuli significantly changed according to the block condition they were presented in. Table 3.1 provides the DSF across the critical blocks. The *t* values in the table test the null hypothesis that the difference between the local *logits* of the stimuli is equal to zero. A Cohen's *d* was also computed for each contrast to give a standardized effect size of the stimuli that mostly contributed in triggering the IAT effect. The stimuli that are associated with a statistically significant negative *t* increase the size of the overall IAT effect. The stimuli that are associated with a statistically significant positive *t* reduce the overall IAT effect.

Compared with their overall speed of categorization, four stimuli mostly triggered the IAT effect (*tumour*, the image of a *DNA filament*, *depression,* and *psychopathy*), as they were categorized faster when presented in the Mental Illness/Psychological versus Physical Illness/Biologic pairing (compatible block); whereas stimuli *flu*, and psycho6 (*family relation*) and psycho1 (*mother/child relation*) images played an aversive role in the emergence of the effect, because their different recognisability favoured the pairing Physical Illness/Psychological versus Mental Illness/Biologic (incompatible block).

Indeed, the difference in recognition between all the Psychological and Biological images was not statistically significant *within* the incompatible block ($t_{(656)}$ = .162, $p > .05$), whereas Biologic stimuli were categorized faster in the compatible sorting condition ($t_{(656)} = 2.749$, $p = .006$). Also Mental Illness and Physical Illness target categories were similarly categorized *within* the incompatible block ($t_{(656)} = -1.113$, $p > .05$), whereas Mental Illness stimuli were consistently categorized faster *within* the compatible pairing ($t_{(656)} = -2.04$, $p = .041$).

The difference in recognisability for target and attribute categories was also tested *between* the critical blocks, to evidence the impact of the four categories in eliciting the IAT effect. In this case the difference *between* critical blocks was expected to be statistically significant. The difference in recognition marginally affected only the Psychological category, which was slightly sorted out quicker in the incompatible block, albeit it didn't reached statistical significance ($t_{(656)} = 1.818$, $p = .069$). The Biologic attribute category and Mental Illness and Physical Illness target categories were similarly categorized between the two critical blocks ($ts_{(656)}$ range = [-.96, -.16], $p > .05$).

To account for relevant individual differences variables that can impact on the semantic associations between mental illness and the psychological and biogenetic domains, relative to physical illness, the previous experience with personal psychological problems and any type of contact (in terms of closeness) with mentally ill people were entered in the model (see §3.3.1, Equation 6) as *facets*, i.e. as independent variables contributing to the latency scores to the stimuli (see Table 3.2).

Table 3.2 Facet measures (*logit*) for blocks order, personal experience and previous contact with mental illness for the compatible and incompatible critical blocks in the mental illness causal beliefs IAT.

	Compatible block		**Incompatible block**	
Facet	**Measure**	***SE***	**Measure**	***SE***
Blocks order				
CI	.00	.02	.06	.03
IC	.00	.03	-.06	.02
	$\chi^2_{(1)} = 0$, $p > .05$		$\chi^2_{(1)} = 9.1$, $p < .01$; G =2.85, $R = .89$	
Personal Experience				
Yes	.01	.07	-.04	.02
No	-.01	.02	.04	.03
	$\chi^2_{(1)} = .1$, $p > .05$		$\chi^2_{(1)} = 4.9$, $p = .03$; $G = 1.99$, $R = .80$	
Contact				
None	-.13	.07	-.03	.07
Scarce	.02	.03	.05	.04
Moderate	.04	.03	-.02	.03
High	.07	.04	.01	.05
	$\chi^2_{(3)} = 6.6$ $p = .08$; $G = 1.69$, $R = .74$		$\chi^2_{(3)} = 2.8$, $p > .05$	

Note: CI stays for compatible block first; IC stays for incompatible block first.

Table 3.2 also evidences that the order of presentation of the two critical blocks was relevant only in the incompatible sorting condition, which was performed quicker when presented first ($\lambda_{IC} = -.06$). However, the same primacy effect of the first block presented did not affect the compatible sorting task, which was equally performed in terms of speed whenever it was presented. A bias/interaction analysis for stimuli × blocks order in the incompatible block resulted in two stimuli differentially categorized in the two ordering conditions: the pictures psycho6 and bio4 (*family* and *test tubes* pictures) were categorized slower when the incompatible block was presented first ($ts_{(326)}$ = [-4.44, -2.06], $p < .05$).

Participants who have or haven't presented psychological problems did not presented any difference in the compatible sorting task, but those who have personally experienced a mental disease or any form of psychological suffering resulted to be more quicker in performing the incompatible task condition (Physical Illness/Psychological versus Mental Illness/Biological). In regard to any form of personal contact and closeness with mentally ill people, the participants who haven't ever had any form of contact with mentally ill people person performed slightly quicker the compatible task condition (Mental Illness/Psychological versus Physical Illness/Biologic; γ = -.13, *SE* = .07; Infit = .89, Outfit = .90) than those who did have (γ range = [.02, .07], *SE* range = [.03, .04]; 1.01 ≤ Infit/Outfit ≤ 1.08). Yet, the differences at the group-level between the elements of this facet didn't reach the statistical significance ($\chi^2_{(3)} = 6.6$ p = .08).

The bias/interaction analysis for stimuli × personal experience in the incompatible block did not evidence any DSF for any of the IAT stimuli; whereas the interaction between indices × previous contact in the compatible block evidenced three stimuli (*Hysteria*, *Tumour*, and the *job meeting* image) with a DSF: respondents with a previous greater contact with mentally ill people categorized them faster (*t*s = [-2.7, 3.46], *df*s = [77, 216], *p*s < .05).

3.4.2 MFRM analysis of mental illness attitude IAT

The analysis of the IAT dataset for the two critical blocks evidenced excellent fit indices for the 20 stimuli (.80 ≤ Infit/Outfit ≤ 1.22) indicating that they are part of a common latent trait. Seventeen respondents (4.77%) presented a misfit to the model requirements (i.e., Infit and Outfit greater than 2 or lower than .5). Differences in the general respondents' speed were satisfactory (β ranged from -3.91 to 3.91, $\bar{\beta}$ = .02, $\overline{SE}$ = .35, *SD* = 1.51; $\chi^2_{(355)} = 3737.3$, $p < .01$), reproducible (*R* = .93) and almost two times greater than the imprecision of their estimates (*G* = 1.32).

The order of presentation of the two critical blocks did not present any relevant difference in categorization speed in the pooled critical blocks ($\chi^2_{(1)} = 0$, p =

.96), meaning that at the group level there was no difference in the sorting task across the two pairings.

Table 3.3 provides overall (i.e., across the two task conditions) and local (i.e., in each task condition block) *logits* of each stimulus (δ_i). Stimuli local measures were computed by estimating them separately for the two critical blocks, compatible and incompatible (see §3.3.1, Equation 6).

The stimuli were generally categorized with different speed rates (δ range = [-.63, .50]; $\chi^2_{(19)}$ =422.2, p < .001; *G* = 4.64, *R* = .96). Also in the mental illness attitude IAT the most recognizable stimulus, in terms of speed of categorization, was *flu* (*Influenza*) for the Physical Illness target category. The least recognizable stimulus was *beautiful* (Positive attribute) which has been somewhat difficult to categorize during the task.

Table 3.3 presents the standardized contrasts between the stimuli local measures (DSF), which evidenced several changing patterns in the ease of categorization of nine stimuli between the two task pairings, in both directions (i.e, increase and decrease of the IAT effect). The Cohen's *d* was also computed for each contrast to give a standardized effect size of the stimuli that mostly contributed in triggering the IAT effect: negative values yields an increase of the IAT effect (faster negative associations with Mental Illness, relative to Physical Illness), whereas positive values bears the decrease of the IAT effect (faster positive associations with Mental Illness, relative to Physical Illness).

Compared with their overall speed of categorization, five stimuli mostly triggered the IAT effect (*depression*, *paranoia*, *schizophrenia*, *sad*, and *diabetes*) as they were categorized faster when presented in the Mental Illness/Negative versus Physical Illness/Positive pairing (compatible block); whereas stimuli *beautiful*, *good*, *tumour*, and *joyful* prompted a quicker classification in the pairing Physical Illness/Negative versus Mental Illness/Positive (incompatible block).

The difference in recognition between all the Positive and Negative words was statistically significant across blocks and within blocks ($ts_{(705)}$ range = [-2.26, -33.899], *p*s < .05), with positive stimuli consistently categorized faster than negative.

Table 3.3 Speed of categorization of stimuli in the mental illness attitude IAT: median latency values and δ estimates across the critical blocks and in each block. standardized contrast values (*t*) and Cohen's *d* effect size between the two critical blocks.

	Across Blocks					Compatible			Incompatible						
Stimulus	**Median**	**δ**	***SE***	**Infit**	**Outfit**	**Median**	**δ**	***SE***	**Median**	**δ**	***SE***	***t***	***df***	***p***	**Cohen's *d***
Depression	789	.15	.07	1.03	1.04	823	-.16	.07	773	.48	.07	-6.465	705	<.001	**-.487**
Paranoia	787.5	.07	.07	.93	.91	837	-.06	.07	731	.26	.07	-3.232	705	.002	**-.243**
Schizophrenia	751	-.29	.07	.86	.84	782	-.34	.07	694	-.02	.07	-3.232	705	.002	**-.243**
Sad	835.5	.38	.07	1.13	1.10	894	.17	.07	770	.42	.07	-2.525	705	.017	**-.190**
Diabetes	825	.36	.07	.97	.98	886.5	.15	.07	764	.39	.07	-2.424	705	.021	**-.183**
Flu	842.5	.51	.07	1.03	1.04	927	.32	.07	760	.43	.07	-1.111	705	.215	-.084
Hysteria	760	-.23	.07	1.05	1.00	849.5	-.12	.07	680	-.09	.07	-.303	705	.381	-.023
Psychopathy	737	-.37	.07	1.00	.92	782.5	-.33	.07	672	-.32	.07	-.101	705	.397	-.008
Marvellous	766.5	.42	.07	1.10	1.03	836	-.06	.07	685	-.06	.07	.000	705	.399	.000
Dangerous	84755	.49	.07	1.04	1.11	915.5	.32	.07	758	.32	.07	.000	705	.399	.000
Pneumonia	824	.32	.07	.91	.89	894	.26	.07	726	.23	.07	.303	705	.381	.023
Heart Attack	781	.03	.07	1.02	.97	851.5	.10	.07	707	.01	.07	.909	705	.264	.068
Horrible	761.5	-.13	.07	1.11	1.22	852.5	-.05	.07	674	-.15	.07	1.010	705	.239	.076
Safe	825	-.45	.07	1.10	1.12	921	.32	.07	719	.14	.07	1.818	705	.076	.137
Ugly	803	-.09	.07	1.13	1.16	892	.14	.07	690	-.05	.07	1.919	705	.063	.145
Bad	777	-.02	.07	1.00	1.01	880	.11	.07	692	-.09	.07	2.020	705	.052	.152
Beautiful	723	-.63	.07	.92	.90	784.5	-.39	.07	634	-.62	.07	2.323	705	.027	**.175**
Good	755	.06	.07	.92	.93	837	-.11	.07	681	-.34	.07	2.323	705	.027	**.175**
Tumour	758	-.29	.07	.83	.80	831	-.10	.07	666	-.34	.07	2.424	705	.021	**.183**
Joyful	743.5	-.29	.07	1.05	1.2	796	-.19	.07	643	-.60	.07	4.142	705	<.001	**.312**

Note: The *t* values test the hypothesis that difference between the local δ is equal to zero.

Also Mental Illness words were categorized faster than Physical Illness words across the blocks ($t_{(705)}$ = -3.232, p < .01) and *within* the compatible task condition $t_{(705)}$ = -3.515, p < .001), except in the incompatible block ($t_{(705)}$ = -.828, p > .05).

When computing the contrasts *between* the two critical blocks, Mental Illness and Positive categories only presented a statistically significant difference in the ease of categorization: the first was categorised faster in the Mental Illness/Negative versus Physical Illness/Positive pairing ($t_{(705)}$ = -2.667, p = .01); the second was categorised faster in the reversed pairing ($t_{(705)}$ = -2.1213, p = .03). Physical Illness and Negative categories did not present any difference ($ts_{(705)}$ = [.02, .485], ps > .05).

The impact of the *facets* representing previous personal experience and contact with mental illness was verified on the latency scores of this IAT as well (see Table 3.4).

Table 3.4 Facet measures (*logit*) for blocks order, personal experience and previous contact with mental illness for the compatible and incompatible critical blocks in the mental illness attitude IAT.

	Compatible block		**Incompatible block**	
Facet	**Measure**	***SE***	**Measure**	***SE***
Blocks order				
CI	-.02	.02	-.06	.02
IC	.02	.02	.06	.02
	$\chi^2_{(1)}$ = 1.5, p > .05		$\chi^2_{(1)}$ = 17.7, p < .01; G =4.08, R = .94	
Personal Experience				
Yes	.05	.02	-.07	.03
No	-.05	.03	.07	.02
	$\chi^2_{(1)}$ = 11.1, p > .01; G = 3.18, R = .91		$\chi^2_{(1)}$ = 21.7, p < .01; G = 4.55, R = .95	
Previous Contact				
None	.19	.05	.07	.05
Scarce	-.05	.03	-.08	.03
Moderate	-.07	.04	-.03	.02
High	-.07	.02	.05	.04
	$\chi^2_{(3)}$ = 23.3 p < .01; G = 3.35, R = .92		$\chi^2_{(3)}$ = 11.5, p = .01; G = 1.65, R = .73	

Note: CI stays for compatible block first; IC stays for incompatible block first.

Also in this IAT the order of presentation of the two critical blocks presented a similar pattern of results: no difference for the compatible block and an opposite trend in the incompatible block, which was slower when presented first (λ_{CI} = -.06). The bias/interaction analysis revealed the words *bad* and *safe* to be categorized slower when the incompatible sorting task was presented first ($ts_{(352)}$ = [-3.17, -2.31], p < .05), the word *schizophrenia* to be categorized quicker when the block was presented first ($t_{(352)}$ = 2.90, p = .004).

Participants who haven't ever presented personal psychological problems resulted to be quicker in the sorting task for the Mental Illness/Negative pairing, relative to Physical Illness/Positive contrast pairing (η = -.05, *SE* = .03; Infit = .94, Outfit = .92), and slower in the reversed condition (i.e., Mental Illness/Positive versus Physical Illness/Negative) (η = .07, *SE* = .02; Infit = 1.01, Outfit = 1.01). As regard the facet of previous contact with mentally ill people, the participants who haven't had any form of contact with mentally ill people performed the compatible block much slower (γ = .19, *SE* = .05; Infit = .95, Outfit 0 .97) than those who did have (γ range = [-.05, -.07], *SE* range = [.02, .04]; .95 ≤ Infit/Outfit ≤ 1.02). In the incompatible task a pattern of mixed measures was recovered: the extreme groups (no contact at all and high previous experience with mental illness) were equally slower in the incompatible condition (Mental Illness/Positive versus Physical Illness/Negative).

The bias/interaction analysis for the two-way interaction of stimuli × personal experience did not evidenced any difference in the compatible sorting condition, whereas it did find the word *bad* to be categorized quicker in the incompatible block by those who had psychological problems in their life ($t_{(352)}$ = -2.09, p = .037). The interaction analysis of stimuli × previous contact evidenced one stimulus presenting a DSF in the compatible block, i.e, *marvellous*, which was sorted faster by those who had a greater contact with mentally ill people, when compared to the other groups (*t*s = [-2, 2.96], *df*s = [82, 265], *p*s < .05). The same interaction analysis was carried out for the incompatible condition and evidenced a DSF for the word *depression* related to those who haven't ever had any experience with affected people when compared to those who did have, with the first group categorizing the stimulus more slowly than the others in the Mental Illness/Positive versus Physical Illness/Negative sorting condition (*t*s = [1.87, 2.90], *df*s = [86, 197], *p*s < .05).

3.5 Discussion

The present study involved the psychometric investigation of two implicit measures of automatic semantic and evaluative associations with the mental illness concept, relative to the physical illness, via the use of an IAT procedure. The first indirect measure targeted semantic associations of mental illness with the psychosocial and biogenetic domains, following the hypothesis about the activation of 'implicit' causal attributions when presented with cues related to the mental illness and to the realms. The main hypothesis was that whenever one holds in his/her associative network representations of mental illness connected to psycho-social or bio-genetic causes, (s)he should be quicker in sorting cues in the categories pairing congruent (e.g., Mental Illness/Psychological or Mental Illness/Biologic) with his/her implicit association. The second IAT was designed according to previous studies that had already developed IATs tapping on automatic positive and negative evaluative associations to mental illness, as indirect measures of a presumed negative attitude towards people with a mental disease (Menatti., Smyth, Nosek, & Teachman, 2012; Rüsch, Corrigan, Todd, & Bodenhausen, 2010).

The two implicit measures were then analysed within a Rasch modelling framework to open a window on the inherent functioning of the measures, by decomposing the well-known IAT effect into its main components or, metaphorically speaking, the 'ingredients' that makes the IAT an implicit measure of automatic associations.

With regard to the IAT on the associations of psychological or biologic aspects to mental illness, the MFRM evidenced the following pattern of results:

1) The MFRM retrieved a common underlying measurement dimension wherein all of the 22 IAT stimuli were located and ordered according to their latency parameter estimates, which describe the stimuli ease of categorization into the categories they belong to.
2) The analysis of the differential functioning of stimuli in the two sorting conditions of the task, namely the hypothesized association-congruent pairing of Mental Illness/Psychological versus Physical Illness/Biologic and the association-incongruent reversed pairing, recovered the effect of those stimuli that mostly triggered the IAT effect: the words *tumour*, *depression*, and *psychopathy*, and the image of a *DNA filament*. These stimuli were categorized faster when presented in

the compatible block; hence, they consistently contributed to the emergence of the association between mental illness and psychological attributes. Noteworthy is the fact that the associative effect was partly driven by cues pertaining to the target categories (*tumour* for the Physical Illness category, *depression* and *psychopathy* for the Mental Illness category), suggesting that the presumed association between the mental disease and the psychological domain might not directly root in the associative links generated by the activation of Mental Illness representations in mind, which in turn should activate connected representations of psychological or biologic elements according to the strength of their bonding associative links (e.g., Greenwald et al., 2002; Strack & Deutsch, 2004). Rather, it seems that the reversed occurred. Mental Illness exemplars, such as *depression* and *psychopathy*, and Physical Illness exemplars, such as *tumour*, were categorized faster when the respective categories shared a common response button with the attribute categories of Psychological and Biologic, respectively. Beyond the inherent features of diseases like depression and psychopathy, which might have aroused associations more related to psychosocial aspects, the direction of the association activation appears to be reversed, for any psychological attribute was classified quicker in the compatible condition. A second pattern of results supports this claim, i.e., the different categorization speed of the categories themselves. Instead of observing a differential categorization effect of targets and attributes pairs *between* the two tasks – just as it was hypothesized – the four categories were classified with the same speed in both critical blocks. Rather, when the ease of categorization of the four categories was compared *within* each block, only in the compatible pairing the Biologic category was sorted faster than the Psychological one, likewise the Mental Illness category was categorized quicker than the Physical Illness. This pattern of results points to what actually may drive the implicit association of mental illness with the psychological realm: apparently, this does not result from the attribution of psychosocial elements to the mental illness concept; rather, it depends on refraining from associating biological aspects.

A similar pattern was found for a measure of implicit preference for white people over black people with a typical racial IAT: it was found that the preference for white people displayed by white individuals especially resulted from the attribution of positive traits to Whites, rather than of negative traits to Blacks, for the stimuli that contributed most to the measure have been the positive ones, rather than the negative ones. Therefore, it was argued that the implicit measure of racial attitude might not necessarily imply black derogation, but could be mostly related to white favoritism (Anselmi et al., 2011). A similar line of reasoning might be applied also in this case, but with the difference that the stimuli that contributed most weren't *attribute* stimuli but mostly *target* stimuli (which questions what is the target and what is the attribute) and that the categories that induced most the IAT effect were the Mental Illness target and the Biologic attribute presented in two contrasting pairings (Mental Illness/Psychological versus Physical Illness/Biologic) in the block hypothesized to elicit the enquired association. Therefore, it is arguable that causal associations with mental illness do not imply an association with the psychosocial *or* the biogenetic attributive realm but it is the association or dissociation with the biogenetic domain that play a key-role.

A second potential account of this result could be related to task-recoding strategies (Teige-Mocigemba et al., 2010) applied by the participants during the performance of the task: respondents could have re-coded the task from the classification of four elements into a pooled binary classification of mental illness and biological cues, by focusing on elements of this two categories, actively pairing the stimuli along another salient dimension available at the time, and consequently creating a 'compatibility effect' on the task that was unrelated to the associations between mental illness and psychological versus biologic aspects. This issue could also give an explanation to the differential effect exerted by the procedural norm of counterbalancing the critical blocks, which facilitated the task for the respondents who completed the incompatible task first, instead of adhering to the traditional 'compatibility effect' of making the task easier when the block presented first is the compatible (e.g., Klauer & Mierke; 2005; Teige-Mocigemba et al., 2010; Teige-Mocigemba, Klauer, & Rothermund, 2008). Yet, the participants displayed a general implicit association between mental illness and the psychological semantic aspects, which counters the retrieved easier performance in the incompatible first block.

To probe these possible accounts of the individual triggering effect of mental illness target concept and biologic attribute concepts a feasible strategy is provided by the application of a *Recoding-Free IAT* (RFIAT – Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009), a *Single Category Implicit Association Test* (SCIAT – Karpinski & Steinman, 2006), or a *Brief IAT* (BIAT – Sriram & Greenwald, 2009) for the assessment of distinct, absolute associations towards each category. The latter has already been recently applied in mental illness implicit evaluation research (Rüsch, Corrigan, et al., 2010; Rüsch, Todd, Bodenhausen, & Corrigan, 2010a,b; Rüsch, Todd, Bodenhausen, Olschewski, & Corrigan, 2010).

3) A related result, worthy to be mentioned, cover the retrieval of three stimuli that mostly contributed to *decreasing* the IAT effect, by being processed more quickly in the incompatible block, relative to the compatible: the word *flu* and the pictures of a *family* and of a *mother/child relationship*. Unluckily, the decreasing effect of the word *flu* depends on the Italian double meaning of the word (influenza): the word means a physical disease but also influence. When the word was sorted out in the Physical Illness/Psychological pairing the misunderstanding came out; and the MFRM meticulously found it.
The two images that decreased the IAT effect are two psychological stimuli that still may have been connoted by a biological foundation: both the family and the relationship between mother and child have a biological genesis, in their more conservative and conceptual representation. Hence, the pairing with physical illness could have been activated by associative links with these two pictures. Or, the re-coding strategies could have been applied in this case as well, by focusing on a different classification dimensions besides the enquired associations. Whether or not this was the case, the MFRM signaled the differential functioning of these stimuli that contradict the expectations. The application of different implicit measures, such as those listed at point 3), can be useful to check it out.

The application of the MFRM to the IAT on the automatic evaluative associations towards mental illness, evidenced similar patterns of results:

1) The MFRM retrieved a common underlying measurement dimension wherein all of the 20 IAT stimuli were located and ordered according to their latency

parameter estimates, which describe the stimuli ease of categorization into the categories they belong to.

2) The contrasts between the stimuli local measures of their ease of categorization between the hypothesized association-congruent sorting condition (i.e., Mental Illness/Negative versus Physical Illness/Positive) and the association-incongruent pairing (i.e., Physical Illness/Negative versus Mental Illness/Positive), recovered the positive contribution of five stimuli to elicit the IAT effect: four stimuli were exemplars of the target categories (*depression*, *paranoia*, *schizophrenia*, and *diabetes*) and one stimulus was a negative attribute (*sad*). Also in the case of the attitude IAT the stimuli that mostly triggered the associative effect were target stimuli and not attributes. That suggests the effect of the specific exemplars used to represent the mental illness concept, which are then sensitive to the activation of different associations probably according to specific features. Noteworthy is the recurrence of *depression* and *psychopathy* stimuli, in addition to *schizophrenia* (which is a highly representative mental disease in everyone's mind and denoted by strong negative attitudes at the explicit level as well). It is arguable that automatic negative evaluations of mental illness are not linked to a general evaluation of the mental illness as a broad, overarching category; rather, the specificity of the diagnostic categories drives different reactions. The differential function of which mental illness exemplar is presented is consistent with recent studies that evidenced disorders-specific effects on stigmatizing attitudes behaviours towards people with mental illness (e.g., Angermeyer & Dietrich, 2006; Angermeyer, Holzinger, Carta, & Schomerus, 2011; Boffo & Mannarini, 2012; Feldman & Crandall, 2007).

