Babeș-Bolyai University

Faculty of Psychology and Educational Sciences

Doctoral school "Applied Cognitive Psychology"

IMPLICIT LEARNING OF COGNITIVE STRUCTURES WITH EMOTIONAL COMPONENTS

Summary of the PhD thesis

Doctoral candidate

Răzvan Călin Jurchiș

Scientific coordinator

Professor Adrian Opre, PhD.

Table of Contents

KEYWORDS: implicit learning; cognitive unconscious; emotion; affective states; cognitive-behavioral therapies; dual processing

1. INTRODUCTION

1.1. Implicit learning

Implicit – or unconscious – learning is the process through which one acquires information about the regularities that govern a stimulus domain, the acquired information being unavailable to conscious awareness. However, the learned information exerts some measurable effects on behavior. More simply, it is learning without the awareness of what has been learned (e.g., Frensch & Runger, 2003; Reber, 1967, 1993).

Probably the most intuitive example of implicit learning is that children are able to use, with high accuracy, the grammar rules of their native language, long before they are aware that there are any grammar rules at all. Thus, they have extracted the rules unintentionally and are unaware of them. Also, for persons that know how to ride a bicycle, it close to impossible to verbalize what it is that they do in order to keep their balance.

In cognitive psychology, this phenomenon has been studied in several experimental paradigms. Probably the best known one is the Artificial Grammar Learning task (Reber, 1967). In this task, participants are exposed to several meaningless strings of letters (e.g., XMVTMTR). Unknown to them, these strings are generated from a very complex rule (i.e., an artificial grammar; see Figure 1 for two examples of artificial grammars). After this exposure phase, participants are able to classify new strings into strings that follow the grammar, and strings that do not, even though they are unable to describe the grammar. Therefore, participants have learned the rule, but they are unaware of it.

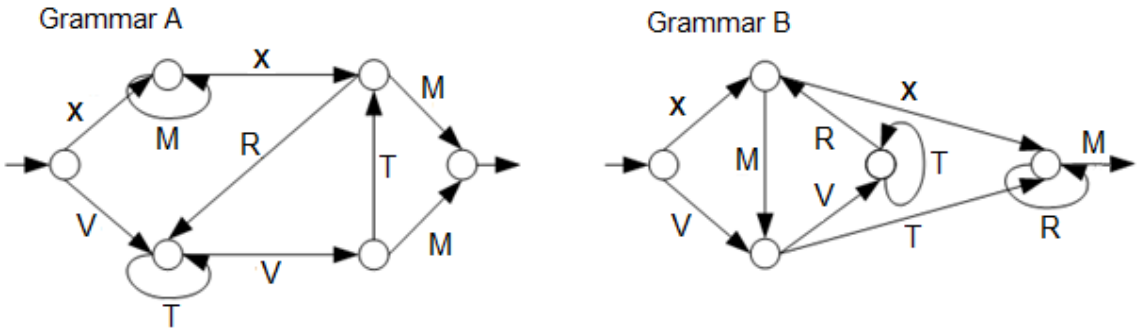

Figure 1. The artificial grammars used in the present thesis (Dienes et al., 1995; Reber, 1967; Norman et al., 2016). Strings are generated by following the order permitted by the arrows. For example, XMXRTVM is consistent only with grammar A and XMVTRXM is consistent only with grammar B.

1.2. The relevance of learning implicit emotional structures

An important task of the human mind is the processing of emotional stimuli. These emotional stimuli have the highest relevance for survival, as they signal dangers, losses, achievements, opportunities, etc. Thus, *if implicit learning detects regularities from the environment in order to enhance the adaptation to these regularities, then it should be particularly predisposed to detect the regularities followed by stimuli with emotional relevance*. In this way, the human cognitive system could anticipate and adapt more rapidly and robustly to the dangers and opportunities existing in the environment.

However, the vast majority of studies from implicit learning have used meaningless and emotionally-neutral stimuli. This is mainly because researchers need to minimize the probability that participants consider as "grammatical" a stimulus due to some previous processing of the stimulus. Thus, stimuli typically need to be as neutral and "sterile" as possible. Consequently, this constraint has resulted in neglecting the possibility of learning implicitly regularities with emotional relevance.

The problem of learning implicit emotional structures has also emerged in a field that is quite distal from the study of implicit learning: the field of cognitive behavior therapies (CBTs). In the treatment of psychopathology, the main objective of these therapies is to change dysfunctional patterns of thinking into more functional ones, via *cognitive restructuring*. However, the restructuring of cognitions expressed by the client sometimes succeeds, but the emotional and behavioral responses remain relatively unchanged. One of the hypotheses often advanced by CBT researchers and practitioners for this paradox is that some symptoms are sustained by knowledge structures that are not accessible to consciousness, hence are not easily modifiable through cognitive restructuring. Moreover, implicit learning has been advanced as the mechanism responsible for the genesis of these structures (Brewin, 1996; Dowd, 2006; Dowd & Courchaine, 2002; Longmore & Worell, 2007; Williams, Wats, MacLeod, & Matthews, 1997). Accordingly, *discovering the processes through which the human mind acquires implicit emotional structures, and the properties of these structures, has the potential to inform the theory and practice of CBT,* by shedding light on how implicit knowledge should be conceptualized and targeted in therapy.

