Preprint: The Version of Record of this manuscript has been published and is available in *Philosophical Psychology* <07.04.2022> <http://www.tandfonline.com> <https://www.tandfonline.com/doi/full/10.1080/09515089.2022.2061342>

René Baston
rene.baston@hhu.de
ORCID iD: 0000-0001-9904-3729

**What underlies Death/Suicide Implicit Association Test measures and how it contributes to suicidal action**

**Abstract**

Recently, psychologists have developed indirect measurement procedures to predict suicidal behavior. A prominent example is the Death/Suicide Implicit Association Test (DS-IAT). In this paper, I argue that there is something special about the DS-IAT which distinguishes it from different IAT measures. I argue that the DS-IAT does not measure weak or strong associations between the implicit self-concept and the abstract concept of death. In contrast, assuming a goal-system approach, I suggest that sorting death-related to self-related words takes effort because death-related words trigger avoidance-impulses, which suicide ideation weakens. The DS-IAT taps into weakened automatic responses from the self-preservation system. Additionally, the suggested cognitive structure, illuminated with the selfish-goal theory, explains predictable suicidal behavior.

**1 Introduction**

While suicide is one of the leading causes of death worldwide (Nock et al., 2012), it is still unclear why individuals kill themselves. Only a fraction of those sincerely considering suicide or desiring death actually die by suicide – what exactly accounts for the difference between wanting and doing in this case is still unclear.

Traditionally, psychologists relied on self-report measures to assess an agent’s risk for suicide. This approach comes with its limitations. For example, a study shows that 78% of patients who die by suicide explicitly deny suicidal thoughts during their last verbal communications before they kill themselves (Busch et al., 2003). Self-report measures are problematic for two reasons. First, agents may intentionally report false statements. It is possible that suicidal agents who decided to kill themselves do not want to unfold their plans. Second, psychological data raises doubts regarding the agents’ introspective powers and their self-knowledge (Bargh, 2017; Bargh & Morsella, 2008; J. Y. Huang & Bargh, 2014; Nisbett & Wilson, 1977; Wilson, 2009). Possibly, unconscious mechanisms generate suicidal behavior. As psychological theories of unconscious processing show, this does not mean that agents who die by suicide perceive their behavior as involuntary. In contrast, it is volitional behavior, while agents are unaware of the major causes of these volitions. These explanatory reasons may be inaccessible for suicidal agents but influential for suicide attempts.

More recently, psychologists investigated implicit cognition to understand suicidal behavior. In contrast to self-report measures, indirect measurements operationalize an agent’s behavioral patterns while undergoing a task to assess a psychological attribute. Here, the agent does not need to communicate the psychological attribute in question, nor does the agent need to be aware of the attribute. A meta-study by Moreno et al. (2022) on implicit cognition tests for the assessment of suicide risks summarizes that indirect measures are generally valid for the assessment of past and future suicidal thoughts and behaviors.

For example, the Self-Injury Implicit Association Test (SI-IAT; Nock & Banaji, 2007), a variant of the Implicit Association Test (Greenwald et al., 1998), was developed to measure a participant’s implicit associations between self-injury and oneself. The authors took such associations as an indicator for self-injurious thoughts, which were robust predictors of suicidal outcomes. There is, furthermore, a variant of the Affective Misattribution Procedure (AMP; Payne et al., 2005) for unconscious suicidal cognitions, called the Suicide Affect Misattribution Procedure (Tucker et al., 2018). Additionally, there is the Death/Suicide Implicit Association Test (DS-IAT; Nock et al., 2010), which I want to focus on.

The DS-IAT is a variant of the IAT and indirectly measures an agent’s suicide risk. Studies show mixed results regarding the DS-IAT’s predictive powers. On the one hand, Nock and colleagues (2010) report that 157 individuals who were seeking treatment at a psychiatric emergency department underwent the DS-IAT measurement. Subjects who scored high on the DS-IAT were associated with an approximately 6-fold increase in the odds of making a suicide attempt in the following 6 months. In another study (Barnes et al., 2017) with 176 psychiatrically hospitalized veterans, the DS-IAT significantly predicted suicide attempts during a 6-month follow-up above and beyond other known risk factors for suicidal behavior. On the other hand, some studies raise doubts regarding the DS-IAT’s predictive powers (for example, see Harrison et al., 2018; Rath et al., 2021). A meta-study on the DS-IAT (Sohn et al., 2021) draws a positive conclusion regarding the usefulness of the DS-IAT for clinical psychology.

Nock and colleagues suggest the IAT taps into a cognitive structure which shows that an agent is “unwilling or unable to report their intentions” to end their own lives: “More specifically, an implicit association with death/suicide may represent one of the final steps in the pathway to suicide that is activated when a person is deciding how to respond to extreme distress.” (Nock et al., 2010, p. 5) The authors state that “a person’s implicit cognition may guide which behavior he or she chooses to cope with extreme distress.” (Nock et al., 2010, p. 5) Accordingly, Nock and colleagues claim that the measured construct is causally responsible for the predicted suicide attempts, which is an assumption that I will share. Moreover, in this text, I will not follow the possible interpretation that individuals who score high on the DS-IAT are *unwilling* to report their suicidal intentions. While this is a reasonable assumption, in this text, I want to develop the other explanatory option, namely that agents are *unable* to self-report their suicidality.

But what does it mean to have intentions while being unable to report them? According to most philosophers, having an intention goes hand in hand with being able to report them (Anscombe, 1957; Mele & Moser, 1994). Furthermore, in which sense does an *association* between the implicit self-concept and death lead to a decision? In which sense is intentional action guided by implicit cognition?

The goal of this manuscript is to develop an alternative account of the underlying psychological structure which the DS-IAT measures. In Section 3, I will show that there is something special about the DS-IAT measurement procedure which distinguishes it from different IAT measurements. I argue that the DS-IAT does not measure weak or strong associations between a self-concept and the abstract concept ‘death’. In contrast, the DS-IAT taps into weakened automatic natural prepotent responses from the self-preservation system. In Section 4, I suggest how to understand a suicide attempt as intentional action, while agents are unaware of the psychological mechanisms which partly guide their intentional action (without undermining its intentionality).

I will rely on a goal-system approach from psychology (Kruglanski et al., 2002). Such approaches assume that goal-directed behavior grounds on a network of goal-representations which have associations with different goals, sub-goals or means. In particular, I will refer to a special version of a goal-system approach, namely on the selfish-goal theory (J. Y. Huang & Bargh, 2014). The selfish-goal theory claims that human behavior depends on conscious and unconscious goals that compete for cognitive resources. In the process of competition, activated goals ‘try’ to be as influential as possible by biasing information processing. The theory explains why agents’ behavior is sometimes inconsistent and hard to understand from a folk-psychological perspective.