Similarly to the causal belief IAT, a second pattern of results supporting this claim is the different categorization speed of the categories themselves. Apparently, Mental Illness category was generally processed faster when paired with negative attributes, compared to the Physical Illness category. Further, the results suggest that the effect of the mental illness attitude IAT was consistently elicited by the general tendency to classify positive words quicker than negative within both critical blocks and in the Mental Illness/Positive and Physical Illness/Negative condition – pointing to the potential emergence of a 'positive association primacy' effect (e.g., Anselmi et al., 2011). Once again, the MFRM allowed spelling out the meaning of the implicit measure: participants' average IAT score did not indicate any polarization towards mental illness, but they were somewhat inclined to not associate positive attributes to physical illness, though they not even implicitly display any positive associations towards the mental illness. This result emerged also in the analysis of the possible 'compatibility effect' caused by the IAT counterbalanced blocks order: both critical blocks were similarly affected when presented first, indicating no definite 'preference' (in terms of facilitated performance) for one block over the other.

The impression is that participants manifested a "better a mental disease than a physical illness" attitude, which is inherently trivial. However, one can suppose an underlying ambivalence, which could be resolved by separating the general "illness" effect from the two associative targets and thus index the distinct, absolute automatic evaluations of the concept of mental illness via, for instance, the above-mentioned SCIAT (Karpinski & Steinman, 2006), or through an evaluative priming task (Fazio, Sanbonmatsu, Powell, & Kardes, 1986) with mental illness primes presented subliminally.

Last but not least, the MFRM further proved the sensitivity of the implicit measures in detecting the effect of individual variables that intervene in the task performance and that are susceptible to influence the to-be-measured implicit associations. The two facets for personal experience and prior contact with mental illness displayed a clear influence on the completion of the critical sorting conditions:

a) Respondents that did not exhibit a personal history of psychological problems and/or mental disease were quicker when confronted with pairing mental illness with psychological aspects contrasted to the pairing physical illness/biologic and implicitly negatively evaluated much faster the mental illness concept. The individual personal experience with psychological suffering and/or with being diagnosed with a mental disease appears to temper probable negative evaluations of mentally ill people. Conversely, that seems to prompt the endorsement of biogenetic associations with mental illness. Whether or not this is related to the attribution of personal control over and responsibility for the problem, as stated by the biogenetic approach to mental illness (see Chapter 2), needs to be tested.

b) Respondents who have never had any contact with mentally ill people performed slightly better the categorization of stimuli in the mental illness/psychological sorting block and were rather ambivalent towards mentally ill people, for it was equally difficult for them to associate positive or negative evaluations to mentally ill people, probably for the absence of any previous life direct contact with affected people. Of interest is the fact that people with a considerable experience with mental problems within their family circle (i.e., high prior contact) were more prone to implicitly negatively evaluate mental illness. The counter effect of a steady and pervasive experience with family members presenting psychological problems could then exacerbate people's reactions towards a very negative stance.

Altogether, the emerged result trends do evidence the influence of personal prior experiences with mental illness, both individually and interpersonally, which have been receiving a growing interest in the mental illness stigma for their potential moderating role in the manifestation of stigmatizing reactions towards affected people (e.g., Rüsch, Angermeyer, & Corrigan, 2005).

Although the present study produced interesting findings, some limitations should be noted. First, the two IATs were administered to a group of psychology students at the end of their university course, which could have contributed to the mixed pattern of mental illness associations. The respondents' considerable mental health literacy and strong psychological background are arguable to have influenced the associative network in which the concept of mental illness resides. Furthermore, it is also conceivable that people willing to undertake a mental health caring profession might manifest ambivalent evaluations of mental problems (e.g., Peris, Teachman, & Nosek 2008). In order to refine the predictive properties and sensitivity of the two measures, such analyses should be replicated with a laymen group.

Second, the comparative nature of the Implicit Association Test could have limited the relevance and interpretability of the results, because of the background effect of the simultaneous activation of the broader, overarching representation of 'illness', which could have interfered with the expression of bias towards either type of illness (mental or physical). An indirect measure of distinct and absolute associations with mental illness can probably offer a more precise proxy of the semantic and evaluative structures that are paired in the individual's memory.

Overall, our findings demonstrate the usefulness of considering, besides the usual global IAT scores, the specific contributions of specific stimuli and categories to the expression of implicit associations. The MFRM has proved to be a valid tool for this purpose. The model represents a rigorous frame of reference in which estimating and comparing the speed of categorization of the stimuli. Moreover, by allowing the analysis of differential stimulus functioning, the model unravelled the contribution of each stimulus to the overall IAT measure. Such an analysis is advisable when distinct and opposing drives are involved, as it can be the case of highly prejudicial and discriminatory towards a minority group.

PART 2

IMPLICIT MEASURES IN THE TREATMENT OF MENTAL HEALTH

Chapter 4

Combined Cognitive Bias Modification training in alcohol addict outpatients: for whom is the combination most effective?

Protocol of a Randomized Clinical Trial.

4.1 Introduction

People with an addiction disorder often describe their substance (ab)use as a somehow "unconscious" decision, something that usually happens "by chance" and without any intentional planning or awareness. It seems that they fall to lapse and relapse in the substance consumption almost accidentally, before or even devoid of individual's account for what is happening. The overwhelming irruption of impulsive tendencies, even when people are aware of their condition and willing to tackle it, is a frequent feature in dependence disorders. The paradox between the conscious intention to avoid the substance, because for instance the costs outweigh the benefits associated with continued substance use, and the perpetration of actions towards the substance and/or towards social situations and locations in which the substance is likely to be present, is one of the key-point for the understanding of addiction mechanisms and for the design of effective treatment intervention.

The widespread application of implicit cognition principles and measures in health psychology and experimental clinical psychology suggested the idea that this theoretical and applied research framework may also add something to the field of addiction research (e.g., Roefs et al., 2011; Rooke, Hine, & Thorsteinsson, 2008; MacLeod, 2012; Wier, Teachman, & De Houwer, 2007). Implicit cognitions might be part of the processes leading to addiction. The idea was brilliant, and clinically productive.

This chapter presents the concept, operationalization and implementation of a double-blind Randomized Controlled Trial (RCT) for the experimentation of two new computerized training interventions, i.e. Cognitive Bias Modification (CBM) paradigms, targeting maladaptive impulsive, or *implicit*, cognitive processes in alcohol addiction.

4.1.1 Background

Traditional theoretical and clinical research on the development, maintenance, and treatment of addiction disorders has usually conceptualized them as resulting from deliberate and rational decision-making processes (Wiers & Stacy, 2006a,b), aimed at weighting the pros and cons of a certain behaviour. According to this view, people continue with the substance consumption behaviour as long as the usually short-term benefits prevail over the often long-term, severe, and harmful consequences. The main idea underneath this perspective is that people are rational decision-makers, and that analytical processes should be applied to health-related behaviours as well. However, drug seeking and consumption continue despite of the negative health outcomes and personal and interpersonal consequences. It is well known that although many drug users are completely aware of the detrimental effects of the substance misuse, and further explicit the clear disposition for treatment seeking and compliance to drug abstinence, the risk for lapse and relapse remains extremely high. Such a paradoxical and destructive pattern of behaviour in addiction yields then the reflection on the motives and mechanisms underlying the drug-seeking conduct, even when explicit motivations to quit it are present. Rational and conscious cognitive processes do not solely guide the behaviour, which appears to be also affected by other mechanisms that go beyond individual intentionality. Hence, what are the motives and mechanisms driving these contrasting patterns of behaviour? Why people should continue in engaging in such harmful and dysfunctional behaviours?

Recently, a new theoretical framework was proposed, which posed that implicit, or relatively automatic processes, may provide additional clues in addiction understanding (Stacy & Wiers, 2010), for they may partly drive human behaviour outside the individual's conscious control and "implicitly" affect the outcome of the decision-making process related to a certain conduct. Several *dual-process* models of addiction state the existence of two interacting information processing systems underlying and jointly predicting a behaviour execution: namely, a fast, associative, and impulsive system, which operates through associative links, and emotional and motivational associations; and a slow, relatively controlled, reflective system, which includes the "rational" decision-making and emotion regulation processes described in the earlier developmental models of addiction (e.g., Strack & Deutsch, 2004; Wiers & Stacy, 2006a, 2006b).

According to this perspective, addiction problems can result from an imbalance between strong, impulsive, and automatic reactions to substance-related cues and weak reflective processes and cognitive control. This imbalance between the two operating systems makes then the individual more at risk for being triggered by drug-cues and automatically prompted to the addictive behaviour loop (Bechara, 2005; Wiers, Teachman, & De Houwer, 2007). Furthermore, substance use itself has an impact on the performance of the two systems, by strengthening impulsive reactions to drug-related cues and weaken cognitive and executive control over the impulses (Wiers, Gladwin, Hofmann, Salemink, & Ridderinkhof, 2013).

More specifically, the reflective system includes the individual's ability of taking control over the impulsive system (i.e., cognitive control), the motivation to exert it, the explicit motives driving one's behaviour, and the beliefs and expectancies on the long-term behaviour outcomes (Hofmann, Friese, & Wiers, 2008; Wiers, Houben, Roefs, De Jong, Hofmann, & Stacy, 2010). In the impulsive system, several automatic cognitive processes, or *cognitive biases*, are distinguished, including *attentional* processes (e.g., alcohol addicts usually have an attentional bias towards alcohol-related cues), substance-related *automatic associations* in memory (e.g., alcohol repeatedly and automatically associated to positive or negative evaluations), and automatically triggered *action tendencies* to the substance (e.g., tendency to approach alcohol in heavy drinkers). The reflective system lies on fast and flexible symbolical processes with a limited capacity and related to the working memory capacity, whereas the impulsive system lies on slow learning associative memory processes, which are by nature automatic and difficult to change (Deutsch & Strack, 2006).

Both systems interact with each other in alcohol-related problems and addiction disorder onset. Recent findings in alcohol misuse research indicated that these automatic, or implicit processes, are a better predictor of alcohol use, moderated by the individual's executive function ability of impulse regulation (e.g., Grenard et al., 2008; Houben & Wiers, 2009; Peeters, Wiers, Monshouwer, van de Schoot, Janssen, & Vollebergh, 2012).

The general idea of *dual-process* models of addiction is that when both impulsive and reflective processes, called also implicit and explicit processes, influence addictive behaviours, both classes of processes can be targeted in interventions. Reflective explicit processes are usually the focal target of standard treatment interventions, such as cognitive-behavioural, counselling, and motivation interventions, in which, for instance, the therapist and the patient make an explicit analysis of patient's alcohol use pros and cons, and related motives and expectancies. On the other side, a recent, outstanding body of research is developing new interventions aimed at the treatment and modification of the impulsive and implicit processes, or cognitive biases, involved in addiction (e.g., Fadardi & Cox, 2009; MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002; Schoenmakers, Wiers, Jones, Bruce, & Jansen, 2007; Schoenmakers, De Bruin, Lux, Goertz, Van Kerkhof, & Wiers, 2010; Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011; Wiers, Gladwin, Hofmann, Salemink, & Ridderinkhof, 2013). The Cognitive Bias Modification (CBM) methods are computerized training interventions of these relatively implicit cognitive motivational processes in addiction behaviour, of which patients may not be aware and which are difficult to control and change through standard interventions. CBM methods are implemented in the clinical setting by modifying and adapting to the retraining procedure the same assessment procedures used to evaluate the individual cognitive bias(es), namely, indirect measurement procedures devised in the implicit cognition and implicit memory fields, such as the *Implicit Association Test* (IAT – Greenwald, Mcghee, & Schwartz, 1998) and the *Approach Avoidance Task* (AAT – Rinck & Becker, 2007; Wiers, Rinck, Dictus, & Van den Wildenberg, 2009).

First clinical applications of CBM re-training paradigms add-on to standard CBT interventions did evidence promising results (e.g., Fadardi & Cox, 2009; Wiers et al., 2011; Wiers et al., 2013). Two Randomized Clinical Trials (RCTs) with addicted inpatients did succeed in re-training away the cognitive approach bias towards alcohol stimuli, with a further generalization of training effects outside the experimental procedure context (Wiers et al., 2011; Eberl, Wiers, Pawelczacka, Rinck, Becker, & Lindenmeyer, 2013). Moreover, patients in the training group showed a statistically significant percentage of less relapse one year after discharge, compared to patients in the control group: 13% in Wiers et al. (2011) and about 9% in Eberl et al. (2013). Also an attentional bias re-training paradigm was successful in modifying the triggered alcohol-related stimuli attentional bias in addict inpatients, with a strong avoidance bias at the post-test and significantly longer time of relapse for the experimental group (Schoenmakers et al., 2010).

No studies have yet covered the investigation of the effects of a CBM intervention with alcohol addict outpatients. The substance dependent outpatient is a highly prevalent client category in the public health care system, differently characterized in terms of addiction severity and treatment process, and likely to receive potential benefits from this type of interventions.

Furthermore, no study has yet been published on the potential effects of combining different CBM paradigms, though it is arguable to increase the treatment efficacy.

4.2 The present study: aims and hypotheses

The aim of the current study is to investigate the effectiveness of two computerized CBM retraining paradigms among adult alcohol addicted outpatients: the alcohol attentional bias and approach bias re-training. Participants receive 11 sessions of either the active or placebo version of the two training programs combined with a brief motivational interview prior to intervention, which serves the function of supporting and tracking the training experience.

The perspective of "*what* works best for *whom*?" (Wiers et al., 2013) guides the study main hypotheses. The main goal is to test the main and added effects of the CBM interventions on the remission progress from the alcohol addiction disorder immediately after the intervention and after three months, with changes in the number of lapse or relapse episodes, in the treatment status and in the therapeutic outcome as the primary outcome measures. It is expected that, for each of the two CBM trainings, participants in the intervention condition will show a lower percentage of lapse or relapse and a positive modification of their treatment status than participants in the control condition (e.g., Eberl et al., 2013; Schoenmakers et al., 2007, 2010; Wiers et al., 2011, 2013).

Crossover effects of each CBM paradigm to the other bias are explored, as well as the additive effect of the exposure to the combination of the two CBM trainings. It is expected that each CBM paradigm will decrease or reverse the specific targeted bias and that these changes can possibly mediate the effects on the clinical outcome. Further, it is expected that the joint exposure to both active CBM retrainings will have a greater beneficial effect when compared to the other intervention conditions.

The moderating effect of response inhibition executive function and of the strength of cognitive bias(es) on the CBM training and clinical outcome relation is taken into account. It is expected that participants with strong automatic biases and/or low inhibitory control will benefit more from CBM retraining than participants with weaker biases and/or stronger executive functions, in line with dual-process models of addiction and consistent with previous results (e.g., Eberl et al., 2013; Peeters et al., 2012; Wiers et al., 2011).

The effect of several independent clinical variables (e.g., age, duration of the addiction disorder, previous detoxifications) on the primary and secondary clinical outcomes will be further explored, in particular the type of standard treatments participants are undergoing (medication intake and/or other psychotherapeutic interventions).

4.3 Methods

4.3.1 Participants and Procedure

Participants are at least 120 adult outpatients with main diagnosis of alcohol addiction disorder, recruited in the public health addiction service of San Donà di Piave (VE), Italy (Servizio per le Dipendenze, ULSS10).

Participants are screened for eligibility according to the following criteria:

- Inclusion criteria: adult outpatients with primary diagnosis of alcohol addiction disorder according to DSM-IV-TR diagnostic criteria, alcohol abstinence for at least two months.
- Exclusion criteria: neuro-cognitive problems, visual or hand-motoric handicaps, severe neurological disorders (e.g., Korsakoff syndrome), comorbidity with psychotic disorders, low fluency in the Italian language.

Participants are recruited by the clinicians according to the inclusion and exclusion criteria and invited to participate to the study. The refereed clinician will also supervise patients' activity and progress along both the entire standard treatment and the experimental intervention.

At invitation, the refereed clinicians provide the patients a brief introduction to the study, explaining that addiction disorders are partly due to uncontrolled and automatic processes which can substantially increase the risk for relapse, and that the main objective of the research is to test the effectiveness of new treatment interventions, which can help the patient in gaining and increasing control over these underlying mechanisms (see Appendix A for the study presentation sheet). Each participant will receive the same general information about the study, to avoid suspects between patients and give standardized information to the participants.

Once informed the patient about the research objectives, the norms for confidential data treatment and for participating in the study (see Appendix A), interested participants can complete the informed consent procedure and create a research account at the training website after which they can read an extensive description of the trial.[6]

Participants will participate to a total of 15 sessions: two baseline measurement sessions, 11 training sessions, a post-intervention measurement session, and a follow-up measurement session after 3 months. Participants can arrange a flexible calendar for their intervention sessions during their regular visits at the public addiction service.

The pre-intervention stage is divided in two sessions, during which a demographics questionnaire, several baseline clinical measures, the two cognitive bias assessment tasks, and two computerized tasks for the assessment of executive functions and alcohol implicit associations, respectively, are administered in fixed order. Once the baseline assessment is complete, participants can start their first CBM training session. Participants have five days to complete each session (two sessions per week), allowing them to complete the 11 training sessions and the post-intervention session in about 6 weeks. A follow-up assessment is conducted at three months after the intervention.

Each CBM session consists of a first part of brief motivational interview (about 15 minutes), after which the motivation to training is briefly assessed, and a second part in which participants complete the two retraining tasks (about 15 minutes each). During the interview, participant and researcher principally focus their attention on the experimental intervention experience and on the related

[6] www.test.uva.nl/lotus/toptraining_serd/registration

feelings and thoughts. The interview has the main objective of reviewing the previous training session and the related perceptions, of introducing the incoming session objectives (decreasing error rates and/or increasing response speed), and of renewing and strengthening participants' motivation in performing the upcoming session. The interview follows a semi-structured protocol for a brief motivational interview, in order to ensure a standard setting to all participants (§4.4).

The study was approved by the Ethics Committee of the School of Psychology of the University of Padova (February 2013; Pr. 1242) and registered at Current Control Trials (ISRCTN01005959).

4.3.2 Trial Design

A 2x2 factorial design is adopted to study the effectiveness of attentional bias retraining and approach bias retraining. This design allows exploring possible additive and multiplicative effects of the combination of the two CBM re-trainings, as well as to what extent one retraining will produce changes in the other automatic bias. The placebo and real training versions for each cognitive bias retraining are then matched into four experimental conditions (see Table 4.1): one double re-training experimental group, two experimental groups receiving one re-training and one placebo, and one double-placebo control group. According to the experimental design, the probability of receiving at least one real re-training intervention reaches the 75%. Participants will receive either the active or the placebo version of both the CBM interventions.

Table 4.1 Experimental manipulation design

	Attentional$_r$	**Attentional$_p$**
Approach$_r$	Approach$_r$ -Attentional$_r$	Approach$_r$ -Attentional$_p$
Approach$_P$	Approach$_p$ -Attentional$_t$	Approach$_p$ -Attentional$_p$

$_r$ training version; $_p$ placebo version.

Prior to training, participants complete a pre-treatment demographics questionnaire and a baseline clinical assessment. Interventions effects on the therapeutic outcomes will be then tested directly after the intervention and 3 months later. Participants complete 11 sessions of training with a between-session time-interval of maximum 5 days. The post-intervention assessment takes place between the 10th and 11th training session during a 'masked' session (participants do not know they are starting the post-test evaluation). That is to avoid possible negative feeling related to the intervention final "evaluation" and minimize self-presentation biases and/or preparatory strategies.

The expected timeline for trial completion is expected to be July 2014. Participant flowchart (as per CONSORT statement – Schulz, Altman, Moher, 2010) is presented in *Figure 4.1*.

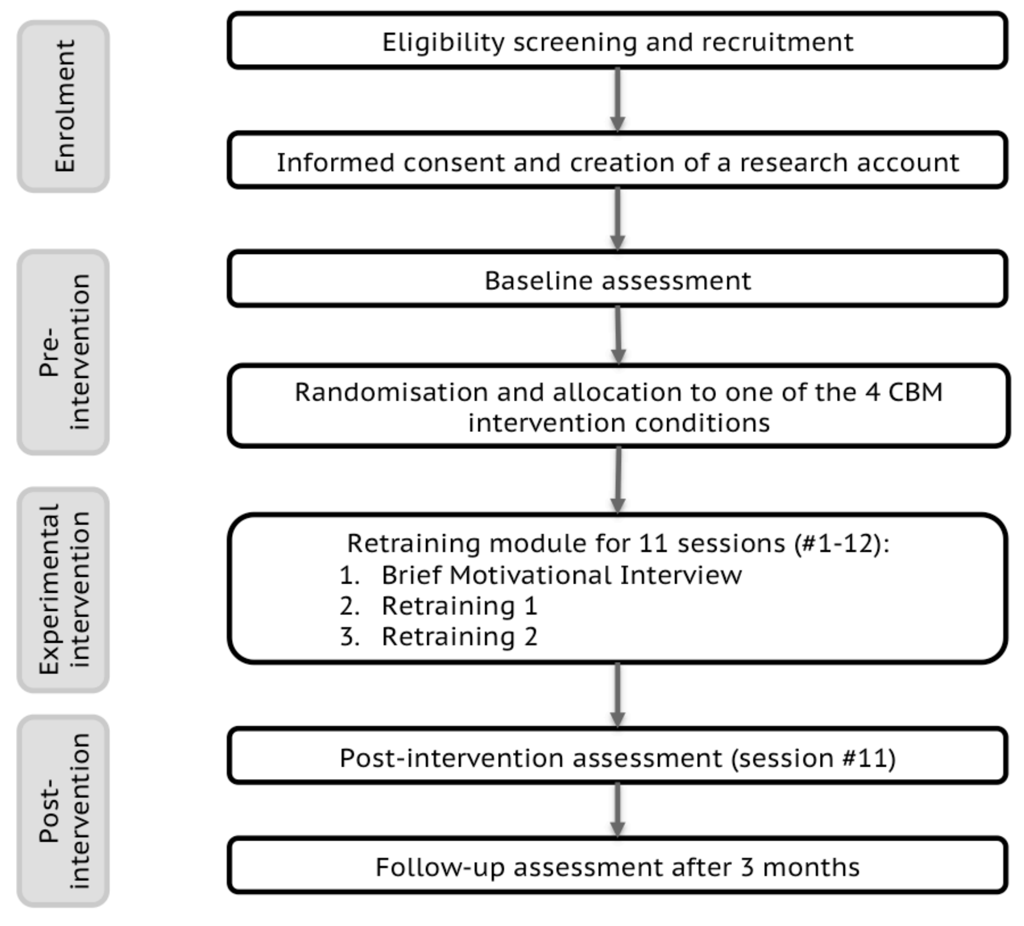


Figure 4.1 Participant flowchart.

4.3.3 CBM Interventions

Each CBM intervention session consists of two tasks: an attentional bias retraining and an approach bias retraining. Each task – both in the active and placebo version – consists of three phases: a brief practice block, an assessment block, and a CBM block. The assessment block serves the purpose of measuring the strength of the bias at the start of every session and tracking any change in the cognitive bias as a result of the CBM training. The practice block presents neutral stimuli (grey geometrical pictures) to practice the task instructions.

In both tasks, each trial starts with a fixation cross in the middle of the screen for a duration randomly picked in the interval $U([500, 1000])$ ms, uniformly distributed. This setting was designed to make the task less boring, to keep the participants' attention focused and to avoid anticipatory responses. Whenever a wrong response is made, a red cross appears on the screen and the same trial is re-started to allow for correction.

A large set of pictures of alcoholic drinks and non-alcohol drinks was created specifically for this study (§0), to be used in all versions of each task (i.e., assessment, training, and placebo).

The two tasks were designed to be as similar as possible (e.g., same stimuli and number of trials) to avoid the effect of any confounding variable on participants' performance. Their order of presentation is counterbalanced between subjects and fixed within subjects. The same tasks order applies to the task assessment versions at the pre-intervention, post-intervention, and follow-up measurement points.