Beforehand, I want to narrow down the relevant type of suicidal behavior. Suicidal behavior “is not a homogeneous phenomenon but a spectrum of self-destructive behaviors that may differ from each other in terms of lethality, planning, and intent.” (van Heeringen, 2018, p. 2) In this text, I am interested in cases which are relevant for clinical psychology: suicide as an abnormal reaction to a normal situation (van Heeringen, 2018, p. 23). Van Heeringen (2018, p. xi), for instance, starts his book with the disturbing case of Valerie who took her bicycle, went to a highway bridge and jumped down because her uncle looked angry at her.[[1]](#footnote-1)

**2 The Implicit Association Test and the Death/Suicide Implicit Association Test**

In this Section, I want to introduce to the DS-IAT. Because the DS-IAT is an adaption of the IAT, I will introduce the classic IAT first. Afterwards, I will highlight some of the DS-IAT’s characteristic properties.

The IAT is a computer-administered test which operationalizes the participant’s reaction times to measure (unconscious) psychological attributes. When a subject undergoes an IAT, the subject has to sort semantic or pictorial stimuli to the left or the right side of the screen. Subjects sort a stimulus to the left by pressing the key ‘e’ and sort a stimulus to the right by pressing ‘i’. Usually, the left and the right side of the screen are labeled with a category and a value-based concept. For example, the category could be ‘insects’ on the left, and ‘flowers’ on the right. In addition, there is a value concept on the left, say ‘good’, and on the right there is ‘bad’. In the current example, there are ‘insects/good’ on the left and ‘flowers/bad’ on the right. During a test, in the middle of the screen, various words or pictures show up, one after another. The participant must sort the appearing stimuli as fast as she can to the left or the right side. If the word ‘delicious’ appears, the participant has to order it to the left because it is labeled with the concept ‘good’. However, there is the concept ‘insect’ as well, and many people do not associate insects with ‘delicious’. Because the stimulus is incongruent with the value-based concept, the participant will take slightly longer to sort the stimulus. In contrast, if the stimulus in congruent with the labels, the participant can quickly solve the task. After a few trails, the labels (values and categories) on the sides change (for example to ‘insects/bad’ and ‘flowers/good’). Psychologists assume that the IAT measures the relative strength of a participant’s association between a category and a value-based concept.

It is not clear that the IAT measures associations, even though IAT stands for Implicit *Association* Test. As Dacey (2016) argued, when psychologist talk about associations, they are often using an abstract filler term which stands in for a causal relation between representational states in a system. These ‘associations’ could be implemented by different mechanisms. For instance, the psychologist De Houwer (2014) suggests a stronger ‘association’ between *insects* and *good* reflects a propositionally structured automatic evaluation of insects. In philosophy, Gendler (2008) thinks the IAT measures associations, Mandelbaum (2016) argues the IAT measures belief-like propositionally structured mental states, Levy (2015) argues the measured mental states are propositional but not belief-like, while Baston (2018) suggests that implicit attitudes are neither propositional nor conceptual representations.

Whatever the IAT actually measures, *theorists assume that the measured mental structure is responsible for response latencies during the measurement and (partly) for the behavior that is predictable* (De Houwer & Moors, 2010; Gawronski et al., 2020, p. 17). The IAT is often used to measure participants’ implicit prejudices and implicit stereotypes against social groups (Banaji, 2001; Greenwald et al., 2002). For example, if a participant who undergoes an IAT measurement takes longer to sort pleasant words to ‘African-American’, the subject has the (unconscious) automatic evaluation in mind that people of color are bad. This explains why the subject needs longer to solve the sorting task, *and* it can explain unintentional discriminatory behavior towards people of color (Greenwald et al., 2009; Rooth, 2010).

A participant undergoing a DS-IAT has to rapidly categorize words that represent the construct of ‘death’ (die, dead, deceased, lifeless, suicide) and ‘life’ (alive, survive, live, thrive, breathing) to attributes of ‘me’ (I, myself, my, mine, self) and ‘not me’ (they, them, their, theirs, other) (Nock et al., 2010). Like in a classic IAT, the DS-IAT is a computer-based test, and the left and the right side indicate an attribute (life or death) and an object (me or not me). Scoring high on the DS-IAT represents a stronger association between *death* and *self*. Accordingly, a high score indicates faster responding on the ‘death/me’ blocks.

Theorists assume that the DS-IAT assesses an agent’s implicit self. The implicit self is an associative structure in memory that operates automatically (Devos et al., 2012). For a subject, the implicit self’s influences can be inaccessible by introspection. Usually, indirect measurements of the implicit-self access the strengths of the self-concept and a group association. When psychologists use the IAT in this way, a strong group association with the self indicates an automatic identification with the social group. Studies suggest that once agents automatically identify with a social group, agents associate related group-attributes with the implicit self. For instance, studies (Nosek et al., 2002) reveal automatic associations between the concept ‘arts’ for women and the concept ‘math’ for men. So, psychologists measure a subject’s automatic self-identification, which relates to the subject’s social identity. This, however, makes the DS-IAT even more puzzling. Clearly, it is not part of a person’s *social identity* to identify with death. What does it mean, then, to implicitly identify oneself with death and how does it translate to behavior?

In sum, there is the challenge to suggest a cognitive structure that the DS-IAT’s measurements reflect and which is explanatory for the operationalized participants’ behavior while undergoing a measurement procedure as well as the predicted suicidal behavior.

**3 About the implicit structure which the DS-IAT measures**

I want to suggest that the DS-IAT does not measure an association between *me* and *death*, because an association between *me* and *death* cannot explain suicidal behavior. I do not want to criticize the general idea of measuring psychological attributes of the implicit self. However, given how an IAT measurement works, there is potential that distinct psychological processes interfere with the agent’s sorting task.

In an IAT measurement, a participant has to sort stimuli to a label that is incongruent with the stimuli. From a phenomenological perspective, it feels harder to sort a stimulus to one side than to the other. To sort an incongruent stimulus to an unfitting category, an agent needs to overcome her first intuition. It takes effort to solve the IAT, and the agent has to engage in executive functions. Response inhibition is a cognitive process which enables an agent to inhibit their impulses, namely habitual or natural responses to a stimulus. Thereby, an agent can select a more appropriate behavior which is consistent with the primary behavioral goal.

To overcome a prepotent response, an agent engages in the executive functions of response inhibition and overriding (Cohen, 2017). In a Stroop test, agents have the task of naming the stimulus’ color, while the stimulus can be incongruent. If the word ‘GREEN’ is displayed in red, it takes effort in self-control to name the displayed color. Here, color naming is a controlled process, while the word reading is automatic. There is a conflict between the voluntary goal to name the color and the automatically triggered action-plan to read the word. The agent inhibits the prepotent response to read the word, and this enables the agent to override the automatic response.

I suggest two ways to explain what kind of prepotent responses are at play when an agent undergoes an IAT. Firstly, there are *habitual* prepotent responses. These responses capture what psychologists traditionally conceptualize as associations which are measured with the IAT. Associative learning through the mechanisms of operant conditioning, classic conditioning, and observational learning is habitual. All types of associative learning can gain new behavioral responses. The association’s gained content depends on an agent’s personal learning history and can differ greatly from individual to individual.

Secondly, there are *natural* prepotent responses. The source of these prepotent responses are different psychological processes interfering with the participant’s goal to sort incongruent stimuli while undergoing an IAT. Besides influences of weak or strong associations, certain stimuli automatically active evolutionary developed cognitive modules which influence a participant’s reactions.

According to Öhman and Mineka (2001), humans have a natural tendency to fear certain stimuli. Because of the evolutionary history of the human species, an encapsulated fear module developed, which originates in a dedicated neural circuitry, centered on the brain’s amygdala. This response system made it possible for our ancestors to quickly attack, freeze, or escape in dangerous situations. Because of evolutionary contingencies, some stimuli, including snakes and threatening faces, are more powerful stimuli to induce fear. In addition, the fear-responses are hard to change with known procedures, like extinction or counter-conditioning, and the fear response for such stimuli is quickly learned. While fear responses towards snakes or spiders depend on learning, the gained fear-response is functionally and neurologically different.

I claim that if an IAT triggers *natural* prepotent responses, then it taps not only into weak associations but into naturally evolved cognitive processes. If an agent who is undergoing an IAT has to sort aversive stimuli, such as rotten meat, feces, or a wolf showing teeth, to the self-referential attribute *me*, the agent will only succeed by putting effort into the task. The agent needs executing functioning to override the first natural reaction, which is to sort disgusting or threatening stimuli away from oneself. There is an incongruent stimulus, but the learned association between the implicit self-conception and the stimulus does not explain the inhibitory effect. Surely, a theorist can rephrase the response latency by saying that the agent has a weak association between himself and, say, a hissing cat. Even though this is a correct description, it misses the central point. The stimulus’ incongruity stems from a natural prepotent response and not from incongruous associations with the self-concept in memory.

Recent data speaks in favor of this interpretation. A study (Rivers & Hahn, 2019) suggests that the IAT taps not only in stored mental associations but also in cognitive processes of self-control. A subject’s own introspective access to implicit attitudes’ contents (Hahn et al., 2014) is tunneled through a combination of activated associations and immediate control of behavioral impulses, which result from those activated associations. This means that subjects often successfully predict their own IAT results because they experience the association’s value and, in addition, their ability to self-regulate. This, in turn, means that the IAT actually measures more than mental associations because the measurement result accounts for processes of self-regulation as well.

The self-accessed ability to overcome a biasing influence depends on the assessed psychological attribute. A study (Rivers & Hahn, 2019) suggests that the ability to overcome a bias does not depend on general executive functioning abilities, but on the specific psychological attributes. The self-control processes which contribute to an IAT measurement and to an agent’s access to her implicit attitude change with the measured psychological attribute, like *flowers*, *cutting* or *death*. Different psychological attributes trigger different biasing influences. The brain is a parallel processing system and perceiving a stimulus can trigger various processes simultaneously. For instance, studies suggest that IAT scores reflect not only the target categories’ valence, but can also be influenced by stimulus effects (Bluemke & Friese, 2006; Gast & Rothermund, 2010). Thus, an IAT measurement result can reflect an automatic fear response towards a single presented stimulus, in contrast to reflecting the negative valence of the related stimulus’ category.

The Self-Injury Implicit Association Test (SI-IAT; Nock & Banaji, 2007) is the DS-IAT’s ancestor and illuminates what the DS-IAT measures. An agent undergoing an SI-IAT sees self-injury related stimuli (pictures of cut skin) or neutral stimuli (pictures of non-injured skin) and has to order them to the corresponding left or right side. The construct label is *cutting* and *no* *cutting*, and the attributes are *me* and *not me*. Predictive powers of the SI-IAT were tested on 89 adolescents, including suicide attempters, non-suicidal subjects, and suicide ideators.[[2]](#footnote-2) The study shows that non-suicidal adolescents have a negative association between self-injury and themselves, suicide ideators show weak positive associations, and suicide attempters show a strong positive association between *cutting* and *me*. The SI-IAT incrementally improved the prediction of future suicide attempts beyond the use of known risk factors.[[3]](#footnote-3)

What makes it harder to sort self-injury related stimuli towards the self-referential attribute? Arguably, there is a weak association between the stimuli and the self-concept, but this misses the crucial point: Affective stimuli trigger behavioral dispositions of avoidance or approach behavior. Empirical studies indicate that individuals show better performances in tasks in which the stimuli-response pairing is congruent with positive-approach and negative-avoid assignments. Then again, if the stimuli-response pairing at a task is incongruent, namely negative-approach and positive-avoid pairings, then individuals’ performance is lower. For instance, participants faster pull a lever towards themselves in response to positive stimuli, while they are faster to push a lever away in response to negative stimuli (Duckworth et al., 2002).

However, the avoidance-approach tendencies do not depend on an intrinsic valance-connotation of muscle activations, like arm flexion (approach) or arm stretching (avoidance). In contrast, the behavioral impulses serve to *decrease or increase the distance* between the person and the evaluated stimulus. In a study by Lavender and Hommel (2007), the experimental setting changed the lever movements’ reference point from the physical self to the computer screen that depicted emotional pictures. Participants were faster to classify negative pictures with arm-flexing lever movements away from the screen and positive pictures with arm-extending lever movements towards the screen.

A study by Markman and Brendl (2005) supports the assumption that the SI-IAT measures primarily the strength of avoidance (and approach) impulses in response to affective stimuli. In that study, participants viewed a computer screen that showed their name in the middle (a self-referential attribute), while positive and negative *words* were presented either in front of or behind the participant’s name. The participant’s goal was to move the words with pressing or pulling a lever towards or away from their name, depending on the word’s valance. No matter whether they had to pull or push the lever, they were faster to move positive words towards their name than away, while the study reveals the opposite pattern for responses to negative words. First, the study reveals no intrinsic valence effect of arm-stretching and arm-extending. Second, the study shows that self-referential words displayed on a monitor function as a reference point for avoidance and approach impulses, which are triggered by affective words.

I claim that a high score at the SI-IAT indicates a weakened self-preservation system. The self-preservation system is an umbrella term for different developed systems which ensure the organism's survival. It incorporates fear processing, the disease avoidance system, and more systems to regulate hunger, thirst and the body temperature. In addition, it includes generation, anticipation and processing of pain (Hansen & Streltzer, 2005). I suppose it is a natural prepotent response to avoid stimuli which show physical injuries. Against this background, usually, there is not only a weak association between *cutting* and *me*, but an avoidance-impulse, which indicates an automatic goal to keep flesh wounds away from oneself. This goal, however, competes with the deliberate goal to sort aversive stimuli towards *me*. For subjects who score high on the SI-IAT, this competition between cognitive processes is basically absent.