4.3.3.1 Attentional bias retraining

Attentional bias is assessed and trained through the *Visual-Probe Task* (VPT – MacLeod, Matthews, & Tata, 1986; MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002; Schoenmakers et al., 2007, 2010; van Deursen et al., 2013). The VPT is a computerized speeded reaction-time task in which participants are asked to respond to probes located in two different positions on the computer screen (i.e., irrelevant-feature implicit measure). During the task, a picture of an alcoholic drink and a picture of non-alcoholic drink are presented next to each other on the screen for 500 ms. After the stimuli presentation, a small arrow (8.3% of the width/height of the picture) pointing upwards of downwards replaces one of the two pictures – measuring speeded detection of alcohol-related stimuli (*attention engagement*) –, or is positioned on top of one of the pictures – measuring the difficulty to disengage from alcohol-related stimuli (*attention disengagement*). Participants are instructed to respond as fast as possible to the direction of the arrow, by pressing the corresponding key on the keyboard (U and N) (for an example of a trial, see *Figure*

4.2). Response window is set to 4000 ms; in case of no response the trial is restarted after repeating the task instructions.

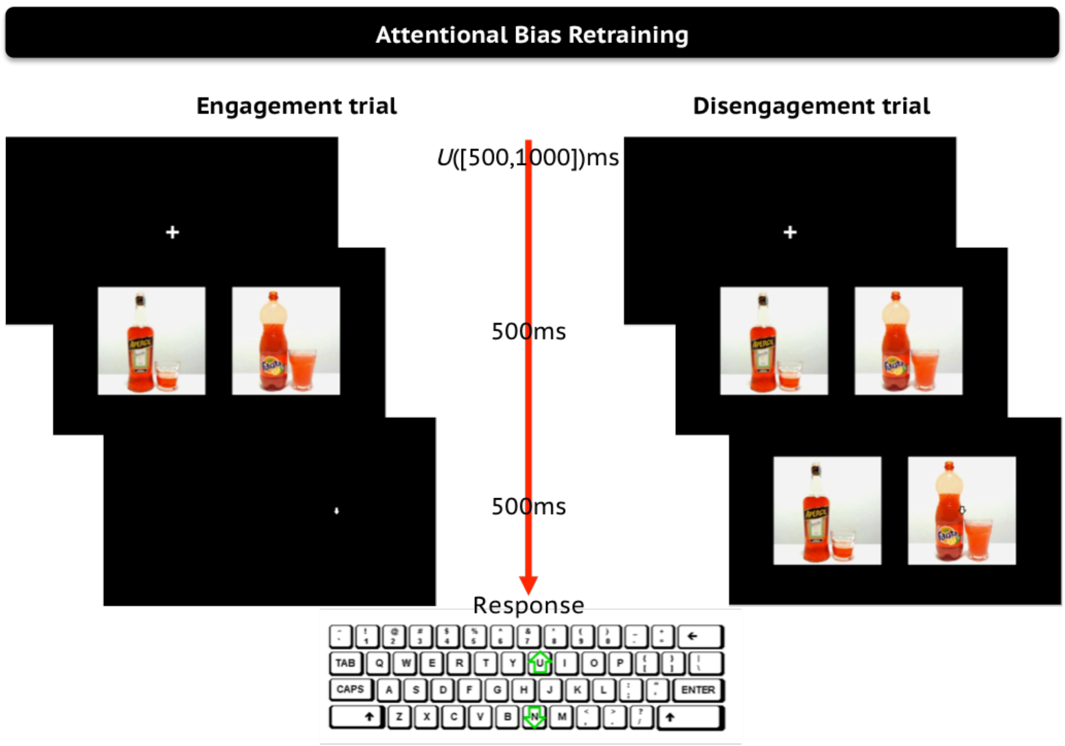


Figure 4.2 Example of two trials in the attentional bias retraining.

In the assessment version of the task and in the assessment block of the CBM and placebo versions, the arrow replaces the picture of alcoholics (alcohol trials) and non-alcoholics (non-alcohol trials) equally often. Attentional bias is computed by subtracting response times (RTs) on alcohol trials from those on non-alcohol trials separately for the two formats of arrow presentation. In the CBM block, participants in the experimental condition are trained to direct their attention away from alcoholic drinks towards non-alcoholic drinks by exposing them only to non-alcohol trials, whereas participants in the placebo condition receive 50% alcohol and 50% non-alcohol trials (as in the task assessment version and in the assessment block).

The task structure in both versions (assessment and training) is presented in Table 4.2. Stimuli are pairs of matched alcohol/non-alcohol pictures (§0), which are counterbalanced for task settings with a 2x2x2 design in the assessment stage (stimuli presented on the left and on the right, formats of arrow presentation, and arrow location at the alcohol or at the non-alcohol picture), and with a 2x2 design in the CBM block (stimuli presented on the left and on the right and formats of arrow presentation). The probe direction is set randomly upwards or downwards.

Table 4.2 Task structure for the attentional bias assessment and retraining: stimuli, number of trial formats and repetitions, and total number of trials for each task block.

Task version	Blocks	Stimuli	No. of stimuli	No. Trial format	Trial reps	No. Trials
Assessment	Practice	Neutral	1	8	1	8
	Test 1	Alcohol/Non-alcohol	20	8	1	160
	Test 2	Alcohol/Non alcohol	20	8	1	160
						328
CBM	Practice	Neutral	1	8	1	8
	Assessment	Alcohol/Non-alcohol	8	8	1	64
	Training	Alcohol/Non-alcohol	12	4	2	96
	Training	Alcohol/Non-alcohol	12	4	2	96
						264
Placebo	Practice	Neutral	1	8	1	8
	Assessment	Alcohol/Non-alcohol	8	8	1	64
	Placebo	Alcohol/Non-alcohol	12	8	1	96
	Placebo	Alcohol/Non-alcohol	12	8	1	96
						264

Note: In the VPT, the stimuli presented on screen are pairs of pictures. The column No. of stimuli refers to the number of pairs used.

4.3.3.2 Approach bias retraining

Alcohol automatic approach tendencies are assessed and trained with the modified Approach-Avoidance Task (AAT – Eberl et al., 2012; Wiers et al., 2009, 2011; Wiers, Rinck, Kordts, Houben, & Strack, 2010). The AAT is a computerized speeded reaction-time task in which participants are asked to react to stimuli presentation format and ignore stimuli content (i.e., irrelevant feature implicit measure).

In this task, a picture of an alcoholic or non-alcoholic beverage is presented in the centre of the screen. The picture is three degrees tilted to the left or to the right. Participants are instructed to respond to the tilt direction of the picture, by pushing pictures tilted to the left away from them and pulling pictures tilted to the right towards them. The combination of the format of the picture and the response (left = push and right = pull, versus left = pull and right = push) is counterbalanced across participants. Participants' response comes along with a zooming effect, which increases picture size in the pulling closer response and decreases it in the pushing away response (for an example of a trial, see *Figure 4.3*). Stimulus stays on screen for 3000 ms; in case of no response the trial is restarted after repeating the instruction.

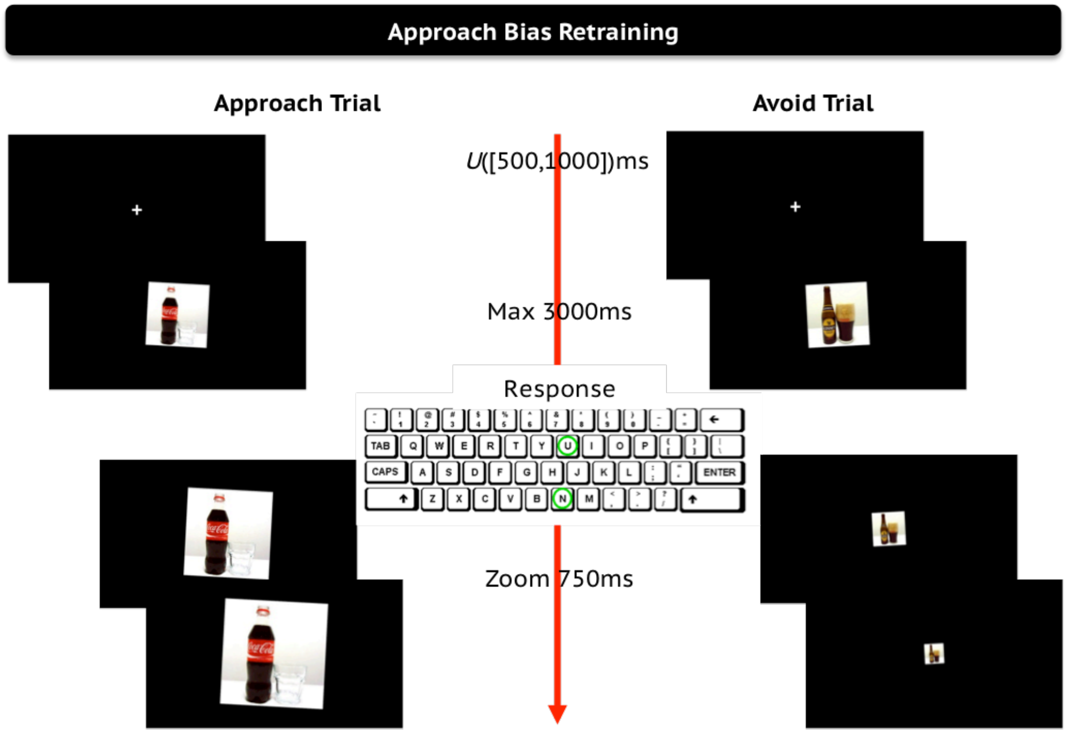


Figure 4.3 Example of two trials in the approach bias retraining.

In the assessment version of the task and in the assessment block of the CBM and placebo versions, the pictures of alcoholics and non-alcoholics are presented equally often in both formats. Approach bias for alcohol is computed by comparing RTs for push, pull, alcohol and non-alcohol trials (alcohol/push –alcohol/pull) and (non-alcohol/push – non-alcohol/pull). In the CBM block, participants in the experimental condition are trained to avoid alcohol by exposing them only to alcohol/push and non-alcohol/pull trials, whereas for participants in the placebo condition alcoholics and non-alcoholic are equally presented in both formats.

The task structure in both versions (assessment and training) is presented in Table 4.3. Stimuli are pairs of matched of alcohol and non-alcohol pictures (§0), which are counterbalanced for presentation format only in the assessment stage.

Table 4.3 Task structure for the approach bias assessment and retraining: stimuli, number of trial formats and repetitions, and total number of trials for each task block.

Task version	Blocks	Stimuli	N of stimuli	N Trial format	Trial reps	No. Trials
Assessment	Practice	Neutral	2	2	2	8
	Test 1	Alcohol/Non-alcohol	40	2	1	80
	Test 2	Alcohol/Non alcohol	40	2	1	80
						172
CBM	Practice	Neutral	2	2	2	8
	Assessment	Alcohol/Non-alcohol	8	2	2	64
	Training	Alcohol/Non-alcohol	12	1	4	96
	Training	Alcohol/Non-alcohol	12	1	4	96
						264
Placebo	Practice	Neutral	2	2	2	8
	Assessment	Alcohol/Non-alcohol	8	2	2	64
	Placebo	Alcohol/Non-alcohol	12	2	2	96
	Placebo	Alcohol/Non-alcohol	12	2	2	96
						264

Note: In the AAT, the stimuli presented on screen are single pictures. The column No. of stimuli refers to the number of pictures used.

4.3.4Tasks stimuli

The pictures used in the AAT and VPT tasks were developed specifically for this study according to a stimuli-recording protocol, which was designed similarly to existing CBM stimuli (van Deursen et al., 2013).

The pool of stimuli was designed to have 144 pairs of alcohol and non-alcohol pictures matched by structural, visual, and pictorial features and photographed in both static (beverage only) and dynamic (presence of a human in interaction with the drink) contexts. Alcohol pictures depict common wine, beer, and spirits brands in Italy (such as Chianti red wine, Prosecco white wine, Moretti beer, Montenegro liquor, etc.), eight brands per category, highly familiar and easily recognizable. A common non-alcohol drink was selected for each alcohol beverage by matching as much as possible the type of packaging (bottle, can, jar, carton), packaging size, and colour.

Drinks were then photographed in a neutral setting (windowless room with a table on a white background, full illumination on the centre of the table, various glasses for the different drinks, a tray, and a bottle opener) and according to the following criteria: drinks in the foreground of the picture, consistent framing to shoot pictures from the same angle, and use of a standard digital camera (photo size 500x500 pixel, no flash, saturation in manual setting).

For each context three scenarios were put on (static: open beverage only, open beverage with empty glass, open beverage with full glass; dynamic: woman serving the open drink on a tray, woman/man opening the drink, woman/man drinking). In the two woman/man dynamic scenarios, alcohol drinks – as well as their matched non-alcohol drinks – were counterbalanced for drink category (wine, beer, and spirits) and gender; whereas in the three static scenarios each picture was shot in each of them.

Stimuli were then processed in Photoshop to adjust for size, exposure, brightness, contrast, and to correct minor image imperfections (see Table 4.3 for some examples of stimuli).

4.3.5 Baseline measures

Socio-demographics information (gender, birthdate, annual income, educational level) and clinical case history details (duration of alcohol addiction, previous detoxifications and treatments, duration of current abstinence, medication intake) are collected during participants' research registration. In the first baseline assessment session, other substances use (integration of CORE *Alcohol and Drug Abuse Survey* – CORE Institute, http://core.siu.edu/; and IPSAD *Italian Population Survey on Alcohol and other Drugs* questionnaire – National observatory for Drug Use, www.epid.ifc.cnr.it./), self-esteem (*Rosenberg Self-Esteem Scale,* RSES – Rosenberg, 1965), anxiety (*State-Trait Anxiety Inventory-Y,* STAI-Y – Spielberger, 1989), and depressive symptoms (*Beck Depression Inventory-II,* BDI-II – Beck, Steer, & Brown, 1996; Italian version by Ghisi, Flebus, Montano, Sanavio, & Sica, 2006), are evaluated. After the questionnaires participants performed the alcohol-related approach and attentional cognitive bias assessment tasks.

In the second baseline assessment session, the alcohol abuse (*Alcohol Use Disorders Identification Test*, AUDIT – Saunders, Aasland, Babor, & Grant, 1993; Babor, Higgins-Biddle, & Saunders, 2001; Italian version by Piccinelli et al., 1997), craving (*Obsessive-Compulsive Drinking Scale*, OCDS – Anton, 2000; Italian version by Janiri et al., 2004), and motivation to treatment (MAC2-A – Spiller, Zavan, & Guelfi, 2006) are evaluated.

Table 4.3 Examples of task stimuli photographed in the six scenarios (three static and three dynamic): target and matched control stimuli for each alcohol category.

		WINE		BEER		SPIRITS	
Context	**Scenario**	***Target***	***Control***	***Target***	***Control***	***Target***	***Control***
Static	Drink only						
	Drink with empty glass						
	Drink with full glass						
Dynamic	Serving						
	Opening						
	Drinking						

After the questionnaires, a computerized version of the classical Stroop task (Stroop, 1934) is used to assess response inhibition executive function (e.g., Eberl et al., 2013; Peeters et al., 2012; Stetter, Chaluppa, Ackermann, & Straube, 1994). In this task, participants have to classify words and symbols according to their ink colour and ignore the content. The task starts with a practice block, in which participants have to learn the correct key-colour combination (only neutral and incongruent trials are presented). The second block consists of a second practice block task with grey key reminders on the bottom of the screen. The third block is a test task composed of 112 trials in which the key reminders disappear and 16 neutral trials (#### in blue), 48 congruent trials (red in red), and 48 incongruent trials (red in yellow) are presented.

The second assessment session ends with a *Brief Implicit Association Task* (BIAT – Sriram & Greenwald, 2009) measuring the strength of approach/avoidance associations with alcohol (e.g., Wiers et al., 2011; Wiers van de Luitgaarden, van den Wildenberg, & Smulders, 2005). In the BIAT, participants are required to choose whether word stimuli presented in the centre of the screen belongs or not to one or two focal categories on top of the screen, by pressing the 'yes' and 'no' corresponding keys (E and I). In the first block (16 trials) participants practice the task by classifying words for alcoholics (wine, beer, vodka, rum), non-alcoholics (pepsi, milk, water, tee), mammals (horse, sheep, cat, elephant) and birds (swallow, eagle, hawk, pigeon), as belonging to alcohol or mammals (focal categories) or not ('anything else'). In the subsequent four blocks (20 trials each), the alcohol focal category is alternatively paired with approach (block 2 and 4) or avoid (block 3 and 5) attribute category. Test attribute stimuli for approach (grab, approach, closeness, touch) and avoidance (flee, push, avoidance, elude) have been adapted from Wiers et al. (2011) and Ostafai and Palfai (2006). The order of the combined blocks for the alcohol/[attribute category] pairings within the BIAT and the contingency between the response and the assigned key (E and I), are counterbalanced across participants. The outcome measure is computed as the standardized difference in latencies between the different combined blocks (modified D-score algorithm – Nosek, Bar-Anan, Sriram, & Greenwald, 2013). As a control measure, participants subsequently rated BIAT stimuli on valence with a Visual Analogue Scale from 0 (*extremely negative*) to 10 (*extremely positive*).

An overview of all measurement instruments along the trial measurement time-points is presented in Table 4.4.

Table 4.4 Measurement instruments: measurement domain and time-points.

Domain	**Measure**	**Baseline**	**Training**	**Post-intervention**	**Follow-up**
Cognitive bias assessment	• Visual-Probe Task • Approach-Avoidance Task	✓	✓	✓	✓
Generalization of training effects	• Brief IAT	✓		✓	
Executive Function	• Stroop Task	✓		✓	
Baseline measures	• Demographics • Case-history details • RSES	✓			
Primary outcome	• CORE Alcohol/Drug use[a] • Treatment status • Clinical status (lapse/relapse)	✓		✓	✓
Secondary outcome	• AUDIT[a] • OCDS • STAI-Y • BDI-II	✓		✓	✓
Motivation to treatment	• MAC2-A • Motivation to training	✓	✓[b]	✓	

[a] *The questionnaire refers to the last 12, 1, and 3 months in the baseline, post-intervention, and follow-up measurement points, respectively.*
[b] *Motivation to training questionnaire only.*

4.3.6 Primary and secondary outcome measures

The main outcome measure is the change in the participants' clinical status, as assessed by the presence of any lapse or relapse during the three months after the intervention, and in their treatment status (medication intake, other form of therapeutic interventions). In addition to participants' self-reported clinical status, the refereed clinicians will provide an evaluation of the patients' therapeutic progress (successful or not).

Secondary outcome measures include changes in the automatic cognitive biases as assessed with the attentional bias and approach bias tasks (§4.3.3) at post-intervention and after three months, by using the same stimuli of the post-intervention session, to check for the duration of the training effects.

Generalisation effects of the two retrainings to other measures are assessed at post-intervention with the BIAT (e.g., van Deursen et al., 2013; Wiers et al., 2010; 2011) and the Stroop task. Other secondary outcome measures (assessed at each measurement point) also include other substances abuse (CORE questionnaire), alcohol-related problems (AUDIT), craving (OCDS), anxiety (STAI-Y), and depression symptoms (BDI-II).

Intervention credibility and expectancies are also assessed with the *Credibility/Expectancy Questionnaire* (CEQ – Devilly & Borkovec, 2000), to evaluate the general participants' experience with the study.

4.37 Randomisation

Participants meeting the inclusion and exclusion criteria will be automatically assigned at the pre-intervention stage to one of the four experimental conditions with equal likelihood and stratified by gender and category of medication intake, as specified here under:

- Category A (alcohol agonists and antagonists with side effects on attention and executive functions): Disulfiram (Antabuse®), Naltrexone (Revia®, Depade®, or Vivitrol®), Acamprosate (Campral®), GHB (Alcover®);
- Category B (psychoactive medication with slight side effects on attention and executive functions, like anxyolitic, antidepressant, and neuroleptic medications);
- Category C (any medication not effecting attention and executive functions).

Participants will be randomly allocated to one of the conditions to which the fewest participants of their gender and medication category have been so far assigned.

Participants will be excluded if (a) they do not complete the baseline assessment, or (b) if they disclose the intention to discontinue the study, or (c) if they present an episode of relapse during the experimental intervention.[7]

4.3.8 Blinding

The trial has a double-blind design; hence, both participants and researchers do not know which experimental condition the participant is assigned to. Randomisation of participants is completely automatized and implemented in the experiment delivery online platform. In order to keep participants blind to which intervention they receive, they are required to respond to an irrelevant feature in both CBM training paradigms (e.g., the orientation of the picture) instead of reacting to the content of picture (alcoholic or non-alcoholic drinks) (e.g., Eberl et al., 2013; van Deursen et al., 2013; Wiers et al., 2011). Participants' awareness about which experimental condition they are assigned to is assessed at the follow-up measurement point.

4.4 Brief Motivational Interview

According to the approach of the Transtheoretical Model of Behavior Change (TTM – e.g., Prochaska, DiClemente, & Norcross, 1992; Prochaska, Norcross, & DiClemente, 1994), which assumes that the changing process in health behaviour treatment is composed of six stages of *readiness*, from the pre-contemplation (avoidance and denial of a problem) to the maintenance (maintaining the successful changes into the daily life) stage, the participants of the present study are supposed to be in the fourth action stage of changing. This stage is characterised by the pursuit of concrete decisions and behaviours aimed to tackle the addiction problem and to change the status quo, and by the establishment of an intentional commitment to the treatment process.

At the beginning of each retraining session participants take part in a brief interview with a trained researcher (about 15 minutes), aimed at introducing the participant to the upcoming experimental session. The brief motivational interviewing protocol here devised is based on the Motivational Interviewing

[7] A single episode of lapse shorter that three days and ended by the patient without any further negative consequence, is admitted.

approach (MI – Miller & Rollnick, 1991/2002), developed on the wake of Prochaska et al.'s TTM approach.

The main objective of a MI during the Action stage is to support the person's changes and progress reached so far in a constructive and open-minded way, by explicitly sustaining the efforts (s)he is making to concretely face the addiction problems. In this context the efforts the participants are making include the participation in the clinical trial and in the alcohol-related cognitive biases re-training sessions. The brief interview is then devised to prepare and introduce the participant to the upcoming experimental session, to rehearse the objectives of the experimental treatment intervention, to empower the motivation to engage in the therapeutic process, and to bring back the patient's attention to the potential benefits of this kind of retrainings. Namely, the strategies adopted in this phase are: 1) reviewing progress, 2) renewing motivation, and 3) redoing commitment (Centre for Substance Abuse Treatment, 1999).

In particular, the brief MI interview here devised should cover the following topics:

- Review of the preceding session (except the first retraining session) and positive feedback. The interview should start by asking for the permission to talk about the previous session (respectful and free attitude): how did the participant perceive the last session? Was s(he) satisfied with his/her performance? In which of the two tasks, or in which part of the session, does (s)he think (s)he was better or more successful? This approach serves the co-structuring of positive framed feedback on the progress so far.
- Renew and support the motivation and compliance to the experimental intervention: according to the participant's report of the previous session, the interview proceeds by shifting the attention to the motives that led the participants to start the change process, to support and/or empower them, and remembering the objectives of this training. The joint rehearse of the motivations that brought the patient to undergo a treatment intervention (e.g., I don't want my loved ones to be ashamed of me; I don't want my children to grow up with an alcohol addicted father/mother; I messed up my life because of the drinking; I don't want to feel as a loser anymore) should work as a fuel for continuing the intervention.

- Empowerment of the self-efficacy and affirmation of the current changing progresses: the retraining interventions are one of the concrete actions pursued by the participant to deal with the alcohol abuse. It is then important to explicitly acknowledge the progress (precision, constancy, and commitment to the tasks) and to reinforce it (e.g., practice effect as a sign of individual efficacy in performing the tasks, increase control over the performance).
- Normalization of any difficulty encountered in the retraining execution and reaffirmation of the commitment (e.g., metaphor of the gym training). Difficulty and boredom are common experiences that are intrinsically part of these kinds of re-training. In particular, the repetition of a certain behavioural pattern is a key-component when the objective is to strengthen ability, or a muscle for example, or to increase the expertise in some life domains, such as a job activity. The gym metaphor is useful in reframing and re-evaluating this topic, in reaffirming the commitment to the practice of a new activity to gain more control and expertise (such as at work). This strategy is functional to the reinforcement of the participant's sense of autonomy and ability to carry out self-chosen goals and plans.
- Collaborative negotiation of the next session goals: what are the expectations for the next session? What is the goal the participant would like to reach? For example the proximal goal of increasing the response speed or reducing the number of errors sounds like a challenge for some participants and consequently stimulating their active involvement in the task.