The DS-IAT takes this approach one step further. It does not depend on aversive pictorial stimuli, but on the abstract concepts of death and related words. Even though words, such as *die* and *dead*, do not appeal to a single evolved survival-instinct system, they appeal to a cascade of self-preservation sub-systems. There is no single system for the so-called survival instinct, but a variety of systems which interact and influence behavior to guarantee self-preservation (Kirkpatrick & Navarrete, 2006). A single evolutionary system with the instruction ‘avoid death’ is of no guidance in determining how to achieve this goal. In contrast, different evolved systems lead to avoidance behavior for kinds of dangers (and to approach behavior for goods which improve self-preservation). The fear of death (see chapter 2 in Joiner, 2007), which is found across all types of populations (Neimeyer et al., 2004), expresses these system’s interactions. Cognitive aspects of the fear of death include the awareness of death indicators, as well as beliefs, attitudes, and thoughts about death and dying (Lehto & Stein, 2009).

The mechanisms which are linked with fear of death include conscious and unconscious cognitive processes (Arndt et al., 2001; Bradley et al., 2012; Greenberg et al., 1994; Z. Huang & Wyer, 2015). Here, I consider cognitive processes as unconscious if one is not aware of crucial influences of a processed stimulus (Bargh & Morsella, 2008). Often, psychologists label this aspect of (un)consciousness *impact awareness* (Gawronski et al., 2006). For this understanding of unconsciousness, it is unimportant whether agents have awareness of a stimulus or a mental state’s content, which is *content awareness*. Data indicates that agents have content awareness of implicit mental states (Berger, 2018; Gawronski et al., 2006, 2020; Krickel, 2018), like implicit prejudices, while it is questionable that they have corresponding impact awareness (Bargh, 2017; Wilson, 2009).

*Death* is a very abstract word, which raises the question of how the self-preservation system could respond to it. The so-called self-preservation system comprises subsystems, like the fear processing subsystem. Being told in a believable way that we are going to face a serious threat triggers fear, which shows how language processing interacts with fear processing. Psychologists suggest different ways to spell out the interconnection from abstract words to emotions, like fear. For instance, abstract words and concepts may ground in our emotional experiences (Vigliocco et al., 2018): While individuals primarily learn words referring to concrete objects or actions through associating sensory-motor experiences with a word, the learning of abstract words depends on associations with emotional states. Tomasello’s (2008) social-cognitive theory of lexical development suggests that grasping abstract words through emotional underpinnings depends on an understanding of others. When individuals understand others, they associate embodied emotional reactions and experiences from others who were responding to events related to, say, *death*.

I suggest that suicidal individuals can more easily sort a death related stimulus to a self-referential attribute because potentially conflicting natural responses from the self-preservation system are weaker. Lester (1967) studied the relationship between fear of death and suicide potential. He expected that suicidal adolescents have less fear of death than non-suicidal adolescents, and he confirmed this hypothesis in two studies. According to D’Attilio and Campbell (1990), the data is, however, inconsistent regarding this aspect. D’Attilio and Campbell (1990) themselves found that adolescents who consider suicide (suicide ideators) show more fear of death than non-suicidal adolescents. The study suggests that the inconsistencies could be resolved by claiming that “high death anxiety among potentially suicidal adolescents may serve as an adaptive cognitive mechanism in preventing actual suicide attempts” (p. 977). In contrast, suicide attempters show less of these adaptive cognitive mechanisms which prevent them from suicidal behavior. There are, so to speak, less automatic cognitive processes which lead to avoidance and distancing impulses. I suggest that the DS-IAT taps into these weakened natural prepotent responses from the self-preservation system. But how is it possible to weaken natural prepotent responses from the self-preservation system?

I suggest that *suicide ideation* weakens natural prepotent responses from the self-preservation system. An agent engages in suicide ideation if the agent repeatedly considers suicide as an option. Suicide ideation includes cognitive and behavioral components (van Heeringen, 2018). Cognitive components include imagining killing oneself or thinking about preparations for suicide attempts, while the behavioral aspects include actual planning or the role play of suicide attempts. Thereby, suicide ideation works like a cognitive behavioral therapy against automatic negative reactions. If, for instance, an individual undergoes a cognitive behavior therapy to solve a fear of snakes, the therapist might start with asking the patient to read about snakes. Afterwards, the patient sees pictures of snakes. At some point, the patient may visit a zoo to see a snake. Finally, the patient has to task to hold a snake in her hands. This process gradually increases the level of exposure to a stimulus, and treatment progress is to feel less anxious about the fear stimulus.

In fact, fatal suicide is often the result of a suicidal process (Joiner, 2007; Retterstøl, 1993; van Heeringen, 2018). Such a process includes observable and unobservable events. The unobservable events are the cognitive components of suicide ideation, like imaging to kill oneself. The observable events include suicidal messages, self-harming behaviors and suicide attempts. According to the interpersonal theory model of suicidal behavior (IPT; Joiner, 2007; Van Orden et al., 2010), humans are not born with the capacity to kill themselves. In contrast, suicide ideation is a precondition to gain control over one’s own behavioral reactions, which naturally prevent humans from suicide.

The goal-system approach (Kruglanski et al., 2002) helps to understand how suicide ideation may establish the capacity for suicidal behavior. In contrast to classic assumptions, a person’s motivations are not static but dynamic, and a goal-system’s motivational constructs can explain how motivation changes dynamically based on an agent’s situation. A motivational construct is a goal, a mental representation of a desired end-state, with associated means, a set of acts to reach an end-state (action-plan).[[4]](#footnote-4) In a goal-system, goals are associated with means or different goals, and means are associated with related goals or distinct means. Accordingly, there is a network of goal-means associations. The more often a subject pairs a certain goal with a means, the stronger gets the association.

With a goal-system approach in mind, first, the cognitive aspects of suicide ideation change an agent’s motivational constructs. An individual who engages in suicide ideation makes the suicidal action-plan highly accessible by repetitively thinking about solving or escaping problems with suicide. Because of the repetition, the association between certain goals and the suicidal action-plan becomes stronger. As a result, an aversive situation may activate a goal which, in turn, activates the suicidal action-plan. Once the association is strong, a goal can trigger the means and *vice versa*.

The more goals an agent can reach by a single means, the higher the agent’s tendency to execute the action-plan (Kruglanski et al., 2002). Agents tend to choose the action-plan which promotes the greatest expectancy of goal attainment. If a single action-plan can attain various goals, this action-plan is valuable. In these cases, even though the expectancy is low, the value is high because different goals are in reach. Accordingly, if agents repetitively consider the suicidal action-plan as a means for different goals, this action-plan associates stronger with a range of distinct goals. If this is the case, the suicidal action-plan gains high subjective utility because different goal values are achievable at once.

The goal-system approach claims that goals, once activated, compete for cognitive resources, like goal-relevant attention (Kruglanski et al., 2002). A series of studies support this interpretation through different experimental setups with a priming procedure. If primes trigger incongruent goals, they affect resources allocated to the agent’s focal goal, influencing the agent’s commitment and progress in attaining it (Shah & Kruglanski, 2002). Because of suicide ideation, goals associated with the suicidal-action plan are easily triggered by suicide-related words, such as *death* or *funeral*. The activated goal-structure pulls resources away from the self-preservation system as they are in a competition for cognitive resources (I will elaborate more on this point in the next section). While sorting death-related to self-related words is usually inhibited due to automatic avoidance-responses, these responses reserve less influence if primes activate competitive goal-structures.[[5]](#footnote-5)

Second, the behavioral aspects of suicide ideation actively decrease natural prepotent responses. By executing behavioral patterns which expose the agent to the suicidal action-plan, the agent learns to control automatic psychological reactions (see chapter 2 in Joiner, 2007). For instance, buying a rope that could be used to hang oneself is a behavioral aspect of suicide ideation. Even though the agent does not have the intention to kill herself, this behavioral pattern increases the agent’s power over her psychological reactions, which are associated with suicide. Similarly, standing on a chair in front of a gallows noose without the intention to kill oneself desensitizes the fear of death. While the action-plan to kill oneself still causes automatic fear-reactions, the behavioral aspects of suicide ideation put an agent in the position to generate effective coping strategies.

In sum, it takes effort to solve a DS-IAT, because agents have to inhibit and override automatic responses. To sort death-related to self-related words, agents must inhibit and override automatic avoidance responses from the self-preservation system, which is a cascade of automatically working sub-systems to protect the body from harm. A goal-system approach explains how suicide ideation weakens responses from the self-preservation system. Goal-means associations compete for cognitive resources, such as attention, while a goal’s influence is determined by its accessibility, expectancy of goal attainment, and by how influential other goals are. Suicide ideation weakens responses from the self-preservation system in two ways. First, because of intense suicide ideation cognitive processes that are causally responsible for avoidance-responses allocate fewer resources. Second, behavioral elements of suicide ideation deliver coping-strategies for overcoming impulses of inhibition.

**4 About unreportable intentions and action-guidance through implicit cognition**

So far, I argued that the DS-IAT measures weakened responses from the self-preservation system. Responsible for this response imbalance is the process of suicide ideation which, unbeknownst to an agent, establishes strong associations between different goals and a suicidal action-plan. In this Section, I want to suggest an explanation for how this implicit cognitive structure guides action.

First, according to Nock (2010, p. 5), a DS-IAT measurement may reflect unreportable intentions, but how can an intention be unreportable? Doing something intentional presupposes that an agent knows what she is doing, and if she knows it, she can report it (Anscombe, 1957; Mele & Moser, 1994). In case an agent does something unknowingly, she does it unintentional. Distinguishing proximal from distal intentions (Mele & Moser, 1994) is a way out of this problem. Distal intentions are long-term intentions, while proximal intentions can occur more spontaneously. Suppose it is 11 am, and John intends to have a beer in the evening. Here, John has a distal intention. John knows his plan for the rest of the day and can report it. If, however, Lars is on his way home but spontaneously visits his favorite bar, Lars has a proximal intention.

I assume, like Nock and colleagues (2010), if high DS-IAT scores pick individuals out who are at risk for suicide, then the DS-IAT measurement reflects a cognitive structure that is causally responsible for that high risk. But as intentions are never unreportable, while suicide is intentional, I suggest that scoring high on the DS-IAT reflects an implicit cognitive structure that generates *proximal* suicidal intentions. This would partly explain why 78% of patients who die by suicide explicitly deny suicidal thoughts during their last verbal communications before they kill themselves (Busch et al., 2003). Arguably, those individuals were hiding their distal suicide intent and were, thereby, lying. However, claiming that individuals are lying demands an extra justification, as it raises moral concerns. Additionally, as mentioned in the introduction, I want to develop the other explanatory option, suggested by Nock and colleagues (2010), according to which individuals cannot report suicidal intentions.

Furthermore, the research literature on suicide shows that many suicides are not planned actions. Psychologists consider some suicides spontaneous: “The evidence strongly suggests that, for any individual, suicidal impulses come in waves. This implies that if such impulses can find no ready expression, they may pass without the person having harmed him- or herself.” (Williams, 1997, p. 207) For instance, a study (Simon et al., 2001) shows that 24% of suicide attempters self-report that between the decision to end one’s life and acting was a time window of five minutes; another study (Paashaus et al., 2021) indicates that for 42% of suicide attempters there was a 10 minute gap between deciding and acting. In such cases, agents could not report a suicidal intention beforehand.

What kind of implicit cognitive structure could guide action by generating proximal intentions? In Section 3, I broadly introduced goal-system approaches (Kruglanski et al., 2002). A goal-system explains human behavior on the basis of an associative network between goal-representations, related sub-goals and means (action-plans). In fact, a huge body of behavioral and neuropsychological data support goal-system approaches (Custers & Aarts, 2010). A special variant of a goal-system approach is the selfish goal theory (SFGT; J. Y. Huang & Bargh, 2014).

According to the SFGT, human behavior is inconsistent across different situations because a set of conscious and unconscious goals are competing for influence on cognition and behavior.[[6]](#footnote-6) Conscious and unconscious goals share the same properties and play the same functional role within a cognitive system. Every cognitive system, however, only has a limited amount of resources. Activated goals ‘act’ selfish because they are biasing the cognitive system in a way that makes the performance of the goal-related action more probable. An activated goal can, for instance, bias a person’s attention to properties of an object which make a certain action execution more likely.

For instance, one study (Hill & Durante, 2011) suggests that activating the unconscious goal for mating leads women to take higher health risks if taking these risks enhances attractiveness. In the study, single women had to evaluate the attractiveness of men living in the local area. Hill and Durante assume this procedure triggered the goal for mating. In contrast to the control group, these women were significantly more willing to use a free tanning membership and take unhealthy diet pills. Furthermore, they rated themselves as being less likely to get skin cancer and less likely to get heart problems from the diet pills. Accordingly, the study shows that women’s beliefs about health risks of attractiveness-enhancements change in response to activated mating goals. The activated goal unconsciously influences risk judgments and intentional behavior, and the resulting behavior can put the person into risk. The underlying cognitive structures are implicit because the agents did not intentionally start these processes, nor are agents aware of the influences.

The SFGT sheds a new light on the motivational component in intentional action. Goal-system approaches explain dynamic motivational strength for distinct actions. Motivations fluctuate from one situation to another, in contrast to a classic static understanding of motivation (Kruglanski et al., 2002). Goal-system approaches suggest to understand motivations as cognition. Consequentially, goals with related means are called *motivational components*. In philosophy of action, motivation is partly conceptualized with cognitive components, namely with desires (Mele, 1992). A goal-system, which includes conscious and unconscious goals, develops this approach further.

Agents sometimes feel motivated to do something, although the agents are unaware of the underlying causes for this motivation (Sandis, 2012). From the SFGT perspective, a situation can trigger and activate conscious and unconscious goals, which are associated with means. These motivational components compete for influence on behavior by biasing attentional processes and suppressing the execution of different goals. Some goals are especially successful during this process. One explanation is that there might be a single action available which fulfills a set of goals. Furthermore, the brain estimates that this single action is probably successful. As a result, an agent feels the motivation to execute the related action-plan.