As a general, rough guideline, the brief MI should start with the open discussion about the previous session; recover of the main objectives and motives related to the participation in the clinical trial; carefully listen to the individual perceptions of the tasks and normalize the experience; stimulate, if and when needed, and generally support the feelings of self-efficacy and self-confidence in actively performing this new "exercises"; reinforce the commitment and the efforts in each session; keep the participants' attention to the concrete advantages and motives that are guiding their therapeutic progress. Generally speaking, the interview should be based on a concrete level of interaction, client-centred and focused on the introduction and practical discussion of the cognitive biases re-training sessions, by avoiding in the meantime the discussion of personal feelings and experiences related to the individual case study and relevant for the patient's therapeutic path. These will be acknowledged but still re-addressed to the standard psychotherapeutic setting.

According to the TTM perspective, the change processes involved in the brief MI here devised deal with behavioural processes, namely, the self-efficacy and management reinforcement, the stimuli control, and the support relationship, leaving the in depth involvement of the experiential processes to the individual clinical setting.

Neither the participant, nor the researcher knows in which experimental condition the participant is assigned to. If the participants enquiry about their intervention condition, the researcher will honestly disclose his/her own unawareness of the participants' allocation and give a feedback on the possible moderate effects of the retraining paradigms even in the placebo condition, as emerged in previous studies (e.g., Wiers et al., 2011).

4.5 Analyses

The analyses strategy splits the analytical process in two stages. The first stage considers the psychometric investigation of the two implicit cognitive biases measures. A preliminary trial data analyses, limited to the consideration of the two cognitive bias assessment measures and training experimental manipulation, will be conducted following a two-fold objective. Firstly, the lack of measurement validity study on the AAT and VPT tasks warrants the investigation of their actual measurement properties, to establish whether they are measuring what they are intended to measure. Secondly, the adaptation of these cognitive bias assessment tasks to training paradigms, with the purpose of experimentally manipulating the to-be-measured psychological attribute, calls for the verification of the changing process following the intervention, and provides the second step for the establishment of measurement validity, i.e., an experimental research framework (see Chapter 1 and Borsboom, Mellenbergh, & van Heerden, 2004).

The second stage, which goes beyond the present work and is far ahead in time, takes into exam the effects of the experimental intervention on the clinical outcomes. Complete analyses will be then conducted in agreement with intention to treatment principle as per CONSORT statement (Schulz et al., 2010). The main effects and interaction effects of the CBM interventions on continuous outcome measures will be analysed with a 2 (attentional bias retraining: active/placebo) × 2 (approach bias retraining: active/placebo) × 3 (time: pre-intervention/post-intervention/3-month follow-up) repeated measures mixed ANOVA. Primary and secondary binary outcome measures will be analyses via a multivariate logistic regression. To answer the question "what works best for whom?", the predictive value of the cognitive bias implicit measures (AAT and VPT) and the clinical assessment questionnaires, as well as the demographics details and the control inhibition executive function (Stroop task), are considered in the clinical outcome effect of the two CBM re-trainings within a moderated mediation analysis (Preacher, Rucker, & Hayes, 2007).

Chapter 5

Measurement of implicit cognitive biases in alcohol addiction: What are we measuring?

A Many-Facet Rasch analysis

5.1 Introduction

Measurement of change represents a difficult challenge. We expect persons (patients, students, experimental participants) to change from Time 1 to Time 2 as a consequence of whatever manipulation administered between the two sessions. The functioning of test items and rating scales, though, may also change even when the same data collection protocol is used. Therefore, the challenge consists in measuring people and items in the same clearly defined frame of reference encompassing both time-points, so that measurements of change will have unambiguous numerical representation and substantive meaning.

The correct estimation along the same continuum of people's ability, both at baseline and post-intervention allows evaluating accurately the rate of change between the different administrations. Furthermore, it might also occur that, at first sight, a certain manipulation did not produce a manifest change in the expected direction or – and this is even worse for those who spend enormous efforts in designing experimental longitudinal studies – that the expected change did not occur. However, what is visible to the researcher's expert eyes does not always concur with what actually happened. Still, they both are complementary aspects of change over time and the use of invariant measures of the crucial underlying attributes is central to monitoring change as well as in attempting to catch the impact of other events or variables that do play a role in driving the change, such as experimental conditions, drugs, instructions, or individual different features that could moderate the desired *change*.

5.2 Many-Facet Rasch Model for longitudinal data

One of the fundamental advantages from constructing measures of psychological attributes is that the estimates derived from a Rasch analysis are located on an interval scale where the measurement unit is maintained at all points along the scale and for all the elements entered in the model.

The use of Rasch models to depict change over time implies to restructure the data by appending the person measures at Time 2 onto the baseline measures at Time 1, resulting in twice as many persons being measured (i.e.. stacking the data; Wright, 2003). Conceptually speaking this transforms both baseline and post-intervention assessment into measures on the same ruler to make interpretations about the effect of the manipulation applied between the two measurement sessions. The assumption is that the targeted psychological attribute or behaviour has changed as a result of the experimental intervention or manipulation administered. As Wright claimed, "[by] stacking the data, we see *who* has changed" (2003, p. 906).

The Many-Facet Rasch Measurement model (MFRM – Linacre, 1989), which belongs to the Rasch models family, has demonstrated to be a powerful modelling framework in tracking change processes, both in experimental social psychology research (e.g., Vianello & Robusto, 2010) and applied clinical psychology research (e.g., Mannarini, 2009), for the flexibility in the inclusion of several elements that can contribute to the outcome of an evaluation process (i.e., *facets*) and for the transformation of observed scores into scalar-invariant, meaningful and comparable measures.

5.3 The present study: objectives

The assessment of automatic approach tendencies and attentional bias in addiction research has seen an enormous growth in the development and experimentation of implicit measures targeting these dysfunctional automatic processes towards the substance object of the dependence (e.g., Stacy & Wiers, 2010).

The present study constitutes a corollary in the investigation of whether these maladaptive implicit processes can be changed or, better say, reversed, for it addresses the psychometric investigation of the two main implicit measures used and the evaluation of the implied implicit processes prior to and after an experimental treatment intervention on the same measurement unit. A preliminary analysis of the data obtained by baseline and post-intervention assessment of the targeted cognitive bias, provided by the Randomized Clinical Trial previously described (see Chapter 4), is conducted with a two-fold objective. First, the lack of studies about measurement validity of the *Approach Avoidance Task* and *Visual Probe Task* calls for the investigation of their measurement properties to establish whether they are measuring what they are intended to measure. Second, the adaptation of the same cognitive bias assessment task to cognitive bias training paradigm, with the purpose of experimentally manipulating the to-be-measured psychological attribute, calls for the verification of the changing process caused by the intervention, and provides the second step for the establishment of their measurement validity, within an experimental research framework (see Chapter 1 and Borsboom, Mellenbergh, & van Heerden, 2004).

This study is the first in attempting to prove the existence of a latent measurement dimension underlying the two implicit measures, wherein the task scores are located and share the same metric, i.e., a measurement unit that remains invariant all along the time-points considered (pre- and post-intervention) and across the elements supposed to contribute to the strength of the approach and/or attentional bias towards alcohol. The main idea behind this study entails the following consideration: it is risky to take for granted that something did change after a certain manipulation because measure A administered at two different moments detected a difference. Even in highly controlled studies, in which the highest standards for scientific research are applied to limit as much as possible any confounding or extraneous factor in the experimental process, the detected change is hardly ascribable to the solely effect of the manipulation. And so, the introduction of moderator and mediator variables and interaction effects with presumed extraneous factors to check for spurious effects. But, what if the measure used to evaluate whether the change occurred or not, *did change itself*? To use Bond and Fox words, "*it is impossible to measure change with a measure that changes*" (2007, p. 164).

Experimental designs offer a perfect breeding ground for exploring at once the functioning of a measure on one side, and the status of the theoretical attribute in question, on the other side.

The present study is explorative in nature, for little is known about the measurement properties of the two tasks and about the inherent features of the processes or constructs the two tasks aimed to quantify. Furthermore, the present analysis can be deemed as a first blueprint for future examination and reflection over the RCT results.

The main study objectives, which focus on both cognitive bias implicit measures, are the following:

1) Verification of the existence of a measurement dimension for each cognitive bias, over which the two implicit measures lie and are expressed on an invariant, interval measurement unit. As far as the theoretical status of the attribute 'attentional bias' and 'approach bias' is concerned, the reflection points to the dimensional nature of the implicit measures and to the domain-general versus domain-specific discussion on what the cognitive bias re-trainings do retrain (for a discussion, see Wiers, Gladwin, Hofmann, Salemink, & Ridderinkhof, 2013).

   Should we consider the VPT attentional bias measure as a unique, undefined measure of the automatic alcohol-triggering effect on attention? Or is it more reasonable to separately consider the two attentional processes (engagement and disengagement) as two components of the 'general' attentional bias? Previous results (Hallion & Ruscio. 2011; Schoenmakers et al., 2010) point to the second option. That's what we're going to test.

   A similar question is put forward for the approach bias measure: can we take it as a 'general' measure of the approach-avoid tendencies towards an X object, which is then sub-specified within the measure? Or, should we keep separate the object-specific approach tendencies? Note that in some cases the approach bias for alcohol did not imply an avoidance bias towards soft-drinks; rather, approach tendencies could co-exist and even be directed to different appetitive stimuli (Wiers, Rinck, Dictus, & Van den

Wildenberg, 2009; Wiers et al., 2013). Therefore the object-specific approach tendencies are to be considered in parallel.

2) Examination of the hypothesized time effect, i.e., did something and/or someone actually change from Time 1 to Time 2?
3) Examination of any individual difference related to the participants' gender.
4) Examination of the effect brought in by the experimental condition participants are assigned to, with the expectation of a better outcome for the double retraining condition.

5.3.1 The model

In the *Many-Facet Rasch Measurement* model (MFRM – Linacre, 1989), each observation is considered to be the outcome of an interaction of elements, such as the individual 'ability', the difficulty of the item, the condition the item is presented in, and so on. These interacting facets are modelled in the MFRM independently one to each other, so that their parameter estimates can be additively combined on the latent trait. The MFRM is an extension of Rasch's seminal *Simple Logistic Model* (SLM – Rasch, 1960), in which two main facets play the actor role when responding to an item: the person's ability and the item difficulty. According to a logistic distribution, the probability of a response *x* to a test, which can be correct (1) or incorrect (0), is a function of the ability β of respondent *v* and difficulty δ of the item *i,* expressed on the *logit* scale ($\beta_v - \delta_i$) (Rasch, 1960) (see Chapter 3).

While retaining the mathematical properties of Rasch models (specific objectivity, local independence, unidimensionality, monotonicity; for details, see Bond & Fox, 2007; Sijtsma, 2012), the MFRM extends the analysis to more complex situations by including other sources of systematic variability (*facets*) that can impact on the probability of a response.

In the present study, several facets were considered in the attempt to model the likelihood of a certain cognitive bias score: besides respondents' 'ability' (i.e., the approach tendency or the strength of the attentional bias) (facet 1) and task indices 'easiness' (facet 2), other facets were entered in the model equation, to account for other variables that may affect the performance of the task, namely, the measurement time-point (facet 3), participants' gender (facet 4), and, for post-intervention data only, the experimental condition participants were assigned to (facet 5). An additional parameter accounting for the cognitive bias score *k = {1. …. m}*, provided by the discretisation of the score distribution of the cognitive bias indices in the two tasks, was embedded in the model. The general MFRM model equations are then formally expressed as follows:

$$ln\left(\frac{P(X_{vibk})}{P(X_{vib(k-1)})}\right) = \beta_v - \delta_i - \lambda_b - \tau_k \quad (1)$$

and

$$ln\left(\frac{P(X_{vibcdk})}{P(X_{vibcd(k-1)})}\right) = \beta_v - \delta_i - \gamma_c - \eta_d - \tau_k \quad (2).$$

Equation (1) specifies the probability that a respondent *v* would respond to the index *i* at time *b* with a score *k* rather than *k-1*; β_v is the person *v*'s automatic approach tendency for the AAT and automatic attentional bias for the VPT parameter, δ_i is the index *i* representativeness of the approach tendency for the AAT and attentional bias for the VPT on the latent trait parameter, λ_b identifies the two time-points, and τ_k is the parameter for the step up to category *k* rather than *k – 1* of the index score.

Equation (2) was applied to data separately in the two time sessions and specifies the probability that a respondent *v* would respond to the index *i* in experimental condition *c* (only for post-intervention data), given his/her gender *d*, with a score *k* rather than *k-1*; β_v is the person *v*'s automatic approach tendency for the AAT and automatic attentional bias for the VPT parameter, δ_i is the index *i* representativeness of the approach tendency for the AAT and attentional bias for the VPT on the latent trait parameter, γ_c lists the four RCT experimental conditions, η_d identifies participants' gender, and τ_k is the parameter for the step up to category *k* rather than *k – 1* of the index score.

According to the MFRM theoretical and mathematical perspective, Equation 1 describes the *logit* (i.e., the log-likelihood) of a certain response *k* as the dependent variable, whereas the various factors entered in the model act as independent variables that influence (or control) the response.

All parameter estimates were positively scaled in the analyses, so that positive values indicate strong approach bias in the AAT and strong attentional bias towards alcohol in the VPT, whereas negative measures indicate the opposite.

To evaluate the goodness-of-fit of the parameter estimates, the MFRM presents several fit indices, which can be informative both for the evaluation of the data adherence to the model requirements – Rasch models are prescriptive in nature – both for the interpretation of the results. The main fit statistics are the *mean square Infit* and *mean square Outfit* statistics, which are expected to be equal to 1 for a perfect fit to the model and in the range .50 – 2 for a an acceptable fit (Linacre. 2009). If an Infit or an Outfit has a value 1 + *x* then there is 100*x*% more variation between the observed and the model-derived response patterns. For instance, an Infit statistic equal to 1.30 indicates 30% more variation in the observed data than the model prediction, whereas a value of .70 signals a 30% less variation in the observed data compared to the model prediction (Bond & Fox. 2007). For a detailed presentation of MFRM goodness-of-fit indices see Chapter 3.

5.3.2 Task scoring and data pre-processing

5.3.2.1 Visual Probe Task

The scoring and data processing of the baseline and post-intervention VPT task data for the MFRM analysis was carried out according to the following steps (cf. MacLeod Rutherford Campbell, Ebsworthy, & Holker, 2002; Schoenmakers et al., 2010; Schoenmakers, Wiers, Jones, Bruce, & Jansen, 2007):

1. Practice and error trials were removed;
2. Latencies smaller than 200ms and greater than 3000ms were removed;
3. The median values of response latencies to the alcoholic and non-alcoholic trials in both formats (i.e.. probe after the pictures and probe after the pictures. respectively), separately for trials presenting static and dynamic pictures, were computed, for a total of 2x2x2 = 8 median values per participant;
4. The median values were subtracted one to each other (non-alcoholic trials – alcoholic trials), separately for static and dynamic pictures, to

obtain two indices for each attentional component: engagement/dynamic, engagement/static, disengagement/dynamic, and disengagement/static. The score was computed so that positive values indicate a stronger attentional bias towards alcohol;

5. Each index scores was successively discretised into five categories according to the quintiles (20^{h}, 40th, 60th, 80th) computed within each attentional bias component (e.g., engagement/dynamic and engagement/static), to index:
   1 = strong attentional bias towards non-alcoholics
   2 = mild attentional bias towards non-alcoholics
   3 = no definite attentional bias towards either alcoholics or non-alcoholics
   4 = mild attentional bias towards alcoholics
   5 = strong attentional bias towards alcoholics.
   This discretization procedure complies with the Rasch modelling requirement of entering only discrete variables in the model (Blanton & Jaccard, 2006).
6. A first P ✕ T ✕ I matrix with the four attentional bias indices was created, where P identifies participant *n.* T identifies time-point *t*. and I identifies the attentional bias index *i*; in this matrix, post-intervention data for each participant were appended to the baseline data, resulting in a matrix twice as many participants that completed both measurement sessions;
7. The P ✕ T ✕ I matrix was then split into the pre- and post-intervention matrices P ✕ G ✕ I and P ✕ G ✕ C ✕ I, where P identifies participant *n,* G codes for participant's gender *g,* C lists the experimental condition *c* (post-intervention only), and I identifies the attentional bias index *i*;
8. A row of other 4 matrices were built by separating the distinct processes involved in the attentional bias in each assessment session: two matrices P ✕ G ✕ D and P ✕ G ✕ E with baseline data only, where P identifies participant *n,* G codes for participant's gender *g*, and D and E identifies the attentional disengagement index *d* and the attentional engagement index *e*, respectively; and two matrices P ✕ G ✕ C ✕ D and P ✕ G ✕ C ✕ E with post-intervention data only, where P identifies participant *n,* G codes for

participant's gender *g*, C lists the experimental condition *c*, and D and E identifies the attentional disengagement index *d* and the attentional engagement index *e*, respectively.

5.3.2.2 Approach-Avoidance Task

A similar pre-processing procedure was applied to the baseline and post-intervention data of the alcohol AAT, except for the scoring algorithm, which followed an adapted version of the standard D-score designed for the *Implicit Association Test*. The improved D algorithm standardizes the difference in response latencies by dividing an individual's difference in RTs by a personalized standard deviation of these latencies (Greenwald, Nosek, & Banaji, 2003; Sriram, Greenwald, & Nosek, 2010).[8] The advantage of such standardized scores over simple difference scores is that they are less vulnerable to biases due to differences in average reaction time. In recent studies with the AAT (Eberl et al., 2012; Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011) the algorithm performed better than the original scoring algorithm. The algorithm yields an approach bias score for each drink type (alcoholics. non-alcoholics). Positive scores indicate an approach tendency, negative ones an avoidance tendency. The larger the score, the stronger the approach tendency.

The AAT data processing was done as follows:

1. Practice trials and latencies lower than 300ms were removed;
2. Incorrect responses were replaced by the mean of correct responses in that same trial format (e.g., mean alcohol/push/static trials) plus a penalty of twice the standard deviation of the same correct responses;
3. The D-score was computed for alcohol trials and non-alcohol trials ((mean alcohol/pull – alcohol/push)/SD(alcohol)), separately for trials using static and dynamic pictures, for a total of four D-scores indicating the approach tendencies towards alcohol and soft-drinks (two for static and tow for dynamic pictures);

[8] Until now, the AAT effect has been calculated as the difference between the median scores for pushing pictures of one category (alcohol or soft drinks) and the median scores for pulling pictures of that category (cf. Rinck & Becker, 2007; Wiers et al., 2009; Wiers, Rinck, Kordts, Houben, & Strack, 2010).

4. Each D-score was successively discretised into five categories according to the quintiles (20^{h}, 40th, 60th, 80th) computed within each drink type (e.g., alcohol approach/dynamic and alcohol approach/dynamic), to index:
   1 = strong avoid bias
   2 = mild avoid bias
   3 = no definite approach/avoid bias
   4 = mild approach bias
   5 = strong approach bias.
5. Similarly to the VPT pre-processing, a first P ⨉ T ⨉ A matrix with the four approach bias indices was created, where P identifies participant *n,* T identifies time-point *t*. and A identifies the approach bias index *i*; also this matrix includes pre- and post-intervention data for each participant;
6. The P ⨉ T ⨉ A matrix was then split into the pre- and post-intervention matrices P ⨉ G ⨉ A and P ⨉ G ⨉ C ⨉ A, where P identifies participant *n,* G codes for participant's gender *g,* C lists the experimental condition *c* (post-intervention only), and A identifies the approach bias index *i*;
7. The last 4 matrices were built by separating the approach bias for the two drink types (alcoholics and non-alcoholics) in each assessment session: two matrices P ⨉ G ⨉ A and P ⨉ G ⨉ S with baseline data only, where P identifies participant *n*, G codes for participant's gender *g*, and A and S identifies the alcohol approach bias index *a* and non-alcohol approach bias index *s*. respectively; and two matrices P ⨉ G ⨉ C ⨉ A and P ⨉ G ⨉ C ⨉ S with post-intervention data only, where P identifies participant *n*, G codes for participant's gender *g*, C lists the experimental condition *c*, and A and S identifies the alcohol approach bias index *a* and the non-alcohol approach bias index *s*.

5.4 Results

The MFRM analyses of the attentional bias and approach bias implicit measures were carried on 53 participants (69.8%male; mean age 51.8 SD = 8.2), who completed pre- and post-intervention assessment, without any episode of lapse during the experimental intervention. The analysis followed several steps, which will be presented progressively for each task. FACETS software (version 3.66.0) was used for the analyses (Linacre. 2009).

5.4.1 Alcohol attentional bias: Visual Probe Task

Although presented within the same task and sharing similar structural features, the engagement and disengagement indices for the attention bias towards alcohol were analysed jointly as a first step and then separated in two parallel analyses flow according to two main reasons:

1) a theoretical reason, which considers the two attentional processes to be different, to operate independently, and to be susceptible of being differently affected by the experimental training intervention (cf. Schoenmakers et al.. 2007. 2010);

2) a methodological reason, which adheres to the unidimensionality property of Rasch models and, in case of multidimensional measures, it envisages the consideration of one sub-dimension at once (Bond & Fox. 2007).

Participants' records in time 1 and time 2 for the four indices, engagement indices, and disengagement indices, were stacked into three long files. In this stage the facet time coded for the measurement time-point to check for a difference the measurement two sessions.

The analysis did not recover any difference between pre- and post-intervention for both the general, overarching attentional bias dimension and the two sub-dimensions for the engagement (E) and disengagement (D) attentional processes ($\chi^2_{(1)}$s = 0, *p*s > .05).

The subsequent step took into consideration the separate analysis of pre- and post-intervention data, following the same triplet logic.

Participants' parameter estimates recovered at pre-test are summarised in Table 5.1. At pre-test, participants did not present significant differences in their attentional bias measures, both in their general attentional bias towards alcohol (β ranged from -1.74 to 1.33, $\bar{\beta}$ = -.04, $\overline{SE}$ = .27, *SD* = .27; $\chi^2_{(52)}$ = 46.5, *p* = .69) and within the bias components of attentional engagement (β ranged from -2.75 to 3.01, $\bar{\beta}$ = .00, $\overline{SE}$ = .87, *SD* = 1.30; $\chi^2_{(52)}$ = 55, *p* = .36). When considering the attentional disengagement process in the expression of bias, participants presented a spread in their measures at the limit of statistical significance (β

Table 5.1 Participants parameter estimates (β_v) at baseline: attentional bias and attentional sub-components.

	General Attentional Bias		**Engagement**		**Disengagement**	
Subject	**β**	***SE***	**β**	***SE***	**β**	***SE***
4	**1.33**	**0.65**	**0.93**	**0.77**	**2.85**	**1.82**
15	**0.99**	**0.52**	**0.93**	**0.77**	**1.63**	**1.02**
23	**0.83**	**0.52**	**0.30**	**0.66**	**2.91**	**1.82**
21*	0.73	0.46	2.88	1.85	0.02	0.64
2	0.59	0.46	0.30	0.66	1.68	1.02
9	**0.56**	**0.46**	**0.44**	**0.66**	**1.58**	**1.02**
35	0.54	0.42	-0.10	0.62	1.70	1.02
52	**0.54**	**0.42**	**2.88**	**1.85**	**-0.40**	**0.66**
20	0.40	0.42	0.80	0.77	0.43	0.67
24	0.40	0.42	2.88	1.85	-0.41	0.66
34	0.40	0.42	1.59	1.05	0.01	0.64
14*	0.39	0.40	0.93	0.77	-0.05	0.64
33	0.35	0.42	1.60	1.05	-0.03	0.64
50	0.35	0.42	-0.09	0.62	1.64	1.02
8	0.23	0.40	2.88	1.85	-0.90	0.75
49*	0.23	0.40	0.30	0.66	0.43	0.67
41	0.20	0.4	3.01	1.85	-1.01	0.75
22	0.18	0.40	-0.48	0.64	1.64	1.02
27	0.18	0.40	-0.48	0.64	1.64	1.02
37	0.12	0.38	-0.50	0.64	0.49	0.67
6	0.07	0.38	0.31	0.66	-0.40	0.66
17	0.07	0.38	-0.10	0.62	0.02	0.64
36	**0.07**	**0.39**	**0.80**	**0.77**	**-0.41**	**0.66**
18*	-0.06	0.38	-0.36	0.64	-0.05	0.64
26	-0.06	0.38	0.02	0.62	-0.47	0.66
11	-0.07	0.38	-0.49	0.64	0.43	0.67
5	-0.10	0.38	0.44	0.66	-0.52	0.66
45	-0.10	0.38	-0.36	0.64	0.32	0.67
48	-0.10	0.38	-0.80	0.71	0.83	0.76
10	-0.18	0.40	-1.61	0.97	0.49	0.67
30	-0.21	0.40	-2.62	1.78	0.88	0.76
12	-0.22	0.38	-0.10	0.62	-0.41	0.66
13	-0.22	0.38	-0.49	0.64	0.01	0.64
25	-0.22	0.38	-2.75	1.78	1.68	1.02
39	-0.22	0.38	-0.10	0.62	-0.41	0.66
51	-0.23	0.40	-0.49	0.64	-0.40	0.66
1	-0.25	0.38	-1.47	0.97	0.83	0.76
31	-0.25	0.38	0.44	0.66	-1.01	0.75
19	-0.27	0.38	0.31	0.66	-0.94	0.75
40*	-0.38	0.42	-0.36	0.64	-0.95	0.75
29	-0.40	0.42	-0.10	0.62	-1.61	1.00
47	-0.40	0.40	-0.80	0.71	-0.10	0.64
32*	-0.42	0.40	-0.48	0.64	-0.45	0.66
46	-0.54	0.42	-0.94	0.71	-0.41	0.66
53	-0.54	0.42	-0.10	0.62	-1.62	1.00
38	-0.57	0.42	-0.80	0.71	-0.52	0.66
43	-0.59	0.42	-1.60	0.97	-0.03	0.64
44	-0.73	0.46	-0.49	0.64	-1.62	1.00
42	-0.76	0.46	-0.80	0.71	-1.01	0.75
3	-0.78	0.52	-0.94	0.71	-1.57	1.00
28	-0.78	0.52	-0.50	0.64	-2.75	1.8
16	-1.14	0.64	-0.81	0.71	-2.86	1.80
7	-1.74	0.94	-2.74	1.78	-1.61	1.00

* Participant estimates with misfitting Infit and Outfit statistics.

ranged from -2.86 to 2.91, $\bar{\beta}$ = -.01, $\overline{SE}$ = .85, *SD* = 1.22; $\chi^2_{(52)}$ = 64.8, *p* = .11). Only six participants (3.18%) presented a misfit to the model requirements (i.e., Infit and Outfit greater than 2 or lower than .5).