The claim that some goal is pulling resources away from other goals may feel purely metaphorical. With the example of attentional processing, I want to briefly illustrate what this can actually mean. The Competition Theory of Selective Attention (Beck & Kastner, 2009; Duncan, 1996) claims that representations in the visual system get attention through a competitive process. A representation of an object will be at the expense of other object’s representations. The competition of what is represented in the visual field happens automatically and unconsciously in parallel, while top-down and bottom-up biasing mechanisms influence the ongoing competition. An agent’s occurrent goal is a top-down influence pre-activating the internal ‘template’ of relevant objects. In this sense, a top-down influence is a biasing neuronal signal which helps to resolve the competition. However, bottom-up mechanisms bias the competition as well. Triggered and automatically working goal-structures amplify those stimuli that are goal-related. Attention, then, results from the competition process, which gains dominance across neuronal systems. For instance, when an agent’s attention is distracted, bottom-up influences, like competing goals, are amplifying aspects of the situation that are incongruent with the agent’s focal goal.

The SFGT claims that a goal’s agenda can be at odds with an individual’s self-related values and overall interests (J. Y. Huang & Bargh, 2014). This potentially leads to non-intentional or unintentional consequences through conscious goal pursuits. Activated goals change a person’s perceptions and expectations in ways which encourage the attainment of the goals’ own end-states. In fact, when persons knowingly act against their self-interest, they feel out of control (Loewenstein, 1996). Such behaviors are commonly coupled with drive states such as thirst, hunger, sexual desire, moods, physical pain or a need for drugs. When persons feel especially motivated to act, the SFGT suggests that related activated goals are successful in reallocating cognitive resources.

In fact, some agents who survived their suicide attempt by jumping from high places reported that they regretted the act in midair. For instance, one person who jumped from the Golden Gate Bridge and survived reported: “I instantly realized that everything in my life that I’d thought was unﬁxable was totally ﬁxable—except for having just jumped.” (Joiner, 2007, p. 53) Another person who survived a jump said: “My first thought was What the hell did I just do? I don’t want to die.” (Joiner, 2007, p. 53) The SFGT explains this behavior in terms of (conscious and unconscious) goals ‘selfishly’ manipulating agents and make persons act against their self-interest.

The outlined structural changes to the motivational system by suicide ideation can be integrated into the SFGT framework. From this perspective, it becomes clear what it means that implicit cognition generates proximal intentions in order to cope with aversive situations. Due to suicide ideation, the suicidal action-plan is highly accessible and functions as a means to different ends (goal end-states). A study (Kleiman et al., 2018) suggests that individuals who engage in suicide ideation experience relief from negative affect after thinking about suicide. In this study, participants reported through their smartphone four times a day on their current affect and whether they were having suicidal thoughts. The data suggest that suicidal thinking decreased sadness and increased positive feelings. However, this repetitive consideration of a suicidal action-plan as a means to an end entrenches the suicidal action-plan in the motivational structure.

For a suicidal individual, an unbearable situation can trigger the goal to escape and thereby the suicidal action-plan, made highly accessible by suicidal ideation, as a means. The SFGT, however, claims more than a purely associative account would. In contrast, it claims that once a goal is activated, it biases attention, memory and perception in a way which contributes to the goal-realization. The highly accessible suicidal action-plan with related goals reallocates cognitive resources to encourage the attainment of the goals’ own end-states. This is important because suicidality is associated with impaired information processing, which, in turn, explains suicidal behavior (for an overview, see Chapter 5 in van Heeringen, 2018). Valerie’s story, mentioned in the introduction, can count as an example. Depending on how persons process information, persons consider a situation as threatening or non-threatening. While Valerie noticed her uncle’s facial expression and considered him angry, Valerie’s biased attention may be causally responsible for this automatic impression formation.

Suicidal individuals have an impaired attention ability, which is the ability to actively select information for processing (van Heeringen, 2018). A variant of the classic Stroop test, briefly introduced in section 3, is the emotional Stroop test, in which, for instance, participants who fear spiders show slower reactions times to name the color of emotional words, like *spider*. The Suicide Stroop is variant of the emotional Stroop test in which participants react to suicide-related words, such as *death* or *funeral*. Priming suicidal agents with such words leads to poor performances (Keilp et al., 2008), like the SFGT would suggest. In fact, studies suggest that suicidal individuals have an attentional bias towards suicide-relevant information (Richard-Devantoy et al., 2016).

Potentially, the SFGT explains why cognitive impairments that are typically associated with suicidality exist or are stronger, like an impaired problem-solving ability (Pollock & Williams, 2004) or memory problems (Williams, 1997).

In sum, I argued for two aspects in this section. First, some suicidal agents cannot report their suicidal intention, because the intention is a *proximal* intention, which spontaneously develops. Second, an implicit cognitive structure is causally responsible for these suicidal intentions. This structure is a network of goal-means associations, which work automatically and the agent is unaware of caused impacts. The selfish goal theory describes this network as competitive, because activated conscious and unconscious goals compete for limited cognitive resources, like attention. Activated goals bias cognitive resources to make the realization of the goal’s end-state more probable. Suicide ideation explains how a goal becomes highly accessible, which makes the goal more influential on perception, attention, and memory. Goals ‘selfishly’ highlight reasons for actions by making certain information salient. Additionally, goal-system approaches suggest that if a goal successfully allocates resources, the agent feels motivation to act correspondingly (basically, a pro-attitude, like an urge, towards an action).

**5 Conclusion**

The assumption that the DS-IAT measures weak associations between death-related and self-related concepts misses the crucial point about the measurement, and it does not explain how the associations relate to suicidal behavior. In contrast, I suggested a cognitive structure that is (a) explanatory for why suicidal agents show less response latencies when undergoing a DS-IAT measurement and (b) that is explanatory for suicidal behavior.

Regarding the first aspect (a), I argued that scoring high on the DS-IAT indicates a weakened self-preservation system. I contend that suicide ideation weakens responses from the self-preservation system in two ways. First, due to intense suicide ideation cognitive processes that are causally responsible for avoidance-responses when sorting death-related to self-related words allocate less resources. Second, behavioral elements of suicide ideation deliver coping-strategies for overcoming impulses of inhibition. Additionally (b), against the background of the Selfish-Goal Theory, the approach helps to explain a higher risk for suicide. The process of suicide ideation generates an influential goal-structure, which, once activated, biases attention, memory and perception in a way which contributes to the goal-realization.

Why do some studies speak in favor of the DS-IAT’s predictive powers (Nock et al., 2010), and other studies do not (Rath et al., 2021)? I hypnotize that unmonitored variables primarily account for the DS-IAT’s weaker or stronger predictive powers. Behavior is not caused by a single factor, but by multiple causes. For example, the main reason to investigate attitudes was the hypothesis that attitudes play an essential role in human behavior. However, it turned out that attitudes alone are not powerful predictors of behavior (for example, see Wicker, 1969). Nevertheless, the research on attitudes continued, and psychologists understood how and under what circumstances attitudes have effects (Maio et al., 2018). I suspect a similar development in clinical psychology for indirect measurements.