In Table 5.1, participants highlighted in bold at the upper limit of the latent trait are those presenting a strong attentional towards alcohol. Although at the group-level the facet time did not evidence any statistically significant difference, these participants have been selected to qualitatively see the shift of their measures along the measurement time-points. The change in their measures is presented in *Figure 5.1.*

Males and females presented a marginal difference in the strength of their general attentional bias towards alcohol, which disappeared when looking at the implied engagement and disengagement processes (see Table 5.2).

At pre-intervention male (η = .10) presented a slightly stronger attentional bias towards alcoholics than women (η = -.10).

Table 5.2 Gender facet parameter estimates (η) at baseline for general attentional bias and attentional sub-components.

	General Attentional Bias		**Engagement**		**Disengagement**	
Gender	**η**	**SE**	**η**	**SE**	**η**	**SE**
Male	.10	.07	.00	.17	-.03	.18
Female	-.10	.10	.00	.13	.03	.13
	$\chi^2_{(1)}$ = 2.5, *p* = .11; G = 1.23, R = .60; .99 ≤ Infit/Outfit ≤ 1.05		$\chi^2_{(1)}$ = 0, *p* > .05;		$\chi^2_{(1)}$ = .1, *p* > .05;	

The four indices were hypothesised to be part of the same measurement dimension of alcohol attention bias, due to their function of measuring the automatic tendency to focus the attention on highly salient cues. However, different attentional processes are implied in this bias. It was then expected a difference between the disengagement and engagement indices. Further, a minor difference between indices computed on static and dynamic task trials was hypothesized. The results across the general and specific features of the attentional bias are presented in Table 5.3.

Table 5.3 Engagement (E) and disengament (D) indices estimates (δ) at baseline: general attentional bias and attentional sub-components.

	Attentional Bias		Engagement		Disengagement	
Index	**δ**	***SE***	**δ**	***SE***	**δ**	***SE***
E-dynamic	.07	.11	.11	.14	-	-
E-static	-.07	.11	-.11	.14	-	-
D-dynamic	-.15	.12	-	-	-.25	.15
D-static	.15	.12	-	-	.25	.15
	$\chi^2_{(3)} = 4.3$, $p > .05$; G = .67, R = .31; .87 ≤ Infit/Outfit ≤ 1.15		$\chi^2_{(1)} = 1.2$, $p > .05$; G = .49, R = .19; .95 ≤ Infit/Outfit ≤ .99		$\chi^2_{(1)} = 5.6$, $p = .02$; G = 2.15, R = .82; .90 ≤ Infit/Outfit ≤ .99	

The four indices presented a satisfactory fit in both the general attentional bias dimension and in its components (.87 ≤ Infit/Outfit ≤ 1.15). As expected they are not clearly different one to each other ($\chi^2_{(3)} = 4.3$, $p > .05$) and seem to index similar aspects of the latent trait. However, a more careful look at the indices estimates can give us a clue on their functioning: it is already evident in the general attentional bias measures that the indices are paired one to each other according to the specific attentional component they imply.[9] This was further proved when separating the analysis for each component: in both of them the fit statistics and estimates trend are alike.

The MFRM analysis also retrieved the difference between the active and dynamic pictures used in the task. Engagement and disengagement indices appear to act differently according to the stimuli presented, and this was particularly pronounced when attentional disengagement processes are singled out ($\chi^2_{(1)} = 5.6$, $p = .02$). When alcoholic and non-alcoholic static pictures were presented in the disengagement trial format (probe on top of the pictures) participants had more impediments in shifting their attention away to detect the probe ($\delta_{Ds} = .25$), when compared to the presentation of active pictures in the same trial format ($\delta_{Dd} = -.25$). The opposite occurred in engagement trials: the presentation of alcoholic and non-alcoholic static pictures just before the appearance of the probe ($\delta_{Es} = -.11$) seems to be 'easier' than the presentation of dynamic pictures ($\delta_{Ed} = .11$).

[9] The log-linear mathematical properties of Rasch models make the sum of parameter estimates equal to 0. In this case the parameter estimates cancel in pairs (E-active – E-dynamic = (.07) - (.07)). suggesting a paired functioning of the four estimates.

Table 5.4 Participants parameter estimates (β_v) at post-intervention: general attentional bias and attentional sub-components.

	General Attentional Bias		Engagement		Disengagement	
Subject	**β**	***SE***	**β**	***SE***	**β**	***SE***
3*	1.31	0.66	1.92	1.06	1.10	0.97
27	1.27	0.67	0.85	0.78	2.76	1.81
14	1.24	0.66	2.98	1.85	0.77	0.70
16	0.78	0.54	1.45	1.06	0.61	0.70
44	0.59	0.47	0.45	0.68	1.18	0.97
34*	0.55	0.43	1.89	1.06	0.11	0.59
17*	0.53	0.43	1.81	1.06	-0.15	0.59
19	0.51	0.40	3.32	1.85	-0.33	0.71
2*	0.49	0.47	0.20	0.68	1.02	0.97
5	0.46	0.47	0.48	0.78	0.95	0.70
31	0.42	0.47	0.50	0.78	0.69	0.70
53	0.39	0.43	3.06	1.85	-0.61	0.61
25	0.33	0.43	-0.25	0.65	1.28	0.97
1	0.30	0.40	-0.10	0.65	1.22	0.70
26	0.25	0.39	0.48	0.68	0.36	0.59
22*	0.20	0.38	0.27	0.65	0.46	0.59
23	**0.11**	**0.40**	**-0.24**	**0.65**	**0.36**	**0.70**
46	0.11	0.40	-0.24	0.65	0.36	0.70
10	0.10	0.39	0.33	0.68	-0.23	0.59
18*	0.10	0.39	-0.07	0.65	0.35	0.61
30	0.10	0.38	-0.38	0.66	0.72	0.61
13	0.06	0.39	0.01	0.65	0.10	0.61
28	0.06	0.39	-0.52	0.66	0.28	0.70
39	0.06	0.39	-0.41	0.66	0.52	0.70
41	-0.01	0.39	0.75	0.78	-0.28	0.61
43	-0.06	0.38	0.86	0.78	-0.85	0.71
8*	-0.09	0.39	0.13	0.65	-0.24	0.61
11	-0.09	0.39	-0.30	0.66	0.11	0.59
47	-0.16	0.38	-0.21	0.65	0.08	0.59
7	-0.17	0.42	-1.21	1.00	0.38	0.59
33	-0.17	0.39	-0.11	0.65	-0.17	0.61
35*	-0.22	0.40	-0.38	0.66	-0.25	0.61
29	-0.26	0.40	-1.58	0.99	0.21	0.61
37*	-0.30	0.42	-0.64	0.74	-0.32	0.61
38	-0.32	0.40	-0.10	0.65	-0.34	0.71
12	-0.35	0.39	-1.15	0.74	-0.06	0.61
24	-0.35	0.39	-0.24	0.65	-0.77	0.61
50	-0.37	0.40	0.86	0.78	-2.70	1.82
20	-0.41	0.42	-1.50	1.00	0.11	0.59
9	**-0.42**	**0.39**	**-0.89**	**0.66**	**-0.08**	**0.59**
32	-0.43	0.42	-1.47	0.99	0.09	0.59
49*	-0.46	0.40	-0.67	0.66	-0.51	0.61
48	-0.48	0.42	0.34	0.68	-2.18	1.82
6	-0.52	0.46	-2.50	1.8	0.01	0.59
52	**-0.52**	**0.46**	**-2.50**	**1.80**	**0.01**	**0.59**
15	**-0.53**	**0.42**	**-0.07**	**0.65**	**-1.46**	**0.98**
4	**-0.59**	**0.42**	**-0.76**	**0.66**	**-0.69**	**0.71**
51	-0.60	0.52	-2.39	1.81	0.02	0.61
40	-0.63	0.42	-0.32	0.65	-1.62	0.98
21*	-0.82	0.52	-0.38	0.66	-2.52	1.82
42	-0.93	0.46	-1.38	0.74	-0.87	0.71
36	**-0.99**	**0.52**	**-1.61**	**0.99**	**-1.03**	**0.71**
45	**-1.39**	**0.64**	**-3.02**	**1.81**	**-0.71**	**0.71**

* Participants estimates with misfitting Infit and Outfit statistics.

A similar analytical strategy was applied to the exploration of VPT post-intervention data, except for the introduction of an additional facet accounting for the experimental condition participants were assigned to.

Participants' parameter estimates recovered at post-intervention are summarised in Table 5.4. At post-intervention participants still did not present significant differences in their attentional bias measures both in the general attentional bias towards alcohol (β ranged from -1.39 to 1.31, $\bar{\beta}$ = -.04, $\overline{SE}$ = .44, *SD* = .55; $\chi^2_{(52)}$ = 59.3, *p* = .23) and within the bias components of attentional engagement (β ranged from -3.01 to 3.32, $\bar{\beta}$ = .09, $\overline{SE}$ = .89, *SD* = 1.31; $\chi^2_{(52)}$ = .59, *p* = .24) and attentional disengagement (β ranged from -2.70 to 2.76, $\bar{\beta}$ = -.05, $\overline{SE}$ = .76, *SD* = .94; $\chi^2_{(52)}$ = 64.8, *p* = .11). Eleven participants (5.83%) presented a misfit to the model requirements (i.e., Infit and Outfit greater than 2 or lower than .5).

At post-intervention, males and females did not present any difference in the strength of the attentional bias towards alcohol ($\chi^2_{(1)}$ = .8, *p* > .38), generalized to the two attentional processes.

Table 5.5 Gender facet parameter estimates (η) at post-intervention for general attentional bias and attentional sub-components.

	General Attentional Bias		**Engagement**		**Disengagement**	
Gender	**η**	**SE**	**η**	**SE**	**η**	**SE**
Male	.05	.07	.07	.13	-.04	.12
Female	-.05	.10	-.07	.18	.04	.15
	$\chi^2_{(1)}$ = .8, *p* > .05;		$\chi^2_{(1)}$ = .4, *p* > .05;		$\chi^2_{(1)}$ = .2, *p* > .05;	

The four indices maintained satisfactory good fit statistics in both the general attentional bias dimension and in its components (.88 ≤ Infit/Outfit ≤ 1.30) (see Table 5.6). However, the pattern of results within the general attentional bias and particularly in its attentional components did change. Although they still kept a paired structure, the four indices did not sum up perfectly to zero and started to vaguely differentiate ($\chi^2_{(3)}$ = 5.1, *p* = .17).

Table 5.6 Engagement (E) and disengament (D) indices estimates (δ) at post-intervention: general attentional bias and attentional sub-components.

	Attentional Bias		Engagement		Disengagement	
Index	**δ**	***SE***	**δ**	***SE***	**δ**	***SE***
E-dynamic	-.18	.12	-.28	.15	-	-
E-static	.16	.12	.28	.15	-	-
D-dynamic	.08	.12	-	-	.09	.13
D-static	-.06	.12	-	-	-.09	.13
	$\chi^2_{(3)} = 5.1$, $p = .17$; G = .84, R = .21; .88 ≤ Infit/Outfit ≤ 1.30		$\chi^2_{(1)} = 6.9$, $p = .01$; G = 2.42, R = .85; .94 ≤ Infit/Outfit ≤ .98		$\chi^2_{(1)} = .9$, $p > .35$; G = .00, R = .00; .91 ≤ Infit/Outfit ≤ 1.01	

The examination of the indices estimates within the two attentional components still indicated a different functioning of indices according to the attentional processes they belong to. At Time 2, this was particularly pronounced for the engagement processes ($\chi^2_{(1)} = 6.9$, $p = .01$). The analysis revealed a second switch: the pattern of differences between static and dynamic indices was reversed between and within the engagement and disengagement attentional components. When alcoholic and non-alcoholic static pictures were presented in the engagement trial format (probe after the pictures) the task elicited more the attentional bias towards alcohol ($\delta_{Es} = .28$), whereas the dynamic pictures pushed it towards non-alcoholics ($\delta_{Ds} = -.28$). Conversely, the disengagement format did not trigger the difference in pictures any longer ($\chi^2_{(1)} = .9$, $p > .35$).

Table 5.8 presents the parameter estimates for the facet experimental condition. The four groups did not significantly distinguish one to each other in the strength of the attentional bias they bring in ($\chi^2_{(3)} = 3.1$, $p > .05$). However it is evident that the experimental condition in which participants receive the real intervention (i.e., VPT and AAT retrainings) carries off a decrease of the attentional bias towards alcohol by shifting it towards the soft-drinks. This trend emerges in all constituents of attentional automatic processes (γ_As = [-.29, -.16]). Also the combination of VPT-placebo and AAT-retraining showed a parameter estimate – although not statistically significant – pointing to a reversed bias towards alcohol. Finally, among all the worse case involves the allocation to the double placebo condition, where the attentional disengagement processes push towards an automatic reaction towards alcohol.

Table 5.8 Experimental condition parameter estimates (γ) for general attentional bias and attentional components.

Index	Attentional Bias		Engagement		Disengagement	
	γ	***SE***	γ	***SE***	γ	***SE***
A	-.16	.11	-.21	.21	-.29	.17
B	.00	.12	-.10	.23	.07	.17
C	.06	.15	.16	.23	-.02*	.25
D	.10	.11	.15	.19	.24	.18
	$\chi^2_{(3)}$ = 3.1, p > .05;		$\chi^2_{(3)}$ = 2.4, p > .05;		$\chi^2_{(3)}$ = .4.7, p = .19;	
	G = .84, R = .21;		G = .00, R = .00;		G = .50, R = .20;	
	.70 ≤ Infit/Outfit ≤ 1.13		.66 ≤ Infit/Outfit ≤ 1.20		.25 ≤ Infit/Outfit ≤ 1.45	

A = double retraining; B = VPT placebo and AAT retraining; C = VPT retraining and AAT placebo; D = double placebo. * Infit/Outfit statistics < .50.

For illustrative purposes, a graphical representation of the participants evidenced in bold in Tables 5.1 and 5.4 is attached below. Participants' individual estimates of pre- and post- intervention general attentional bias, attentional difficulty to disengage, and ease of engagement, are plotted in *Figure 5.1*.

5.4.2 Alcohol approach bias: Approach Avoidance Task

The MFRM analysis of the alcohol approach bias implicit measure, the AAT, followed the same step-by-step approach used so far in the VPT. The approach bias indices scored the automatic approach tendencies towards alcohol and soft-drinks in two distinct pairs of indices, each with active dynamic pictures. Also in this case, the major question was: can we consider them as part of the same space? Or should we better off separating them according to the object approached? In the effort to answer this measurement and theoretical questions, the pre- and post-intervention datasets were collapsed into one long file by including a facet time to verify any difference between measurement sessions.

Participants' records in Time 1 and Time 2 for the four approach indices, alcohol approach indices, and soft-drinks approach indices, were stacked into three long files. Afterwards, the analytical approach strategically considered the running of parallel model estimations on the general approach-avoid dimension, which does not envisage any specific object to avoid or to approach, and on the approach-avoid sub-dimensions gravitating around a specific object: alcoholics or non-alcoholics.

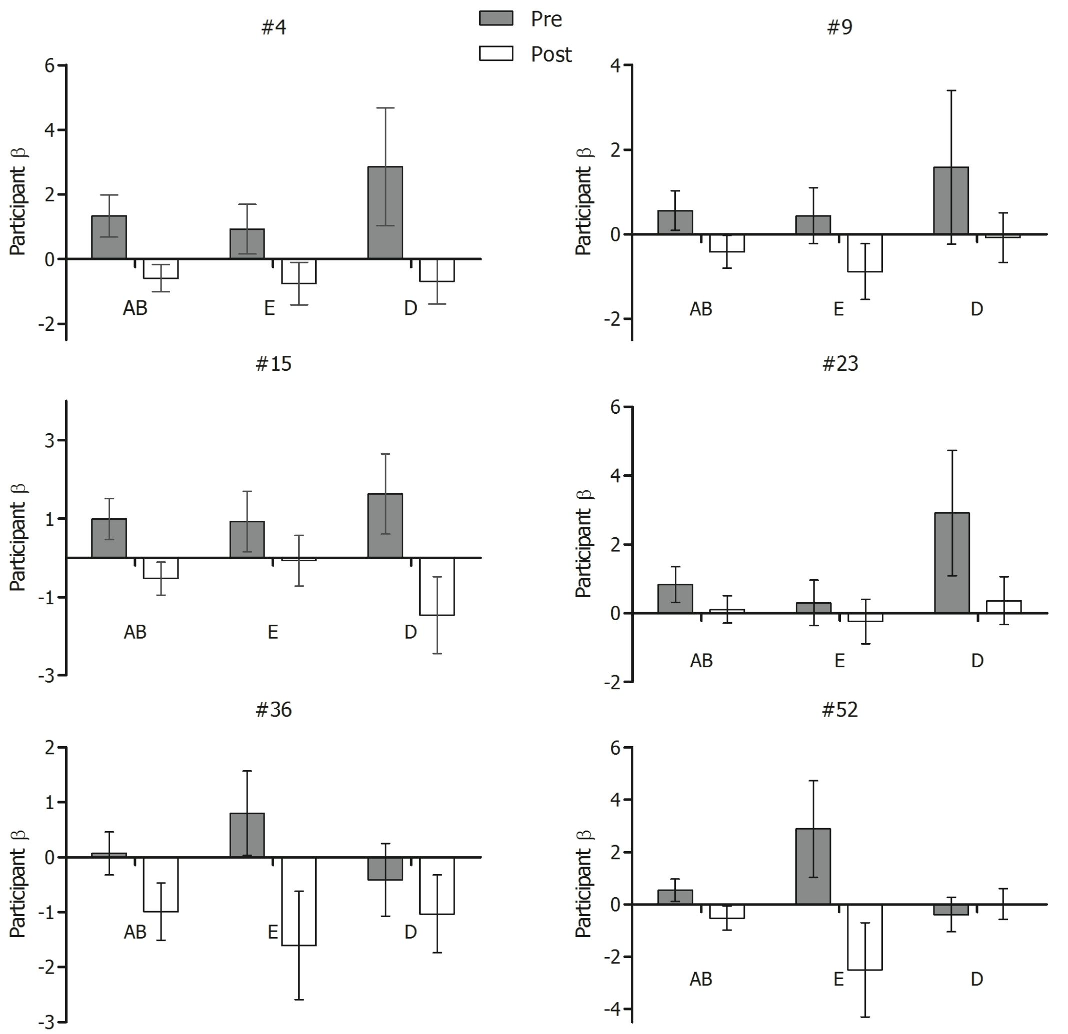


Figure 5.1 Attentional bias estimates (β) at pre- and post-intervention for participants with a strong alcohol attentional bias at baseline (subjects in bold in Tables 5.1 and 5.4): general attentional bias (AB), disengagement (D) and engagement (E). Higher values indicate stronger attentional bias towards alcohol.

The time facet didn't show any difference in any of the three hypothesized approach-avoid dimensions ($\chi^2_{(1)}$s = 0, *ps* > .05). The step further was then to examine baseline and post-intervention data separately.

Participants' parameter estimates recovered at post-intervention are summarised in Table 5.9. At baseline participants presented different approach

Table 5.9 Participants parameter estimates (β_v) at baseline: general approach bias and alcohol/non-alcohol approach tendencies.

	General Approch Bias		Alcohol Approch		Non-alcohol Approch	
Subject	**β**	***SE***	**β**	***SE***	**β**	***SE***
15*	3.03	1.81	2.86	1.83	2.88	1.89
27	**1.66**	**0.97**	**1.49**	**1.02**	**2.60**	**1.89**
47	**1.08**	**0.56**	**0.15**	**0.66**	**3.55**	**1.89**
8	**1.02**	**0.68**	**1.49**	**1.02**	**1.26**	**1.09**
48	**1.02**	**0.68**	**2.71**	**1.83**	**0.39**	**0.82**
14*	0.90	0.49	3.09	1.83	0.00	0.67
46	**0.90**	**0.49**	**1.86**	**1.02**	**0.48**	**0.71**
2	0.69	0.49	0.07	0.63	3.00	1.89
18	0.65	0.56	1.49	1.02	0.39	0.82
19	0.59	0.49	0.89	0.75	0.67	0.82
21	0.59	0.49	0.40	0.66	1.54	1.09
49*	0.54	0.41	-0.03	0.63	1.15	0.71
38	0.47	0.45	0.48	0.66	0.79	0.82
13	0.38	0.49	1.49	1.02	-0.19	0.71
3	0.37	0.41	0.86	0.75	-0.30	0.76
25*	0.37	0.41	0.38	0.66	0.22	0.68
11*	0.28	0.43	0.48	0.66	0.21	0.71
50*	0.22	0.41	0.15	0.66	0.28	0.67
28	0.18	0.43	-0.01	0.63	0.67	0.82
37	**0.18**	**0.43**	**0.89**	**0.75**	**-0.39**	**0.67**
20	0.16	0.45	-0.16	0.63	1.26	1.09
53	0.16	0.45	-0.56	0.65	2.60	1.89
1*	0.05	0.41	-0.26	0.63	0.28	0.67
6	0.05	0.41	0.64	0.75	-0.70	0.76
9	0.02	0.42	-0.89	0.73	0.67	0.67
4*	-0.02	0.41	-0.64	0.73	0.48	0.71
31	-0.02	0.41	-0.18	0.65	0.00	0.67
44	-0.02	0.41	0.22	0.63	-0.46	0.68
17	-0.11	0.41	0.64	0.75	-1.44	1.02
26	-0.16	0.45	-0.03	0.63	-1.05	1.01
33	-0.16	0.45	-0.89	0.73	0.22	0.68
34	-0.16	0.45	-0.43	0.65	-0.30	0.76
52	-0.21	0.41	-1.02	0.73	1.26	1.09
30*	-0.24	0.41	-0.33	0.65	-0.27	0.67
23	-0.38	0.41	-0.16	0.63	-0.67	0.67
29	-0.38	0.41	0.25	0.66	-1.13	0.68
36*	-0.38	0.41	-0.56	0.65	-0.19	0.71
35	-0.39	0.45	-0.64	0.73	-0.46	0.68
42	-0.39	0.45	-0.18	0.65	-0.97	0.76
7	-0.48	0.45	-2.96	1.79	0.76	0.71
24	-0.48	0.45	-1.12	0.73	-0.18	0.68
45*	-0.65	0.55	-0.89	0.73	-1.05	1.01
5	-0.70	0.45	-0.41	0.65	-1.37	0.76
12	-0.70	0.45	-0.87	0.73	-0.85	0.68
22	-0.70	0.45	-0.87	0.73	-0.85	0.68
10	-0.91	0.45	-0.56	0.65	-1.64	0.76
32*	-0.91	0.48	-0.87	0.73	-1.37	0.76
51	-0.91	0.45	-1.02	0.73	-1.13	0.68
39	-1.13	0.48	-0.56	0.65	-2.39	1.02
40*	-1.13	0.48	-1.02	0.73	-1.64	0.76
43	-1.23	0.67	-0.64	0.73	-2.92	1.83
41*	-1.76	0.67	-1.71	0.98	-2.39	1.02
16*	-3.30	1.80	-2.71	1.79	-3.32	1.83

* Participants estimates with misfitting Infit and Outfit statistics.

bias measures (β ranged from -3.30 to 3.03, $\bar{\beta}$ = -.04, $\overline{SE}$ = .52, *SD* = .90; $\chi^2_{(52)}$ = 85.5, *p* < .001). When looking at the approach to the two object categories, participants showed similar approach estimates towards alcohol (β ranged from -2.96 to 3.09, $\bar{\beta}$ = .02, $\overline{SE}$ = .83, *SD* = 1.17; $\chi^2_{(52)}$ = 60. *p* = .24), whereas for soft-drinks participants' measure of their approach tendencies were well disperse (β ranged from -3.32 to 3.55, $\bar{\beta}$ = -.05, $\overline{SE}$ = .93, *SD* = 1.41; $\chi^2_{(52)}$ = 74.8, *p* = .03). Fifteen participants (7.95%) presented a misfit to the model requirements (i.e., Infit and Outfit greater than 2 or lower than .5).