For instance, Rath and colleagues (2021) measure suicide ideation with the German Beck Scale for Suicide Ideation (BSS; Kliem et al., 2017). The BSS score ranges from 0 to 38, with higher values indicating a higher suicide risk. Overall, the measurement value ignores the question of *how* suicide ideation takes place. For example, by repetitively considering suicide as a way to reach *different* goals, the suicidal action-plan becomes more valuable. In tendency (see section 3), agents execute actions that reach various goals at once. It should make a difference whether suicide is only considered to escape suffering or, additionally, for self-punishment, revenge, and communicative purposes. Furthermore, the accessibility of suicidal end-means’ associations makes goals more influential, while accessibility depends on repetitive thinking patterns, which are not captured adequately with the BSS measurement value. I consider factors like these to be relevant for the predictive power’s quality.

**References**

Anscombe, G. E. M. (1957). *Intention*. Blackwell.

Arndt, J., Allen, J. J. B., & Greenberg, J. (2001). Traces of Terror: Subliminal Death Primes and Facial Electromyographic Indices of Affect. *Motivation and Emotion*, *25*(3), 253–277.

Banaji, M. R. (2001). Implicit Attitudes Can Be Measured. In H. L. Roediger III, J. S. Nairne, & I. Neath (Eds.), *The Nature of Remembering: Essays in Honor of Robert G. Crowder* (pp. 117–150). American Psychological Association.

Bargh, J. A. (2017). *Before You Know It: The Unconscious Reasons We Do What We Do*. Atria Books.

Bargh, J. A., & Morsella, E. (2008). The Unconscious Mind. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *3*(1), 73–79.

Barnes, S., Bahraini, N., Forster, J., Stearns-Yoder, K., Hostetter, T., Smith, G., Nagamoto, H., & Nock, M. (2017). Moving Beyond Self-Report: Implicit Associations about Death/Life Prospectively Predict Suicidal Behavior among Veterans. *Suicide and Life-Threatening Behavior*, *47*(1), 67–77.

Baston, R. (2018). Two flaws concerning belief accounts of implicit biases. *Philosophical Psychology*, *31*(3), 352–367.

Berger, J. (2018). Implicit attitudes and awareness. *Synthese*, *197*, 1291–1312.

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*(2), 163–176.

Bradley, K. I., Kennison, S. M., Burke, A. L., & Chaney, J. M. (2012). The effect of mortality salience on implicit bias. *Death Studies*, *36*(9), 819–831.

Busch, K. A., Fawcett, J., & Jacobs, D. G. (2003). Clinical correlates of inpatient suicide. *The Journal of Clinical Psychiatry*, *64*(1), 14–19.

Cohen, J. D. (2017). Cognitive control: Core constructs and current considerations. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 3–28). Wiley Blackwell.

Custers, R., & Aarts, H. (2010). The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness. *Science*, *329*(5987), 47–50.

Dacey, M. (2016). Rethinking associations in psychology. *Synthese*, *193*(12), 3763–3786.

D’Atttlio, J. P., & Campbell, B. (1990). Relationship between Death Anxiety and Suicide Potential in An Adolescent Population. *Psychological Reports*, *67*(3), 975–978.

Dawkins, R. (1989). *The Selfish Gene* (2nd Revised Edition). Oxford Paperbacks.

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). Guilford Press.

Devos, T., Huynh, Q., & Banaji, M. R. (2012). Implicit Self and Identity. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of Self and Identity* (Second edition, pp. 155–179). The Guilford Press.

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science*, *13*(6), 513–519.

Gast, A., & Rothermund, K. (2010). When old and frail is not the same: Dissociating category and stimulus effects in four implicit attitude measurement methods. *Quarterly Journal of Experimental Psychology*, *63*(3), 479–498.

Gawronski, B., De Houwer, J., & Sherman, J. W. (2020). Twenty-Five Years of Research Using Implicit Measures. *Social Cognition*, *38*(Supplement), s1–s25.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are “implicit” attitudes unconscious? *Consciousness and Cognition*, *15*(3), 485–499.

Gendler, T. S. (2008). Alief in Action (and Reaction). *Mind & Language*, *23*(5), 552–585.

Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts in mortality salience effects. *Journal of Personality and Social Psychology*, *67*(4), 627–637.

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*(1), 3–25.

Greenwald, A. G., McGhee, D., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology. General*, *143*(3), 1369–1392.

Hansen, G. R., & Streltzer, J. (2005). The Psychology of Pain. *Emergency Medicine Clinics of North America*, *23*, 339–348.

Harrison, D. P., Stritzke, W. G. K., Fay, N., & Hudaib, A.-R. (2018). Suicide risk assessment: Trust an implicit probe or listen to the patient? *Psychological Assessment*, *30*(10), 1317–1329.

Hill, S. E., & Durante, K. M. (2011). Courtship, Competition, and the Pursuit of Attractiveness: Mating Goals Facilitate Health-Related Risk Taking and Strategic Risk Suppression in Women. *Personality and Social Psychology Bulletin*, *37*(3), 383–394.

Houwer, J. D. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353.

Huang, J. Y., & Bargh, J. A. (2014). The Selfish Goal: Autonomously operating motivational structures as the proximate cause of human judgment and behavior. *The Behavioral and Brain Sciences*, *37*(2), 121–135.

Huang, Z., & Wyer, R. S. (2015). *Diverging effects of mortality salience on variety seeking: The different roles of death anxiety and semantic concept activation*. *58*, 112–123.

Joiner, T. (2007). *Why People Die by Suicide* (1st edition). Harvard University Press.

Keilp, J. G., Gorlyn, M., Oquendo, M. A., Burke, A. K., & Mann, J. J. (2008). Attention Deficit in Depressed Suicide Attempters. *Psychiatry Research*, *159*(1–2), 7–17.

Kirkpatrick, L. A., & Navarrete, C. D. (2006). Reports of my death anxiety have been greatly exaggerated: A critique of terror management theory from an evolutionary perspective. *Psychological Inquiry*, *17*(4), 288–298.

Kleiman, E. M., Coppersmith, D. D. L., Millner, A. J., Franz, P. J., Fox, K. R., & Nock, M. K. (2018). Are suicidal thoughts reinforcing? A preliminary real-time monitoring study on the potential affect regulation function of suicidal thinking. *Journal of Affective Disorders*, *232*, 122–126.

Kliem, S., Lohmann, A., Mößle, T., & Brähler, E. (2017). German Beck Scale for Suicide Ideation (BSS): Psychometric properties from a representative population survey. *BMC Psychiatry*, *17*, 389.

Krickel, B. (2018). Are the States Underlying Implicit Biases Unconscious? – A Neo-Freudian Answer. *Philosophical Psychology*, *31*(7), 1007–1026.

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. In *Advances in experimental social psychology, Vol. 34* (pp. 331–378). Academic Press.