In Table 5.9, participants highlighted in bold are those presenting a strong approach bias towards alcohol (upper limit). Similarly to the VPT, these participants have been selected to visually see the shift of their measures along the measurement time-points. The change in their measures is presented in *Figure 5.2.*

At pre-intervention males (η = .16) presented a general stronger approach bias than females (η = -.16). This effect encompassed the approach tendencies towards both alcohol (η = .11) and soft-drinks (η = .20). However, men and women were not significantly different in the automatic approach tendencies towards the addiction object ($\chi^2_{(1)}$ = 1.3, *p* > .05).

Table 5.10 Gender facet parameter estimates (η) at baseline for general approach bias and alcohol/soft-drinks approach tendencies.

	Approach Bias		**Alcohol Approach**		**Non-alcohol Approach**	
Gender	**η**	**SE**	**η**	**SE**	**η**	**SE**
Male	.16	.08	.11	.17	.20	.15
Female	-.16	.10	-.11	.13	-.20	.17
	$\chi^2_{(1)}$ = 6, *p* = .01; G = 2.23, R = .83; .92 ≤ Infit/Outfit ≤ 1.08		$\chi^2_{(1)}$ = 1.3, *p* > .05;		$\chi^2_{(1)}$ = 3.2, *p* = .07; G = 1.48, R = .69; .71 ≤ Infit/Outfit ≤ 1.34	

The hypothesis about the status of the four approach indices entailed the idea that the approach-avoid continuum could be an overarching dimension which can be sub-framed when encountering a specific object towards or against which reacting. It was then expected a difference between the alcohol and soft-drinks approach indices. Further, in the light of the VPT results, the use of static and dynamic pictures in the AAT task would benefit from a check for their impact on the approach bias the indices elicit. The results across the general approach-avoid dimension and within the two drink types are presented in Table 5.3.

Table 5.11 Alcohol Approach (A) and non-alcohol approach (NA) indices estimates (δ) at baseline: general approach bias and alcohol/non-alcohol approach tendencies.

	Approach Bias		**Alcohol Approach**		**Non-alcohol Approach**	
Index	**δ**	***SE***	**δ**	***SE***	**δ**	***SE***
A-dynamic	.19	.13	.14	.14	-	-
A-static	-.19	.13	-.14	.14	-	-
NA-dynamic	-.14	.13	-	-	-.22	.16
NA-static	.14	.13	-	-	.22	.16
	$\chi^2_{(3)} = 7.2$, $p = .06$; G = 1.19, R = .58; .87 ≤ Infit/Outfit ≤ 1.15		$\chi^2_{(1)} = 5.7$, $p = .02$; G = 2.17, R = .83; .92 ≤ Infit/Outfit ≤ 1.03		$\chi^2_{(1)} = 4$, $p = .05$; G = 1.72, R = .75; .93 ≤ Infit/Outfit ≤ 1.01	

The four indices presented satisfactory fit indices in both the general approach-avoid latent trait and in the object-specific sub-dimension (.87 ≤ Infit/Outfit ≤ 1.15). As expected they are slightly different one to each other when taken all together ($\chi^2_{(3)} = 7.2$, $p = .06$) and well-distinguished when targeting both alcohol ($\chi^2_{(1)} = 5.7$, $p = .02$) and soft-drinks ($\chi^2_{(1)} = 4$, $p = .05$). It is evident from the measures estimates that, similarly to the VPT, the indices are paired one to each other according to the specific object they are intended for.

The MFRM analysis retrieved a consistent difference between the static and dynamic pictures used in the task. Automatic approach towards alcoholics and non-alcoholic appear to act differently according to the stimuli presented: when alcoholic dynamic pictures were presented in the approach format (pull) participants tended to be quicker in their approaching reactions ($\delta_{Ad} = .14$), when compared to the presentation of static pictures in the same trial format ($\delta_{As} = -.14$). The opposite occurred toward non-alcohol cues: the presentation of non-alcoholic static pictures in the approach format ($\delta_{NAs} = .22$) seems to be trigger a stronger prompt reaction to pulling them closer than the presentation of dynamic pictures ($\delta_{NAd} = -.22$).

The third step in the analysis of the AAT involved the similar examination of post-intervention datasets, with the final introduction of the experimental condition facet.

Table 5.12 Participants parameter estimates (β_v) at post-intervention: general approach bias and alcohol/non-alcohol approach tendencies.

	General Approch Bias		Alcohol Approch		Non-alcohol Approch	
Subject	**β**	***SE***	**β**	***SE***	**β**	***SE***
14*	2.65	1.82	2.18	1.88	3.24	1.91
18	2.26	1.00	3.69	1.89	2.10	1.14
35	2.22	1.00	1.81	1.09	4.03	1.91
6	1.89	0.59	1.51	0.83	2.65	1.14
5	1.75	0.99	1.48	1.09	3.06	1.91
38	1.35	0.71	1.30	0.83	4.34	1.91
16	1.07	0.71	2.81	1.88	0.65	0.91
8	**0.77**	**0.47**	**1.05**	**0.73**	**1.06**	**0.85**
24	0.74	0.47	1.59	1.09	-0.64	0.89
11	0.71	0.49	1.90	1.09	0.52	0.85
21*	0.70	0.53	2.27	1.09	-0.08	0.85
45	0.68	0.49	-0.51	0.73	1.77	0.91
23*	0.66	0.49	0.81	0.70	2.76	1.14
49	0.66	0.46	0.20	0.70	1.71	0.85
27	**0.60**	**0.49**	**-0.13**	**0.73**	**3.50**	**1.91**
4*	0.56	0.49	0.33	0.73	1.61	0.91
17	0.53	0.46	-0.40	0.70	0.83	0.85
2	0.51	0.59	2.27	1.88	0.44	0.85
9*	0.46	0.45	2.19	1.09	-1.79	1.21
19	0.45	0.49	0.29	0.70	1.70	1.14
1*	0.32	0.47	-0.27	0.82	0.82	0.85
40	0.24	0.49	0.10	0.73	1.04	0.91
31*	0.22	0.49	0.61	0.73	1.53	0.91
22	0.20	0.45	-0.06	0.73	0.67	0.85
10	0.16	0.46	-0.13	0.73	1.09	0.91
39	0.14	0.46	1.05	0.73	-1.28	0.99
44	0.06	0.53	0.85	1.09	0.04	0.85
33	-0.04	0.47	-1.33	0.82	0.33	0.85
42	-0.07	0.47	0.17	0.73	0.09	0.85
36	-0.08	0.47	0.03	0.73	-0.40	0.89
37	**-0.14**	**0.47**	**0.61**	**0.83**	**-0.83**	**0.85**
25	-0.15	0.54	-0.07	0.70	-3.22	1.94
7	-0.19	0.54	0.32	0.73	-1.96	1.21
51	-0.20	0.46	-0.42	0.70	0.27	0.85
52*	-0.20	0.46	-0.42	0.70	0.27	0.85
13	-0.26	0.46	0.89	0.73	-2.03	0.99
30	-0.34	0.45	-0.25	0.70	-0.23	0.85
20*	-0.41	0.45	-0.93	0.73	0.27	0.85
29	-0.41	0.45	-1.51	0.82	1.04	0.91
48	**-0.59**	**0.54**	**-0.56**	**0.82**	**-1.28**	**0.99**
47	**-0.63**	**0.73**	**-1.40**	**1.08**	**-1.88**	**1.21**
28*	-0.74	0.54	-0.65	0.82	-1.67	0.99
43*	-0.92	0.47	-0.34	0.70	-2.25	0.99
53*	-1.25	0.61	-1.14	1.08	-1.36	0.99
26	-1.32	0.73	-3.49	1.88	-1.29	0.99
32*	-1.40	0.61	-1.23	1.08	-1.75	0.99
15	-1.46	0.61	-0.22	0.73	-5.09	1.94
12	-1.51	0.73	-0.65	0.82	-4.34	1.94
46	**-1.75**	**0.61**	**-1.44**	**0.82**	**-3.41**	**1.21**
41	-1.76	0.73	-2.89	1.88	-2.03	0.99
50*	-2.58	1.84	-2.71	1.88	-3.39	1.94
34*	-2.87	1.84	-3.33	1.88	-3.22	1.94
3*	-3.62	1.85	-4.29	1.88	-4.01	1.94

* Participants estimates with misfitting Infit and Outfit statistics.

Participants' parameter estimates recovered at post-intervention are summarised in Table 5.12. At post-intervention participants still present a significantly diversified variety of approach bias strengths (β ranged from -3.62 to 2.65, $\bar{\beta}$ = -.05, $\overline{SE}$ = .65, *SD* = 1.21; $\chi^2_{(52)}$ = 111.2 *p* < .001), towards both alcohol (β ranged from -4.29 to 3.69, $\bar{\beta}$ = .02. $\overline{SE}$ = 1.00, *SD* = 1.58; $\chi^2_{(52)}$ = 78.2, *p* = .01) and soft-drinks (β ranged from -5.09 to 4.34, $\bar{\beta}$ = -.12, $\overline{SE}$ = 1.15, *SD* = .2.15; $\chi^2_{(52)}$ = 127.5, *p* < .01). Sixteen participants (8.48%) presented a misfit to the model requirements (i.e., Infit and Outfit greater than 2 or lower than .5).

When compared with baseline, at post-intervention males and females presented a reversed approach tendency, with men showing a stronger avoid bias particularly towards alcohol (η = -.31) ($\chi^2_{(1)}$ = 6.4, *p* < .01; see Table 5.13). No gender differences emerged towards non-alcoholics.

Table 5.13 Gender facet parameter estimates (η) at post-intervention for general approach bias and alcohol/soft-drinks approach tendencies.

	Approach Bias		**Alcohol Approach**		**Non-alcohol Approach**	
Gender	**η**	**SE**	**η**	**SE**	**η**	**SE**
Male	-.14	.09	-.31	.15	.09	.18
Female	.14	.12	.31	.20	-.09	.24
	$\chi^2_{(1)}$ = 3.5, *p* = .06; G = 1.59, R = .72; .98 ≤ Infit/Outfit ≤ 1.05		$\chi^2_{(1)}$ = 6.4, *p* < .01; G = 2.32, R = .84; .92 ≤ Infit/Outfit ≤ 1.05		$\chi^2_{(1)}$ = .3, *p* > .05;	

The four indices conserved a satisfactory fit to the model in the general approach-avoid latent trait – yet they still do no differentiate between more or less different approach-avoid tendencies ($\chi^2_{(3)}$ = 4.8, *p* = .19) – and in the soft-drink specific approach bias ($\chi^2_{(1)}$ = 8.2, *p* < .01) (see Table 5.14). That wasn't observed when the measurement dimension was sub-framed on the alcohol cue ($\chi^2_{(1)}$ = .5, *p* > .05): the effect of static and dynamic stimuli declined.

The stimulus effect still was visible on the soft-drinks approach bias, but reversed: soft-drink static pictures elicited more easily avoid reactions (push response) (δ_{NAs} = -.41) than dynamic pics presented in the same trial format (δ_{Ad} = .41).

Table 5.14 Alcohol Approach (A) and non-alcohol approach (NA) indices estimates (δ) at post-intervention: general approach bias and alcohol/non-alcohol approach tendencies.

	Approach Bias		**Alcohol Approach**		**Non-alcohol Approach**	
Index	**δ**	***SE***	**δ**	***SE***	**δ**	***SE***
A-dynamic	-.07	.15	-.08	.17	-	-
A-static	.06	.15	.08	.17	-	-
NA-dynamic	.22	.15	-	-	.41	.20
NA-static	-.22	.15	-	-	-.41	.20
	$\chi^2_{(3)} = 4.8$, $p = .19$; G = .77, R = .37; .70 ≤ Infit/Outfit ≤ 1.20		$\chi^2_{(1)} = .5$, $p > .05$;		$\chi^2_{(1)} = 8.2$, $p < .01$; G = 2.67, R = .88; .98 ≤ Infit/Outfit ≤ 1.00	

Table 5.15 presents the parameter estimates for the facet experimental condition in the approach-avoid dimension underlying the AAT. The four groups did significantly distinguish one to each other in the strength of the general approach-avoid bias they bring in ($\chi^2_{(3)} = .14$, $p < .01$). However, the effect appears to be mixed between the approach-avoid tendencies towards both alcoholics and non-alcoholics. It is evident that the experimental condition in which participants receive the real intervention (i.e., VPT and AAT retrainings) carries off a decrease in the approach bias towards alcohol by shifting it towards the soft-drinks. This trend emerges across the general and object-specific approach tendencies (γ_As = [-.41, -.37]). Interestingly, also the combination of VPT-retraining and AAT-placebo showed a parameter estimate – although not statistically significant – pointing to a reversed bias towards alcohol. Finally, the worse case involves the allocation to the double placebo condition, where the approach bias towards alcohol is still strong.

For illustrative purposes, a graphical representation of the participants evidenced in bold in Tables 5.9 and 5.12 is attached below. Participants' pre- and post-intervention general approach bias estimates and estimates for alcohol approach and soft-drinks approach are plotted in *Figure 5.2*.

Table 5.15 Experimental condition parameter estimates (γ) for general approach bias and alcohol/non-alcohol approach tendencies.

Condition	Approach Bias γ	Approach Bias *SE*	Alcohol Approach γ	Alcohol Approach *SE*	Non-alcohol Approach γ	Non-alcohol Approach *SE*
A	-.37	.15	-.21	.22	-.41	.29
B	.03	.14	-.05	.22	.34	.27
C	-.04	.17	-.48	.27	-.32	.35
D	.38	.14	.74	.24	.39	.25
	$\chi^2_{(3)} = 14$, $p < .01$;		$\chi^2_{(3)} = 13.5$, $p < .01$;		$\chi^2_{(3)} = 6.5$, $p = .09$;	
	G = 1.81, R = .77;		G = 1.95, R = .79;		G = 1.04, R = .52;	
	.72 ≤ Infit/Outfit ≤ 1.24		.60 ≤ Infit/Outfit ≤ 1.23		.74 ≤ Infit/Outfit ≤ 1.25	

A = double retraining; B = VPT placebo and AAT retraining; C = VPT retraining and AAT placebo; D = double placebo.

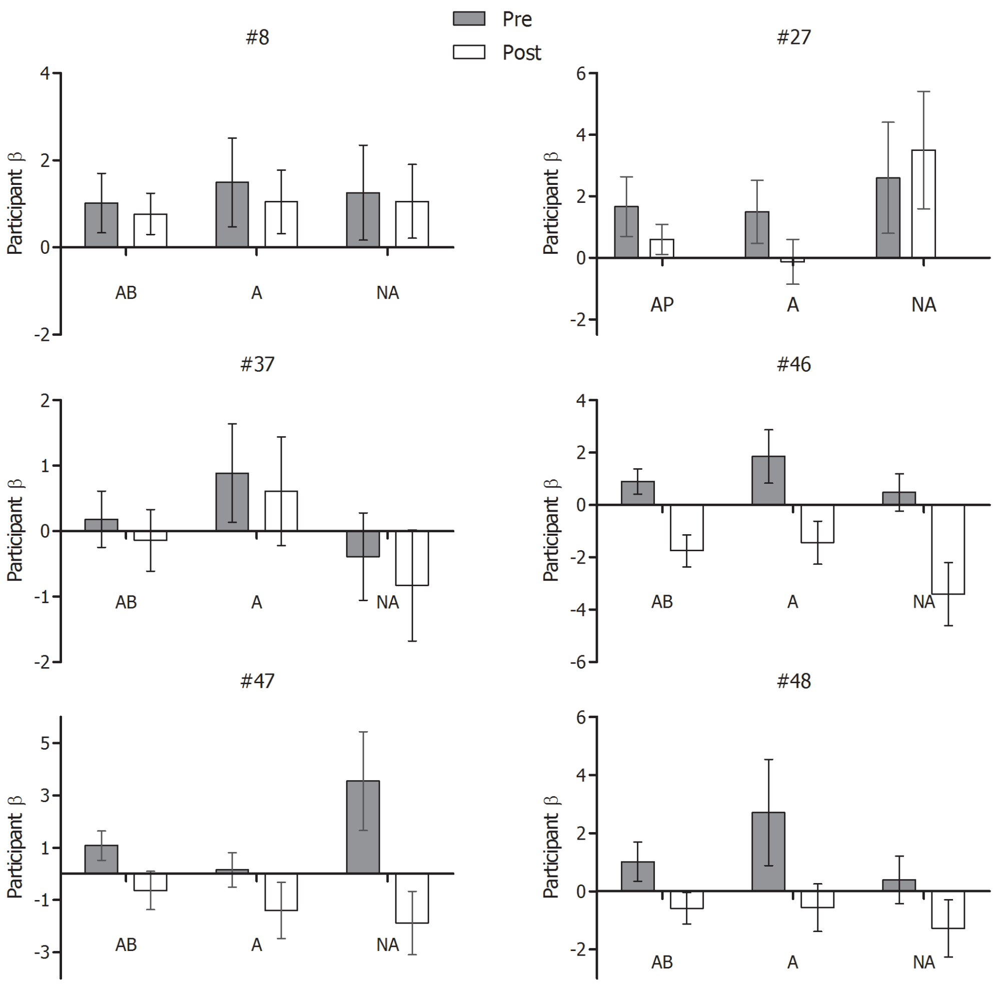


Figure 5.2 Approach bias estimates (β) at pre- and post-intervention for participants with a strong alcohol approach bias at baseline (subjects in bold in Tables 5.9 and 5.12): general approach-avoid bias (AB), alcohol approach (A) and non-alcohol approach (NA). Higher values indicate stronger approach bias towards alcohol.

5.5 Discussion

The present study covered the exploration of preliminary data collected within the experimentation of a combined retraining program targeting dysfunctional impulsive processes in alcohol addiction, namely the automatic approach tendencies and the attentional bias towards alcohol. The combination of two CBM retraining paradigms in the treatment of the cognitive biases involved in the alcohol addiction disorder was experimented for the first time on a group of alcohol dependent outpatients, with the main objective of answering the question whether the joint exposure to a double treatment intervention of implicit processes could be beneficial or not.

The application of a rigorous modelling framework, i.e., the MFRM Rasch model, allowed examining several issues and hypotheses implied in the present study, first of all the proof of the ontological status of the attributes the measures are targeting.

A standard discussion would summarise the results obtained and comment them in the light of previous studies and theoretical literature. However, it might be more effective to start from a general surmise standing above all the present research and acknowledge the results in the light of the current limitations.

It is a compelling enterprise, and somewhat premature, to draw definite conclusions from the corpus of results obtained so far, for three main reasons: first, the explorative nature of the current study limits its generalizability to an all or nothing conclusion. The group of participants used is only a part of the targeted sample for the RCT completion. Therefore, what we have found by now should be considered a work-in-progress result, which, nonetheless, let the reader and the researcher to foresee what actually is occurring and what possible directions and forms the final outcome could take. Second, the participants here examined are still on their way, since a Time 3 measurement session (i.e., follow-up session at three months, cf. Chapter 4) of their cognitive biases towards alcohol is missing. Therefore, the presumed changing progress they should undergo hasn't probably reached a final point.

Third, a changing process can take many different forms. For instance, it could be stage-alike or gradual, or a combination in time of the two of them. The results achieved so far point to the fact that participants did not substantially change from the pre- to the post-intervention measurement session, as the facet time didn't have any effect on the initial measures of attentional and approach bias towards alcohol. However, the separate analyses of participants' data in the two time points evidenced that something is indeed occurring and that, at the group-level, *there is* an on going, subtle changing progress.

In Table 5.15 the main results are summarised. At pre-test participants showed similar levels of alcohol attentional and approach bias, indicating that the strength of the automatic processes towards the addiction substance was similar across participants. That could be considered an expected result given the study inclusion criterion of having a main diagnosis of alcohol dependence disorder. Substance use itself has bee found to strengthen impulsive reactions to drug-related cues and weaken cognitive and executive control over the impulses (e.g., Wiers et al., 2013).

Of interest are the cognitive biases estimates at post-intervention assessment point: the model signalled that participants' parameter estimates were slightly different for both biases, indicating that something has started to change between the two time points. At group-level the direction of change did not substantially emerge. However, by looking at the measures of participants with a strong bias at baseline and plotting them against post-test estimates (see *Figure 5.1* and *5.2*), it is graphically visible the change in the direction of bias.

This qualitative observation resembles what has been found in two recent studies targeting alcohol approach bias only (e.g., Eberl et al., 2012; Wiers et al., 2011): patients who benefited most from the experimental training were those with the strongest initial bias, evidencing a moderating role of the initial cognitive bias level on the effects of the training intervention.

At pre-test, males showed generally higher levels of attentional bias and unspecific and soft-drinks-related approach bias. Approach automatic reactions towards alcohol resulted to be gender-independent. At post-intervention this difference disappeared for the attentional bias, whereas it was reversed towards the avoid bias, both in domain-general and alcohol-related approach-avoid tendencies.

Table 5.16 Summary of main results for attentional (VPT) and approach bias (AAT) at pre- and post-intervention for each MFRM facet.

	VPT		AAT	
FACET	**Pre**	**Post**	**Pre**	**Post**
Time	No differences emerged.		No differences emerged.	
Subject Measures	Similar levels of alcohol attentionl bias. Slightly more variability in disengagement processes.	Similar levels of alcohol attentional bias in all its sub-components.	Different general and soft-drinks approach-avoid tendencies but similar levels of alcohol-specific approach bias.	Very different levels of approach-avoid tendencies, both domain-general and object-specific.
Gender	Males present a slightly stronger general attentional bias, apart from the engagement and disengagement processes.	No differences emerged.	Males present stronger general-domain and soft-drinks-specific approach tendencies. No gender differences in the alcohol approach bias.	Males present a marginal domain-general avoid bias and a stronger avoid bias towards alcohol. Women shifted towards an approach bias. No differences in soft-drinks approach- bias.
Task Indices	Indices measure similar attentional bias aspects. They are paired according to the attentional process implied. Alcohol pics in static context elicit stronger attentional bias for disengagement. No difference between the stimuli contexts for engagement processes.	Indices still measure similar attentional bias aspects and are paired according to the attentional process implied. Alcohol pics in static context elicit stronger attentional bias for engagement in the task. No difference between the stimuli contexts for disengagement processes.	Indices entail both general-domain and object-specific approach-avoid tendencies. They work in pairs according to the targeted object.Static alcohol stimuli trigger more alcohol avoid tendencies than dynamic pics, whereas static non-alcohol pics elicited stronger soft-drinks approach tendencies.	Indices still entail both general-domain and object-specific approach-avoid tendencies and work in pairs. The stimuli context does not impact on the alcohol approach-avoid tendencies, whereas static soft-drinks pics elicit stronger avoid bias towards non-alcoholics.
Experimental condition	-	Marginally difference in the elicited attentional bias when disengagement processes are implied: double training triggers a reversal of the bias towards soft-drinks, whereas double placebo displays the opposite.	-	Differential effect on domain-general and alcohol-specific approach-avoid tendencies: double placebo prompts the strongest avoid bias, followed by VPT real/AAT placebo. The double placebo brings in steady domain-general and object-specific approach tendencies.

The examination of the experimental condition effects revealed that the double real intervention outperformed the other three conditions in reversing the positive bias towards alcohol. If we locate the four experimental conditions on a 'success' continuum we would find the double real at the 'very successful' pole of the continuum and the double placebo at the opposite side. The mixed placebo/retraining conditions stay in between, with a surprising VPT-real/AAT-placebo condition also doing moderately well in reducing the approach bias towards alcohol, even though the placebo version of the approach bias retraining is administered. It would be interesting to check for the order of presentation of the two tasks, which is counterbalanced between participants and fixed within participants, to see whether any effect of completing the real attentional training first generalises or not over the approach bias level, by surpassing the hypothesised null or moderate effect of the AAT-placebo.

I chose to comment the analysis of the four indices for each task as a final reflection on what the two implicit measures do actually *measure* and how the two tasks do function. One of the issues in Cognitive Bias Modification (CBM) research is that usually the most indirect measures to assess the cognitive biases, such as the AAT and the VPT, are the most easily adjusted for retraining because of their task-irrelevant feature, i.e, participants react to a different feature of the stimulus, and because contingencies of the stimuli can be manipulated without changing the instructions. However, the paradox is that, at the same time, these measures appear to be suboptimal for assessment, due to a suboptimal reliability (Ataya et al., 2012; De Houwer & De Bruycker, 2007; Field, Caren, Fernie, & De Houwer, 2011). A second issue refers to the processes they are measuring (and training): whether they are domain-general or context-specific. Theoretical perspectives on the training control over domain-general capacities or in relation to a specific domain, where it is triggered by specific stimuli (such as in the CBM), suggest that it would be different to train, for instance, alcoholic patients with fear pictures and anxious patients with alcohol pictures, because an important aspect of training is that control processes are activated in time by stimuli related to the problem domain (Wiers et al., 2013).

To this end, the MFRM analysis can probably give us some clues, which can be better understood by keeping in mind that Rasch models are unidimensional. When analyzing the four indices for each task a double hypothesis guided the process: both implicit measures tap into a general domain, which can be subsequently framed into specific sub-components, analyzable separately and with an own ontological status. For instance, the VPT is an implicit measure of the attentional bias towards whatever object and is composed by both attentional processes of engagement and disengagement. Likewise, the AAT can be an index of general approach-avoid automatic tendencies towards whatever object, but depending on the specific stimuli presented the measurement dimension does specify into a domain-specific approach bias. It follows that alcohol approach tendencies and soft-drinks approach tendencies are also two separate measurement dimensions and can be affected differently by the training intervention.

In the data collected so far, it was systematically found that the four indices are part of the same dimension, but they do not differentiate one to each other and do measure similar aspects of the underlying dimension. When the four were split in pairs, things changed and emerged the different elements composing the cognitive bias. A further specification came from the differences obtained by the usage of static and dynamic stimuli in the two tasks. The two contexts in which alcohol and non-alcohol drinks were presented produced an effect on the performance of the tasks. For instance, in the VPT, alcohol static pictures were more difficult to disengage from at baseline, while at post-intervention they activated a stronger engagement in participants' attention. Similarly, in the AAT, alcohol dynamic pictures elicited quicker pull closer responses at baseline, while at post-intervention this difference disappeared. These different patterns of results related to the type of stimuli used to highlight the domain-specific features described above and add another feature, the presence or absence of a human being in interaction with the drink, which is likely to activate different impulsive reactions and control-related processes. This first evidence calls for further investigation about any mediating or moderating role of stimuli features eliciting different automatic responses.

Altogether, the explorative nature of the present study gave us precious and informative clues about what is going on and where it is going, particularly in relation to the different roles played by the each cognitive bias components and specific features. Several theoretical points still need to be cleared up and reach the “proof of principle” stage. It would be really interesting to see how the 53 participants behave at follow-up session, whether their changing process took a precise form or is still on the run and whether they therapeutic progress is (hopefully) on the right way. But just like all changing processes, it takes time.

CONCLUSION

The introduction section of this work closes with a general, inclusive acknowledgement of what the present work brought about in the research pathway I walked through, passing by measurement theory, implicit measurement, mental health, experimental clinical psychology, and modeling perspectives. It would seem that these elements are a kind of melting pot of different domains and backgrounds, just mixed together without a coherent guideline. Actually, they are different domains and have different backgrounds. Nonetheless, they have in common one, unique and core element: the human mind.

This work started with a reflection on what it means to measure psychological variables. The measurement process in psychology is inherently connected to the scientific research practice, but also to the clinical practice. Just think about a therapist evaluating the extent to which the patient's therapeutic progress is positive: how does (s)he expresses a judgment without any measurement unit to which it is possible to compare the object of the evaluation? The evaluation would be rather subjective and susceptible to any kind of personal and extra-personal bias by the therapist, the context, and the patient herself/himself. Although we may not be always aware of and deliberately involve in, the operation of *measuring* is constantly part of our daily life.

The measurement problem in psychology was one of the key-stones around which this work developed. In the first chapter the conception of measurement in psychology and the related criticisms raised along different psychological mainstreams have been presented, with the purpose of giving an idea of the guiding reflection that brought to life the "melting pot" of the above mentioned elements.

Among the enormous amount of measurement instruments designed along the history of psychology, a relatively recent family of methods has been receiving, and still is receiving, a great interest: the *implicit measures*.

Implicit measures were conceived and designed to target psychological processes and/or attributes that more explicit measures couldn't index or weren't optimal for, because of inherent features of the to-be-measured psychological attributes that limited the reach of traditional measurement instruments. These 'unreachable' psychological attributes are indeed considered as automatic, uncontrollable, unconscious, efficient, effortful, and implicit. Several dual-process models of human cognition have deepen the theoretical conceptualization of the human mind architecture by positing the existence of parallel information processing systems that work jointly in decision making processes, and in the expression and regulation of human behaviours (e.g, Strack & Deutsch, 2004). One of the two systems, the impulsive system, involves these implicit processes, which play a role in subtly triggering behaviour. Implicit measures have been developed in the attempt to catch them or, more precisely, to measure them up.

The widespread use of these measures across a variety of domains in psychology inoculated a question: are these methods so flexible and usable to measure almost everything? If yes, how can they achieve it and how do they work?

That was the starting point of this research project.

The research process has involved the experimentation of several implicit measures in two different contexts within the broader domain of mental health, which have in common the involvement of supposedly automatic, implicit, involuntary, and uncontrollable psychological processes: the automatic components of stigmatizing attitudes and behaviours *towards* people affected by a mental disease (Part 1) and the impulsive, automatic processes implied *within* people affected by a mental disease, more specifically, by an alcohol addiction disorder (Part 2).

Stigmatising attitudes and behaviours towards mentally ill people are difficult to assess, given the particular social minority group and the strong resistance and ambivalence in expressing a prejudicial attitude or behaviour towards someone who is acknowledged as being suffering but nonetheless can elicit negative reactions, for the interaction of stereotypes, beliefs, more or less thoughtful evaluations, and cultural and social norms. The assessment of the various facets of stigma is then a quite challenging issue in stigma research and further poses several theoretical questions about their nature. Two Implicit Association Tests targeting two aspects of mental illness stigma, namely, aetiological beliefs and attitudes have been then designed with a two-fold objective. On one hand, to verify whether they could be used as assessment techniques in this particular framework, on the other hand, to explore the plausible existence of implicit complements of mental illness stigma.

The second line of research called for a additional aspect, which doubled the research topic: the experimentation of implicit measurement techniques as means for change, by adapting them to the function of training of those implicit processes they were initially designed to assess. A Randomised Clinical controlled Trial (RCT) with alcohol dependent outpatients has been then designed and implemented to evaluate the combination of two training paradigms targeting the automatic attentional processes and approach tendencies towards alcohol, by using the same implicit measures for the assessment of the strength of cognitive biases towards alcohol as a way to reduce them, or at least reversing them towards an aversive cue as salient as alcohol (non-alcoholics).

At that point, the set-up for two different observations of the functioning of implicit measures was ready.

The step further has been to plan an analytical strategy that could provide a precise insight into the measures properties, but also into the theoretical underpinnings of that measure. In both studies, the measurement properties of the implicit measures developed and their meaning in relation to the theoretical to-be-measured psychological attributes have been explored within a Rasch modelling perspective, through the application of the Many-Facet Rasch Measurement (MFRM) model.

The application of the MFRM model allowed disentangling the different 'ingredients' contributing to the emergence of the implicit, associative effect in both the IAT for semantic associations with mental illness and in the IAT for the evaluative associations towards mental illness. The model revealed how implicit aetiological beliefs and evaluative associations with mental illness are indeed multifaceted aspects. Semantic implicit associations with mental illness resulted to be dependent on the diagnostic categories presented and automatically determined by the differential endorsement of biologic semantic associations and not by more psychological-related associations. Conversely, evaluative associations towards mental illness presented a reversed pattern of "positive association primacy", which pointed to the fact that evaluative associations aren't triggered by neither positive associations or by negative associations towards mental illness. This result calls for the problem of the relative nature of the IAT, which indeed the MFRM highlighted, and warrants the investigation of similar associations by indexing their distinct absolute component.

Further, the MFRM evidenced the functioning of the IAT at the microscopic level, by evidencing those stimuli that mostly contributed in eliciting the IAT effect and triggering the hypothesised associations. The main advantage of these results stay in the possibility of accurately examine what is the best to test a researcher's hypothesis, for the differential effect of the stimuli used, as signalled in the two IATs, can lead to very different results.

The application of the MFRM to the analysis of a group of participants at pre- and post-intervention assessment sessions gave an interesting cluster of first promising results about the efficacy of the experimental treatment intervention. Although participants did not show a substantial change in their alcohol attentional and approach bias measures, the MFRM evidenced a changing process in action. The four experimental conditions showed to have a differential effect in bringing in a decrease and/or a reversal of the two cognitive biases, with the double real training condition outperforming the others. The MFRM contributed to the exploration of the dimensional and theoretical status of the two cognitive bias implicit measures and provided several informative clues about their general and domain-specific features. The main hint is that bottom-up cognitive processes, domain-general and domain-specific are probably simultaneously present in the expression of bias and, more importantly, are differently affected by the training intervention. A final interesting note goes to the discovery of a substantial effect of the stimuli used in improving control processes over the impulsive reactions towards alcohol, with reversed patterns from pre- to post-intervention measurement sessions.

These results are to be taken cautiously, for two main reasons: first, the explorative and preliminary nature of the study and the still on-going changing process implied in the treatment process limit the possibility of drawing definite conclusions. Second, the application of the MFRM is *one* of the possible strategies in the attempt to get a *representation* of what is going on. This model demonstrated to be a powerful tool in showing both what was expected and what instead was surprising. The enterprise of giving a meaning to what the model tells you put you and your hypotheses to test. Like any longitudinal experimental study, time is a key-variable in the research process, therefore the only thing to do now is probably to wait and, in the light of what has already showed off, think about the possible ends.

I like symmetries. Therefore, I would like to conclude this work in circle, by expressing a thought similar to what I've been writing at the end of the introduction. Doing research in psychological science is difficult. Many variables, many confounders, many things interconnected, intertwined, and dependent one on each other. And you can never disregard the random error component, which, sometimes, messes up everything. However, in the "melting pot" of this work the final achievement was to open a window on the psychological processes of interest with the confidence that this window is quite stable and big enough to let a good sight on what stays outside. And this is the ultimate outcome of the entire process, which, in my opinion, is the conquest when doing psychological research.

References

Abdullah, T., & Brown, T.L. (2011). Mental illness stigma and ethnocultural beliefs, values, and norms: An integrative review. *Clinical Psychology Review, 31*, 934-948.

Angermeyer, M.C., & Dietrich, S. (2006). Public beliefs about and attitudes towards people with mental illness: a review of population studies. *Acta Psychiatrica Scandinavica, 113*, 163-179.

Angermeyer, M.C., Holzinger, A., Carta, M.G., & Schomerus, G. (2011). Biogenetic explanations and public acceptance of mental illness: systematic review of population studies. *The British Journal of Psychiatry, 199*, 367-372.

Angermeyer, M.C., & Matschinger, H. (2005). Causal beliefs and attitudes to people with schizophrenia: trend analysis based on data from two population surveys in Germany. *The British Journal of Psychiatry, 186*, 331-334.

Anselmi, P., Vianello, M., & Robusto, E. (2011). Positive associations primacy in the IAT. A Many-Facet Rasch Measurement analysis. *Experimental Psychology, 58*, 376-384.

Anselmi, P., Vianello, M., Voci, A., & Robusto, E. (2013). Implicit sexual attitude of heterosexual, gay and bisexual individuals: Disentangling the contribution of specific associations to the overall measure. *Plos ONE, 8*(11), e7899. Doi:10.1371/journal.pone.0078990

Anton, R. F. (2000). Obsessive-compulsive aspects of craving: development of the Obsessive Compulsive Drinking Scale. *Addiction, 95,* Suppl 2, S211-217.

Ataya, A.F., Adams, S., Mullings, E., Cooper, R.M., Attwood, A.S., & Munafo, M.R. (2012). Internal reliability of measures of substance-related cognitive bias. *Drug and Alcohol Dependence*, *121*, 148-151.

Babor, T.F., Higgins-Biddle, J.C, & Saunders, J.B. (2001). *The Alcohol Use Disorders Identification Test. Guidelines for Use in Primary Care* (2nd Ed.). World Health Organization; Department of Mental Health and Substance Dependence.

Maristela G. MonteiroBanaji, M.R., & Hardin, C.D. (1996). Automatic stereotyping. *Psychological Science, 7*, 136-141.

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

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

Bargh, J.A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R.S. Wyer & T.K. Srull (Eds.), *Handbook of social cognition* (pp. 1-40). Hillsdale, NJ: Erlbaum.

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist, 32*, 169-177.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. *Nature neuroscience*, *8*, 1458–1463.

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *BDI–II, Beck Depression Inventory: Manual* (2nd Ed.). Boston: Harcourt Brace.

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A. (1995). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*. http://www.ge.ilc.cnr.it/page.php?ID=archCoLFIS&lingua=it

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics redux. *American Psychologist, 61*, 62-71.

Boffo, M., & Mannarini, S. (2012). Pregiudizio e credenze eziologiche dei disturbi mentali Misurazione implicita ed esplicita. [Prejudice and aetiological beliefs about mental illness. Implicit and explicit measurement]. Oral communication presented at the *20th National Congress of Italian Association of Psychology (AIP)*, section of Experimental Psychology, September 20-23, Chieti, Italy.

Boffo, M., Mannarini, S., & Wiers, R.W. (2013). TOP (Training op Onbewuste Processen) training: Retraining automatic cognitive processes in alcohol addict outpatients. *Current Controlled Trials.* Doi:10.1186/ISRCTN01005959

Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model. Fundamental measurement in human sciences. Second edition.* Mahwah: Erlbaum Associate.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press.

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

Borsboom, D., & Mellenbergh, G.J. (2004). Why psychometrics is not pathological. *Theory & Psychology, 14*, 105–120.

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

Bos, A.E.R., Pryor, J.B., Reeder, G.D., & Stutterheim, S.E. (2013). Stigma: Advances in theory and research. *Basic and Applied Social Psychology*, *35,* 1-9.

Boysen, G.A. (2011). Biological explanations and stigmatizing attitudes: using essentialism and perceived dangerousness to predict antistigma intervention effectiveness. *The Journal of Social Psychology, 151*, 274-291.

Brener, L., Rose, G., von Hippel, C., & Wilson, H. (2013). Implicit attitudes, emotions, and helping intentions of mental health workers toward their clients. *The Journal of Nervous and Mental Disease, 201*, 460-463.

Bridgman, P.W. (1927). *The logic of modern physics*. New York: Macmillan.

Brohan, E., Slade, M., Clement, S., & Thornicroft, G. (2010). Experience, of mental illness stigma, prejudice and discrimination: a review of measure. *BMC Health Services Research, 10*, 80. http://www.biomedcentral.com/1472-6963/10/80

Butler, L.T., & Berry, D.C. (2001). Implicit memory: Intention and awareness revisited. *Trends in Cognitive Science, 5,* 192–197.

Cameron, C.D., Brown-Iannuzzi, J.L., & Payne, B.K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review, 16*, 330-350.

Centre for Substance Abuse Treatment (1999). *Enhancing Motivation for Change in Substance Abuse Treatment. Treatment Improvement Protocol (TIP) 35.* Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment. Retrieved from http://motivationalinterview.org/

Chen, M., & Bargh, J.A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25,* 215-224.

Corrigan, P.W. (2000). Mental health stigma as social attribution: implications for research methods and attitude change. *Clinical Psychology: Science & Practice, 7*, 48-67.

Corrigan, P.W. (2004). How stigma interferes with mental health care. *AmericanPsychologist, 59*, 614-625.

Corrigan, P.W. (2007). How clinical diagnosis might exacerbate the stigma of mental illness. *Social Work, 52*, 31–39.

Corrigan, P.W., Tsang, H.H., Shi, K., Lam, C.S., & Larson, J. (2010). Chinese and American employers' perspectives regarding hiring people with behaviorally driven health conditions: The role of stigma. *Social Science & Medicine*, *71*, 2162-2169.

Cousijn, J., Goudriaan, A.E., & Wiers, R.W. (2011). Reaching out towards cannabis: Approach-bias in heavy cannabis users predicts changes in cannabis use. *Addiction, 106*, 1667-1674.

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

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

De Houwer, J., & De Bruycker, E. (2007). The implicit association test outperforms the extrinsic affective Simon task as an implicit measure of inter-individual differences in attitudes. *British Journal of Social Psychology*, *46*, 401–421.

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

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski, & B.K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp.176-193). New York, NY: Guilford Press.

De Houwer, J., & Moors, A. (2012). How to define and examine implicit processes? In R. Proctor & J. Capaldi (Eds.). *Implicit and explicit processes in the psychology of science* (pp. 183-198). NY: Oxford University Press.

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

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009b). Theoretical claims neccesitate basic research: Reply to Gawronski, LeBel, Peters, and Banse (2009) and Nosek and Greenwald (2009). *Psychological Bulletin, 135*, 377-379.

de Jong, P.J., & Veenstra, E. (2007). *Automatic approach–avoidance tendencies and dysfunctional eating patterns.* Paper presented at the Fifth World Congress of Behavioural and Cognitive Therapy, Barcelona, Spain.

Deutsch, R., & Strack, F. (2006). Duality Models in Social Psychology: From Dual Processes to Interacting Systems. *Psychological Inquiry, 17*, 166–172.

Devilly, G.J., & Borkovec, T.D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 31*, 73-86.

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

Eberl, C., Wiers, R.W., Pawelczacka, S., Rinck, M., Becker, E.S., & Lindenmeyer, J. (2013). Approach bias modification in alcohol dependence: Do clinical effects replicate and for whom does it work best? *Developmental Cognitive Neuroscience, 4*, 38-51.

Fadardi, J.S., & Cox, W.M. (2009). Reversing the sequence: reducing alcohol consumption by overcoming alcohol attentional bias. *Drug and alcohol dependence*, *101*, 137–145.

Fazio, R.H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology, 23,* 75-109.

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

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.

Fechner, G.T. (1860). Elemente der psychophysik. Leipzig: Breitkopf & Hartel. (English translation by H. E. Adler, Elements of Pychophysics, vol. 1, D. H. Howes & E. G. Boring (Eds). New York: Holt, Rinehart & Winston.)

Feldman, D.B., & Crandall, C.S. (2007). Dimensions of mental illness stigma: What about mental illness causes social rejection? *Journal of Social and Clinical Psychology, 26,* 137-154.

Ferguson, A., Myers, C., Bartlett, R., Banister, H., Bartlett, F., Brown, W., … & Tucker, W. (1940). Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *Report of the British Association for the Advancement of Science, 2,* 331-349.

Field, M., Caren, R., Fernie, G., & De Houwer, J. (2011). Alcohol approach tendencies in heavy drinkers: Comparison of effects in a relevant stimulus-response compatibility task and an approach/ avoidance Simon task. *Psychology of Addictive Behaviors*, *25*, 697–701.

Gaertner, S.L., & McLaughlin, J.P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23-30.

Gawronski, B., & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731.

Gawronski, B., & De Houwer, J. (in press). Implicit measures in social and personality psychology. In H.T. Reis, & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd edition). New York: Cambridge University Press.

Gawronski, B., Sherman, J.W., Trope, Y. (in press). Two of What? A Conceptual Analysis of Dual-Process Theories. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.). *Dual-process theories of the social mind*. New York, NY: Guilford Press.

Gawronski, B. & Sritharan, R. (2010). Formation, change, and contextualization of mental associations. Determinants and principle of variations in implicit measures. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 216-240). New York, NY: Guilford Press.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40,* 535-542.

Gawronski, B., & Ye, Y. (2014). What drives priming effects in the affect misattribution procedure? *Personality and Social Psychology Bulletin, 40*, 3-15.

Ghisi, M., Flebus, G.B., Montano, A., Sanavio, E., & Sica, C. (2006). *Adattamento italiano di A.T. Beck, R.A. Steer, G.K. Brown (1996). BDI-II, Beck Depression Inventory –II*. Firenze: Organizzazioni Speciali.

Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, NJ: Prentice Hall.

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

Greenwald, A.G., Mcghee, D.E., & Schwartz, J.L.K. (1998). Measuring Individual Differences in Implicit Cognition. *Journal of Personality*, *74*, 1464-1480.

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

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.

Grenard, J.L., Ames, S.L., Wiers, R.W., Thush, C., Sussman, S., & Stacy, A.W. (2008). Working memory capacity moderates the predictive effects of drug-related associations on substance use. *Psychology of addictive behaviors*, *22*, 426-432.

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin*, *137*, 940–958.

Hinshaw, S. (2006). *The mark of shame: Stigma of mental illness and an agenda for change*. New York: Oxford University Press.

Hofmann, W., Friese, M., & Wiers, R.W. (2008). Impulsive versus reflective influences on health behavior: a theoretical framework and empirical review. *Health Psychology Review*, *2*, 111-137.

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

Holder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass. *Berichte iiber die Verbandlungen der Königlich Sächsischen Gesellschaft der Wissenschajten zu Leipzig, Mathematisch-Physische Klasse, 53*, 146.

Houben, K., & Wiers, R.W. (2009). Response inhibition moderates the relationship between implicit associations and drinking behavior. *Alcoholism: Clinical and Experimental Research*, *33*, 626-633.

Janiri, L., Calvosa, F., Dario, T., Pozzi, G., Ruggeri, A., Addolorato, G., …, & De Risio, S. (2004). The Italian version of the Obsessive-Compulsive Drinking Scale:

validation, comparison with the other versions, and difference between type 1- and type 2-like alcoholics. *Drug and alcohol dependence, 74*, 187-195.

Jones, E.E., Farina, A., Hastdorf, A., Markus, H., Miller, D., & Scott, R. (1984). *Social stigma: The psychology of marked relationships*. New York: Freeman.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697-720.

Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement, 2*, 389-423.

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.

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 Implicit Association Test Effect. *Personality & Social Psychology Bulletin, 31*, 208-217.

Kvaale, E.P., Gottdiener, W.H., & Haslam, N. (2013). Biogenetic explanations and stigma: A meta-analytic review of associations among laypeople. *Social Science & Medicine, 96*, 95-103.

Kvaale, E.P., Haslam, N., & Gottdiener, W.H. (2013). The 'side effects' of medicalization: A meta-analytic review of how biogenetic explanations affect stigma. *Clinical Psychology Review, 33*, 782-794.

Kyngdon, A. (2011). Plausible measurement analogies to some psychometric models of test performance. *British Journal of Mathematical and Statistical Psychology, 64*, 478-497.

Lee, A.A., Laurent, S.M., Wykes, T.L., Kitchen Andren, K.A., Bourassa, K.A., & McKibbin, C.L. (in press). Genetic attributions and mental illness diagnosis: effects on perceptions of danger, social distance, and real helping decisions. *Social Psychiatry and Psychiatric Epidemiology.* Doi:10.1007/s00127-013-0764-1

Linacre, J.M. (1989). *Multi-facet Rasch measurement.* Chicago, IL: MESA Press.

Linacre, J.M. (2009). *Facets Rasch measurement computer program (Version 3.66.0)* [Computer software]. Chicago, IL: Winsteps.com.