Lavender, T., & Hommel, B. (2007). Affect and action: Towards an event-coding account. *Cognition and Emotion*, *21*(6), 1270–1296.

Lehto, R. H., & Stein, K. F. (2009). Death anxiety: An analysis of an evolving concept. *Research and Theory for Nursing Practice*, *23*(1), 23–41.

Lester, D. (1967). Fear of Death of Suicidal Persons. *Psychological Reports*, *20*(3\_suppl), 1077–1078.

Levy, N. (2015). Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements. *Noûs*, *49*(4), 800–823.

Loewenstein, G. (1996). Out of Control: Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292.

Maio, G. R., Haddock, G., & Verplanken, B. (2018). *The Psychology of Attitudes and Attitude Change* (3rd ed.). SAGE Publications Ltd.

Mandelbaum, E. (2016). Attitude, Inference, Association: On the Propositional Structure of Implicit Bias. *Noûs*, *50*(3), 629–658.

Mele, A. R. (1992). *Springs of Action: Understanding Intentional Behavior* (1 edition). OUP.

Mele, A. R., & Moser, P. K. (1994). Intentional Action. *Noûs*, *28*(1), 39–68.

Moreno, M., Gutiérrez-Rojas, L., & Porras-Segovia, A. (2022). Implicit Cognition Tests for the Assessment of Suicide Risk: A Systematic Review. *Current Psychiatry Reports*, *24*(2), 141–159. https://doi.org/10.1007/s11920-022-01316-5

Neimeyer, R. A., Wittkowski, J., & Moser, R. P. (2004). Psychological research on death attitudes: An overview and evaluation. *Death Studies*, *28*(4), 309–340.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *48*(3), 231–259.

Nock, M. K., & Banaji, M. R. (2007). Prediction of Suicide Ideation and Attempts Among Adolescents Using a Brief Performance-Based Test. *Journal of Consulting and Clinical Psychology*, *75*(5), 707–715.

Nock, M. K., Borges, G., & Ono, Y. (2012). *Suicide: Global perspectives from the WHO World Mental Health Surveys* (pp. xi, 391). Cambridge University Press.

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science*, *21*(4), 511–517.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*(1), 44–59.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522.

Paashaus, L., Forkmann, T., Glaesmer, H., Juckel, G., Rath, D., Schönfelder, A., & Teismann, T. (2021). From decision to action: Suicidal history and time between decision to die and actual suicide attempt. *Clinical Psychology & Psychotherapy*, *28*(6), 1427–1434.

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*(3), 277–293.

Pollock, L. R., & Williams, J. M. G. (2004). Problem-solving in suicide attempters. *Psychological Medicine*, *34*(1), 163–167.

Rath, D., Teismann, T., Schmitz, F., Glaesmer, H., Hallensleben, N., Paashaus, L., Spangenberg, L., Schönfelder, A., Juckel, G., & Forkmann, T. (2021). Predicting suicidal behavior by implicit associations with death? Examination of the death IAT in two inpatient samples of differing suicide risk. *Psychological Assessment*, *33*(4), 287–299.

Retterstøl, N. (1993). *Suicide: A European Perspective*. Cambridge University Press.

Richard-Devantoy, S., Ding, Y., Turecki, G., & Jollant, F. (2016). Attentional bias toward suicide-relevant information in suicide attempters: A cross-sectional study and a meta-analysis. *Journal of Affective Disorders*, *196*, 101–108.

Rivers, A. M., & Hahn, A. (2019). What Cognitive Mechanisms Do People Reflect on When They Predict IAT Scores? *Personality & Social Psychology Bulletin*, *45*(6), 878–892.

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, *17*(3), 523–534.

Sandis, C. (2012). *The Things We Do and Why We Do Them*. Palgrave-Macmillan.

Shah, J. Y., & Kruglanski, A. W. (2002). Priming against your will: How accessible alternatives affect goal pursuit. *Journal of Experimental Social Psychology*, *38*(4), 368–383.

Simon, O. R., Swann, A. C., Powell, K. E., Potter, L. B., Kresnow, M. J., & O’Carroll, P. W. (2001). Characteristics of impulsive suicide attempts and attempters. *Suicide & Life-Threatening Behavior*, *32*(1 Suppl), 49–59.

Sohn, M. N., McMorris, C. A., Bray, S., & McGirr, A. (2021). The death-implicit association test and suicide attempts: A systematic review and meta-analysis of discriminative and prospective utility. *Psychological Medicine*, *51*(11), 1789–1798.

Stafford, L. D., & Scheffler, G. (2008). Hunger inhibits negative associations to food but not auditory biases in attention. *Appetite*, *51*(3), 731–734.

Tomasello, M. (2008). *Origins of Human Communication*. A Bradford Book.

Tucker, R. P., Wingate, L. R., Burkley, M., & Wells, T. T. (2018). Implicit Association with Suicide as Measured by the Suicide Affect Misattribution Procedure (S-AMP) Predicts Suicide Ideation. *Suicide and Life-Threatening Behavior*, *48*(6), 720–731.

van Heeringen, K. (2018). *The Neuroscience of Suicidal Behavior* (Illustrated edition). Cambridge University Press.

Van Orden, K. A., Witte, T. K., Cukrowicz, K. C., Braithwaite, S., Selby, E. A., & Joiner, T. E. (2010). The Interpersonal Theory of Suicide. *Psychological Review*, *117*(2), 575–600.

Vigliocco, G., Ponari, M., & Norbury, C. (2018). Learning and Processing Abstract Words and Concepts: Insights From Typical and Atypical Development. *Topics in Cognitive Science*, *10*(3), 533–549.

Wicker, A. W. (1969). Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects. *Journal of Social Issues*, *25*(4), 41–78.

Williams, M. (1997). *Cry of Pain: Understanding Suicide and Self-Harm*. Penguin.

Wilson, T. D. (2009). Know Thyself. *Perspectives on Psychological Science*, *4*(4), 384–389.

1. I think that implicit cognition only plays a crucial role for some types of suicidal agents. [↑](#footnote-ref-1)
2. An individual who considers suicide counts as a suicide ideator. [↑](#footnote-ref-2)
3. One of the strongest risk factors are past suicidal attempts, but the SI-IAT outruns this risk factor. This makes it implausible that the SI-IAT does simply measure the memory-traces of past suicidal attempts. [↑](#footnote-ref-3)
4. I am open to the idea that goal-systems use different underlying mechanisms than associations. [↑](#footnote-ref-4)
5. Arguably, when agents are undergoing a classic IAT measure, similar mechanisms are at play. For instance, if participants are hungry, it is harder for them to sort food-related words to the value-label *bad* (Stafford & Scheffler, 2008). The goal to eat, here a motivation that gained power, influences the participants’ IAT performance, which seems unexplainable with memory’s associations between food and value-based concepts. [↑](#footnote-ref-5)
6. The idea of selfish goals is an analogy to Dawkin’s (1989) selfish genes. [↑](#footnote-ref-6)