Linacre, J.M. (2010). *FACET Rasch Model Computer Program: Users Guide.* [Computer software manual]. Chicago, IL: Winsteps.com.

Lincoln, T.M., Arens, E., Berger, C., & Rief, W. (2008). Can Antistigma Campaigns Be Improved? A Test of the Impact of Biogenetic Vs Psychosocial Causal Explanations on Implicit and Explicit Attitudes to Schizophrenia. *Schizophrenia Bulletin, 34*, 984-994.

Link, B.G., Struening, E.L., Neese-Todd, S., Asmussen, S., & Phelan, J.C. (2001). Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people with mental illnesses. *Psychiatric Services*, *52*, 1621-1626.

Link, B.G., Yang, L.H., Phelan, J.C., & Collins, P.Y. (2004). Measuring mental illness stigma. *Schizophrenia Bulletin, 30*, 511-541.

Livingston, J.D., & Boyd, J.E. (2010). Correlates and consequences of internalized stigma for people living with mental illness: A systematic review and meta-analysis. *Social Science & Medicine*, *71*, 2150-2161.

Luce, R.D., & Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1,* 1-27.

MacLeod, C.M. (2012). Cognitive bias modification procedures in the management of mental disorders. *Current Opinion in Psychiatry, 25*, 114-120.

MacLeod, C.M., Mathews, A.M., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15-20.

MacLeod, C.M., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*, *111*, 107-123.

Mannarini, S., & Boffo, M. (2013a). Assessing mental disorders causal beliefs: A latent dimension identification. *Community Mental Health Journal, 49*, 686-693.

Mannarini, S., & Boffo, M. (2013b). Beliefs about the patient's role in the psychotherapeutic relationship: A latent trait approach. *TPM: Testing, Psychometrics, Methodology in Applied Psychology, 20*, 277-294.

Markus, K.A., & Borsboom, D. (2012). The cat came back: Evaluating arguments against psychological measurement. *Theory Psychology, 22*, 452-466.

McClelland, D.C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Bulletin, 96,* 690-702.

Mellenbergh, G.J. (1994). Generalized linear item response theory. *Psychological Bulletin, 115*, 300-307.

Mehta, S., & Farina, A. (1997). Is being 'sick' really better? Effect of the disease view of mental disorder on stigma. *Journal of Social and Clinical Psychology, 16*, 405–419.

Menatti, A., Smyth, F. L., Nosek, B. A., & Teachman, B. A. (2012). Reducing stigma toward individuals with mental illnesses: A brief, online intervention. *Stigma Research and Action*.

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.

Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NY: Lawrence Erlbaum Associates.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*, 355-383.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept.* Cambridge, UK: Cambridge University Press.

Michell, J. (2008). Is Psychometrics Pathological Science? *Measurement: Interdisciplinary Research and Perspectives, 6,* 7-24.

Miller, W.R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people to change addictive behavior (2nd Ed.)*. New York, NJ: Guilford Press. First published in 1991.

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

Monteith, L.L., & Pettit, J.W. (2011). Implicit and Explicit Stigmatizing Attitudes and Stereotypes About Depression. *Journal of Social and Clinical Psychology, 30*, 484-505.

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

Moors, A., Spruyt, A., & De Houwer, J. (2010). In search of a measure that qualifies as implicit: Recommendations based on a decompositional view of automaticity. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 19-37). New York, NY: Guilford Press.

Morgan, C.D., & Murray, H.A. (1935). A method of investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry, 34,* 289–306.

Muchinsky, P.M. (1996). The correction for attenuation. *Educational & Psychological Measurement, 56*, 63-75.

Myford, C.M., & Wolfe, E.W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, *4*, 386-422.

National Alliance for Mental Illness (2008). Understanding schizophrenia and recovery: what you need to know about this medical illness. http://www.nami.org/Template.cfm?Section=By_Illness&template=/ContentManagement/ ContentDisplay.cfm&ContentID6=7729. Accessed 16 January 2014.

National Alliance for Mental Illness (2009). Major depression fact sheet. http://www.nami.org/Template.cfm?Section=Depression&Template=/ContentManagement/ContentDisplay.cfm&ContentID8=8956. Accessed 16 January 2014.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226-254.

Nosek, B.A., & Banaji, M.R. (2005). The Go/No-Go Association Task. *Social Cognition*, 1*9*, 625-666.

Nosek, B.A., Bar-Anan, Y., Sriram, N., & Greenwald, A.G. (2013). Understanding and Using the Brief Implicit Association Test: I. Recommended Scoring Procedures. Unpublished manuscript.

Nosek, B.A., & Greenwald, A.G. (2009). (Part of) the case for a pragmatic approach to validity: Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin, 135,* 373-376.

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

Nosek, B.A., Hawkins, C.B., & Frazier, R.S. (2012). Implicit Social Cognition. In S. Fiske & C.N. Macrae (Eds.) *Handbook of Social Cognition* (pp. 31-53). New York, NY: Sage.

Nosek, B.A., Hawkins, C.B., & Frazier, R.S. (2011). Implicit social cognition: from measures to mechanisms. *Trends in Cognitive Sciences, 15*, 152-159.

Ostafin, B.D., & Palfai, T.P. (2006). Compelled to consume: The Implicit Association Test and automatic alcohol motivation. *Psychology of Addictive Bahaviors, 20*, 322-327.

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

Payne , B. K., & Gawronski , B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski & B.K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1-15). New York, NY: Guilford Press.

Peeters M., Wiers, R.W., Monshouwer, K., van de Schoot, R., Janssen, T., & Vollebergh, W.A.M. (2012). Automatic processes in at-risk adolescents: the role of alcohol-approach tendencies and response inhibition in drinking behavior. *Addiction, 107*, 1939-1946.

Penke, L., Eichstaedt, J., & Asendorpf, J.B. (2006). Single Attribute Implicit Association Tests (SA- IAT) for the assessment of unipolar constructs: The case of sociosexuality. *Experimental Psychology, 53,* 283-291.

Peris, T.S., Teachman, B.A., & Nosek, B.A. (2008). Implicit and explicit stigma of mental illness: links to clinical care. *Journal of Nervous and Mental Disease, 196*, 752-760.

Piccinelli, M., Tessari, E., Bortolomasi, M., Piasere, O., Semenzin, M., Garzotto, N., & Tansella, M. (1997). Efficacy of the Alcohol Use Disorders Identification Test as a Screening Tool for Hazardous Alcohol Intake and Related Disorders in Primary Care: A Validity Study. *BMJ: British Medical Journal, 314*, 420-424.

Popa-Roch, M., & Delmas, F. (2010). Prejudice Implicit Association Test effects. The role of self-related heuristics. *Zeitschrift für Psychologie/Journal of Psychology, 218*, 44–50.

Posner, M.I., & Snyder, C.R.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Preacher, K. J., Rucker, D.D., & Hayes, A.F. (2007). Addressing moderated mediation hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research, 42,* 185–227.

Prochaska, J.O., DiClemente, C.C., & Norcross, J.C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 1102-1114.

Prochaska, J.O., Norcross, J.C., & DiClemente, C.C. (1994). *Changing for good: the revolutionary program that explains the six stages of change and teaches you how to free yourself from bad habits.* New York, NJ: W. Morrow.

Pryor, J.B., & Reeder, G.D. (2011). HIV-related stigma. In J.C. Hall, B.J. Hall & C.J. Cockerell (Eds.), *HIV/AIDS in the Post-HAART Era: manifestations, treatment, and Epidemiology* (pp. 790-806). Shelton, CT: PMPH-USA, Ltd.

Read, J., & Harre, N. (2001). The role of biological and genetic causal beliefs in the stigmatisation of 'mental patients.' *Journal of Mental Health UK, 10*, 223-235.

Read, J., Haslam, N., Sayce, L., & Davies, E. (2006). Prejudice and schizophrenia: a review of the 'mental illness is an illness like any other' approach. *Acta Psychiatrica Scandinavica, 114*, 303-318.

Richardson-Klavehn, A., & Bjork, R.A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475-543.

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

Robusto, E., Cristante, F., & Vianello, M. (2008). Assessing the impact of replication on implicit association test effect by means of the extended logistic model for the assessment of change. *Behavior Research Methods, 40*, 954-960.

Roefs, A., Huijding, J., Smulders, F.T., MacLeod, C.M., de Jong, P.J., Wiers, R.W., & Jansen, A. (2011). Implicit measures of association in psychopathology research. *Psychological bulletin*, *137*, 149-193.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the IAT: The Recoding-Free Implicit Association Test (IAT-RF). *Quarterly Journal of Experimental Psychology, 62,* 84-98.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment test*. Danish Institute for Educational Research, Copenhagen, 1960; reprinted, The University of Chicago Press, Chicago, 1980.

Rosenthal, R., & Rosnow, R.L. (1991). *Essentials of behavioral research: Methods and data analysis. (2nd ed.)*. New York: McGraw Hill.

Rüsch, N., Angermeyer, M.C., & Corrigan, P.W., (2005). Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. *European Psychiatry, 20*, 529-539.

Rüsch, N., Corrigan, P.W., Todd, A.R., & Bodenhausen, G.V. (2010). Implicit self-stigma in people with mental illness. *Journal of Nervous and Mental Disease, 198*, 150-153.

Rüsch, N., Corrigan, P.W., Todd, A.R., & Bodenhausen, G.V. (2011). Automatic stereotyping against people with schizophrenia, schizoaffective and affective disorders. *Psychiatry Research, 186*, 34-39.

Rüsch, N., Todd, A.R., Bodenhausen, G.V., & Corrigan, P.W. (2010a). Biogenetic models of psychopathology, implicit guilt, and mental illness stigma. *Psychiatry Research, 179*, 328-332

Rüsch, N., Todd, A.R., Bodenhausen, G.V., & Corrigan, P.W. (2010b). Do people with mental illness deserve what they get? Links between meritocratic worldviews and implicit versus explicit stigma. *European Archives of Psychiatry and Clinical Neuroscience, 260*, 617-625.

Rüsch, N., Todd, A.R., Bodenhausen, G.V., Olschewski, M., & Corrigan, P.W. (2010). Automatically activated shame reactions and perceived legitimacy of discrimination: A longitudinal study among people with mental illness. *Journal of Behaviour Therapy and Experimental Psychiatry, 41*, 60-63.

Rydell, R.J., & McConnell, A.R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91,* 995-1008.

Saunders, J.B., Aasland, O.G., Babor, T.F., & Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, *88*, 791-804.

Schacter, D.L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.

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

Schoenmakers, T.M., Wiers, R.W., Jones, B.T., Bruce, G., & Jansen, A.T.M. (2007). Attentional re-training decreases attentional bias in heavy drinkers without generalization. *Addiction*, *102*, 399-405.

Schoenmakers, T.M., De Bruin, M., Lux, I.F.M., Goertz, A.G., Van Kerkhof, D.H.T., & Wiers, R.W. (2010). Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. *Drug and alcohol dependence*, *109*, 30-36.

Schomerus, G., Schwahn, C., Holzinger, A., Corrigan, P.W., Grabe, H.J., Carta, M.G., & Angermeyer, M.C. (2012). Evolution of public attitudes about mental illness: A systematic review and meta-analysis. *Acta Psychiatrica Scandinavica, 125*, 440–452.

Schulz, K.F., Altman, D.G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*, 18.

Schulze, B. (2007). Stigma and mental health professionals: A review of the evidence on an intricate relationship. *International Review of Psychiatry, 19*, 137-155.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127-190.

Sijtsma, K. (2012). Psychological measurement between physics and statistics. *Theory & Psychology, 22*, 786-809.

Smith, E.R., & 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.

Spearman, C. (1904) The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72-101.

Spielberger, C.D. (1989). *State–Trait Anxiety Inventory: Forma* Y. Firenze: Organizzazioni Speciali.

Spiller, V., Zavan, V., & Guelfi, G.P. (2006). Assessing motivation for change in subjects with alcohol problems: the MAC2-A questionnaire. *Alcohol and alcoholism, 41*, 616–623.

Sriram, N., & Greenwald, A.G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*, 283-294.

Sriram, N., Greenwald, A.G., & Nosek, B.A. (2010). Correlational biases in mean response latency differences*. Statistical Methodology, 7*, 277–291.

Sritharan, R., & Gawronski, B. (2010). Changing implicit and explicit prejudice: Insights from the Associative-Propositional Evaluation model. *Social Psychology, 41*, 113-123.

Stacy, A., & Wiers, R.W. (2010). Implicit cognition and addiction: a tool for explaining paradoxical behavior. *Annual review of clinical psychology*, *6*, 551-575.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680.

Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Pychology* (pp. 1-49). New York: Wiley.

Stier, A., & Hinshaw, S.P. (2007). Explicit and implicit stigma against individuals with mental illness. *Australian Psychologist, 42*, 106-117.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and social psychology review, 8*, 220-247.

Teachman, B.A., Wilson, J.G., & Komarovskaya, I. (2006). Implicit and explicit stigma of mental illness in diagnosed and healthy samples. *Journal of Social and Clinical Psychology, 25*, 75-95.

Teige-Mocigemba, S., Klauer, K.C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: The Single Block IAT. *European Journal of Psychological Assessment, 24,* 237-245.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to the Implicit Association Test and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117- 139). New York: Guilford Press.

van Deursen, D.S., Salemink, E., Smit, F., Kramer, J., & Wiers, R.W. (2013). Web-based cognitive bias modification for problem drinkers: protocol of a randomised controlled trial with a 2x2x2 factorial design. *BMC Public Health, 13*, 674. http://www.biomedcentral.com/1471-2458/13/674

van Ravenzwaaij, D., van der Maas, H.L.J., & Wagenmakers, E.J. (2011). Does the name–race Implicit Association Test measure racial prejudice? *Experimental Psychology, 58*, 271-277.

Vianello, M., & Robusto, E. (2010). The many-facet Rasch model in the analysis of the Go/No-Go association task. *Behavior Research Methods, 42*, 944–956.

Wang, X., Huang, X., Jackson, T., & Chen, R. (2012). Components of implicit stigma against mental illness among Chinese students. *PLoS ONE, 7*(9), e46016. doi:10.1371/journal.pone.0046016

Weiner B. (1995). *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. Guilford

Weiss, M.G., & Ramakrishna, J. (2006). Stigma interventions and research for international health. *Lancet, 367*, 536 – 538.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95-116). New York, NY: Guilford Press.

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

Wigboldus, D.H.J., Holland, R.W., & van Knippenberg, A. (2005). *Single target implicit associations*. Unpublished manuscript, Raboud University, Nijmegen, The Netherlands.

Wiers, R.W., Eberl, C., Rinck, M., Becker, E.S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological science*, *22*, 490-497.

Wiers, R.W., Houben, K., Roefs, A., De Jong, P., Hofmann, W., & Stacy, A.W. (2010). Implicit Cognition in Health Psychology: Why Common Sense Goes Out of the Window. In B. Gawronski & K. Payne (Eds.), *Handbook of Implicit Social Cognition.* (pp. 463-488). NY: Guilford.

Wiers, R.W., Rinck, M., Dictus, M., & Van den Wildenberg, E. (2009). Relatively strong automatic appetitive action-tendencies in male carriers of the OPRM1 G-allele. *Genes, brain, and behavior*, *8*, 101-106.

Wiers, R.W., Rinck, M., Kordts, R., Houben, K., & Strack, F. (2010). Re-training automatic action-tendencies to approach alcohol in hazardous drinkers. *Addiction*, *105*, 279–287.

Wiers, R.W., & Stacy, A. W. (2006a). Implicit Cognition and Addiction. *Current Directions in Psychological Science*, *15*, 292–296.

Wiers, R.W., & Stacy, A.W. (Eds.) (2006b). *Handbook of implicit cognition and addiction*. Thousand Oaks, CA: SAGE Publishers.

Wiers, R.W., Teachman, B., & De Houwer, J. (2007). Implicit cognitive processes in psychopathology: an introduction. *Journal of behavior therapy and experimental psychiatry*, *38*, 95-104.

Wiers, R.W., Gladwin, T.E., Hofmann, W., Salemink, E., & Ridderinkhof, K.R. (2013). Cognitive Bias Modification and Cognitive Control Training in addiction and related psychopathology: mechanisms, clinical perspectives, and ways forward. *Clinical Psychological Science*, *1*, 192-212.

Wiers, R.W., van de Luitgaarden, J., van den Wildenberg, E., & Smulders, F.T.Y. (2005). Challenging implicit and explicit alcohol-related cognitions in young heavy drinkers. *Addiction, 100*, 806-819.

Williams, J.M.G., Mathews, A., & MacLeod, C.M. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, *120*, 3–24.

Wilson, T.D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117-142.

Wittenbrink, B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 17-58). New York: Guilford Press.

Wittenbrink, B., Judd, C.M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815-827.

World Health Organisation (2001). *Mental Health: new understanding, new hope. The World Health Report*. Geneva: WHO.

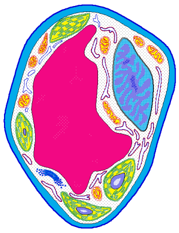
Wright B.D. (2003). Rack and Stack: Time 1 vs. Time 2 or Pre-Test vs. Post-Test. *Rasch Measurement Transactions, 17(*1), 905-906. http://www.rasch.org/rmt/contents.htm

APPENDIX A

Mental Illness Causal Beliefs IAT:

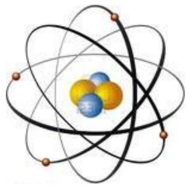
Psychological and Biologic Stimuli

Biologic images:

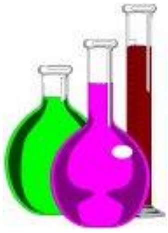


Bio1

Bio2

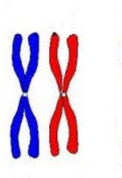


Bio3



Bio4

Bio4



Bio4

Psychological images:

Psycho1

Psycho2

Psycho3

Psycho4

Psycho5

Psycho6

APPENDIX B

Combined CBM Randomized Clinical Trial:

Ethical Procedure Material

UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITY OF AMSTERDAM

PRESENTAZIONE DELLO STUDIO

Tecniche di intervento nei processi impliciti impulsivi nella dipendenza da alcol in pazienti ambulatoriali – Progetto TOP-training

Gentile paziente, con questo studio si desidera verificare l'efficacia di nuovi interventi terapeutici finalizzati al trattamento della dipendenza da alcol. Questa tipologia di trattamenti funziona attraverso il coinvolgimento e la modificazione comportamentale di alcuni meccanismi automatici alla base dell'abuso e della dipendenza dall'alcol, i quali aumentano il rischio concreto di ricadute.

Molte persone descrivano il loro consumo di alcol come una scelta molto spesso non consapevole: sembra che il "cedere" alla tentazione e il ricadere nel consumo di alcolici accada da sé, prima, o addirittura senza, che la persona se ne renda effettivamente conto. La ricerca scientifica in questo campo ha infatti evidenziato come alcuni comportamenti e pensieri tipici nelle dipendenze, spontanei e a volte inconsapevoli, giochino un ruolo rilevante nella dipendenza da alcol. Queste modalità di comportamento e di pensiero, "istintive" e "involontarie", sono così veloci e automatiche nel prendere il sopravvento che l'individuo non ne è consapevole e allo stesso tempo incapace di prestarci attenzione o esercitare un controllo attivo su di esse.

Obiettivo del progetto TOP-training è quello di verificare l'efficacia di alcune nuove procedure di trattamento di questi meccanismi involontari e inconsci, aiutando il paziente a guadagnare ed esercitare un maggiore controllo su di essi.

La informiamo che questi nuovi trattamenti non andranno a sostituire la sua terapia standard, ma saranno ad essa affiancati ed eseguiti in parallelo. Tale scelta deriva da risultati di studi precedenti in cui l'utilizzo combinato di diverse forme di trattamento sembra essere promettente in termini di efficacia del percorso terapeutico e di riduzione del rischio di ricadute.

Lo studio prevede la partecipazione a 15 sedute di training della durata di 45 minuti circa, due volte alla settimana durante le sue visite al SerD. Le sedute consistono nell'esecuzione di due esercizi al computer, in cui Le viene chiesto di classificare gli oggetti visualizzati sullo schermo utilizzando alcuni tasti della tastiera. Questi "esercizi"hanno come obiettivo quello di facilitare l'apprendimento e la pratica delle abilità di controllo nei confronti di stimoli legati all'alcol. All'inizio e al termine delle sedute di training, e durante un controllo a distanza di 3 *mesi, le chiederemo di compilare alcuni questionari e strumenti di valutazione, al fine di verificare i risultati raggiunti.*

Tale progetto di ricerca TOP-training è frutto di una collaborazione tra il Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata (FISPPA) dell'Università degli studi di Padova, il Servizio sanitario per le Dipendenze (SerD) dell'ULSS10 di San Donà di Piave e il gruppo di ricerca ADAPTlab dell'Università di Amsterdam, il quale è specializzato nello studio e sviluppo di trattamenti ad hoc per i disturbi da dipendenza da sostanze.

UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITY OF AMSTERDAM

INFORMAZIONI NORMATIVA DI PARTECIPAZIONE ALLO STUDIO

Si informa che l'adesione allo studio è su base volontaria, e che i suoi dati rimarranno strettamente confidenziali e saranno protetti secondo il Codice in materia di protezione dei dati personali (Dlgs. n. 196/2003;Direttiva Europea 95/46/EC) e dal segreto professionale (Codice Deontologico della Professione di Psicologo). Solamente il personale responsabile della ricerca potrà avere accesso ai dati ai fini della ricerca e divulgazione scientifica, e in nessun modo verranno rese note informazioni sui casi singoli o che possano rendere identificabili i partecipanti alla ricerca.

Si informa inoltre il paziente che è un suo diritto interrompere la partecipazione allo studio in qualsiasi momento, senza fornire alcuna motivazione, senza alcuna conseguenza e penalizzazione ed ottenendo il non utilizzo dei suoi dati.

UNIVERSITÀ DEGLI STUDI DI PADOVA

ULSS10 VENETO ORIENTALE

UNIVERSITY OF AMSTERDAM

MODULO DI CONSENSO INFORMATO 1

Al fine di procedere con lo studio e garantire il suo diritto alla privacy e libertà di partecipazione, La preghiamo di leggere le seguenti voci:

- ☐ *Dichiaro di aver acconsentito volontariamente alla partecipazione allo studio.*
- ☐ *Sono stato informato/a, prima di partecipare al suddetto studio, del mio diritto di interrompere la mia partecipazione allo studio in qualsiasi momento, senza fornire alcuna motivazione, senza alcuna penalizzazione e riottenendo tutti i miei dati.*
- ☐ *Sono stato informato dello scopo del suddetto studio e del fatto che tutti i dati che mi riguardano rimarranno riservati, protetti dal segreto professionale, e accessibili solo ai responsabili del progetto, secondo quanto stabilito dalla legge (Dlgs. n. 196/2003).*
- ☐ *Sono stato informato che solo le persone che conducono la ricerca potranno avere accesso ai miei dati limitatamente ai fini della loro elaborazione e alla pubblicazione anonima dei risultati a fine scientifico.*
- ☐ *Autorizzo i responsabili del presente studio all'utilizzo dei miei dati.*

Data, ___________________________

Nome e cognome del partecipante: __________________________________

Firma

____

UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITY OF AMSTERDAM

MODULO DI CONSENSO INFORMATO 2

La ringraziamo per la partecipazione allo studio.
Al fine di completare la procedura sperimentale, La preghiamo di leggere ed esprimere nuovamente il suo consenso o meno alle seguenti voci inerenti la partecipazione allo studio e la tutela dei suoi dati personali:

- □ *Dichiaro di aver acconsentito volontariamente alla partecipazione allo studio.*
- □ *Sono stato informato/a, prima di partecipare al suddetto studio, del mio diritto di interrompere la mia partecipazione allo studio in qualsiasi momento, senza fornire alcuna motivazione, senza alcuna penalizzazione e riottenendo tutti i miei dati.*
- □ *Sono stato informato dello scopo del suddetto studio e del fatto che tutti i dati che mi riguardano rimarranno riservati, protetti dal segreto professionale, e accessibili solo ai responsabili del progetto, secondo quanto stabilito dalla legge (Dlgs. n. 196/2003).*
- □ *Sono stato informato che solo le persone che conducono la ricerca potranno avere accesso ai miei dati limitatamente ai fini della loro elaborazione e alla pubblicazione anonima dei risultati a fine scientifico.*
- □ *Autorizzo i responsabili del presente studio all'utilizzo dei miei dati.*

Data, ___________________________

Nome e cognome del partecipante: __________________________________

Firma

____