1.3. Previous research on implicit emotional knowledge

Previous studies have investigated whether participants can learn without awareness simple associations between an initially-neutral stimulus (conditioned stimulus; CS) and a positive or negative one (in evaluative conditioning), between an initially-neutral stimulus and a fear-eliciting one (in fear conditioning), or between an action and a positive outcome (in reward learning/instrumental conditioning). Most of these fields have provided mixed, inconclusive results (e.g., Sweldens et al. 2014; see section 1.4. of the thesis, for an updated review). Studies on implicit reward learning seem to provide the strongest support for the possibility of learning unconscious emotional associations (e.g., Leganes-Fonteneau et al., 2018).

However, even studies that have provided reliable evidence for the acquisition of nonconscious emotional associations are subject to several notable constrains and limitations: *First*, they attempted to induce learning of simple CS-US associations, but these simple associations are difficult *not* to be detected consciously. Accordingly, the CS and US have to be altered, weakened, in order to make them difficult to be registered consciously (e.g., through unconscious exposure or by processing them briefly and non-attentionally). This weakening of the CS and US is, at least partially, responsible for the next two limitations. The *second* limitation is that *these created associations did not influence subjective affective ratings of the CS*; the learning effect has been detected only through response-time based measures (e.g., Atas et al., 2014; Leganes-Fonteneau et al., 2018). *Third*, these studies did not provide evidence that the unconscious associations influence responding to other stimuli than the exact CS that has been involved in the conditioning episode. In other words, when participants were unaware of the CS–US association, *the conditioning effect did not generalize to new stimuli that were related to the CS* (e.g., were in the same category). Therefore, it is possible that these lines of study, which need to weaken the stimuli for producing unconscious processing, underestimate the role that non-conscious structures can play in shaping affective processing.

On the other hand, research on implicit learning (of emotionally neutral stimuli) has shown that the human mind is able to extract regularities (a) that are more complex (i.e., involve probabilistic relationships among multiple elements), (b) that influence subjective judgments (thus are not detectable only through subtle, response-time based, measures), and (c) that can influence evaluations of new stimuli that were not involved directly in the acquisition of the regularity. Thus,

it appears that implicit structures can have a stronger impact, when studied under proper conditions. As mentioned above, there are no investigations on whether the human mind can learn implicitly emotionally-loaded regularities that have these properties.

Thus, *studies conducted in implicit learning, which have obtained robust unconscious structures, did not attempt to study structures with emotional components* (as we will see, inducing an emotional component in a structure brings substantial methodological difficulties). *On the other hand, studies on unconscious conditioning have investigated structures with emotional components, but in settings that drastically limit their strength and influence.*

2. GENERAL OBJECTIVES AND OVERVIEW OF THE STUDIES

The overarching objective of the thesis is *to advance our knowledge regarding the implicit learning of structures with emotional components and of the role they could play in clinical phenomena.* From this general objective, we derived four mid-level objectives that are presented next.

As discussed in the introduction, investigations on implicit emotional knowledge have been relatively sparse, and limited to associations between a specific stimulus and an affective component. However, in real life, an important question is whether implicitly-learned structures can be used for evaluating new, previously-unseen, stimuli as having emotional valence. Moreover, learning may occur not by being repeatedly exposed to the exact same situation, but by being exposed to different situations, that follow a common structure.

Therefore, the first mid-level objective is *to investigate whether structures with emotional components can be learned implicitly, and whether these structures can be used for appraising new situations or stimuli that follow the structures.*

A precondition for this objective is to build and validate experimental tasks that allow investigations of implicit structures with emotional components. These tasks would then allow investigating the existence of implicit learning of emotional structures, but also investigations regarding the properties of this phenomenon, or regarding factors that might influence it. In

consequence, the second objective is *to develop experimental tasks that can investigate implicit emotional learning.*

As some CBT researchers have warned (e.g., Dowd, 2006), the CBT theory regarding human cognition is still incomplete, as it focuses almost exclusively on consciously accessible cognitions. Consequently, a third objective is *to facilitate the infusion into CBT theory, of knowledge regarding the functions and properties of implicit structure that has been discovered in cognitive psychology.*

Furthermore, implicit and explicit structures exert their influences through a variety of cognitive processes, mostly through the so-called type 1 processes (automatic appraisals, judgments, heuristics). Research on these type 1 processes can also constitute a resource for improving CBT's theory of human cognition. Hence, the fourth objective is *to synthesize phenomena proposed by the dual-process approach from cognitive psychology, and to use this knowledge for addressing several limitations currently faced by CBTs.*

The first three studies attempt to establish whether it is possible to learn implicitly complex structures with emotional relevance by using three novel experimental paradigms, adapted from existing experimental tasks in implicit learning and evaluative conditioning.

Studies 4 and 5 deal with more ecologically-relevant dimensions of the implicit learning of emotional structures. Specifically, study 4 investigates whether it is possible to learn without awareness socio-emotional regularities, that is, regularities followed by emotional facial expressions. Study 5 investigates the possibility of learning implicitly the structure of scenarios that describe situations which could occur in the real life.

Studies 6 and 7 target the transfer of knowledge regarding implicit knowledge, from cognitive psychology towards the CBT theories of psychopathology and adaptive functioning. *Study 6* synthesizes several properties of implicit structures and explains their relevance for CBT. *Study 7* decomposes the main processes proposed by CBT theories in subjacent processes discovered by the dual process approach to learning, appraisal, judgment, and decision making. Building mostly on *type 1* processes that can be shaped or influenced by implicitly-learned structures, it shows that constructs from apparently distal CBT theories are sustained by common processes. Also, this

reconceptualization offers testable hypotheses and broader research directions that might contribute to the improvement of CBT theory and practice.

3. ORIGINAL RESEARCH

3.1. Study 1. Unconscious Learning of Cognitive Structures with Emotional Components: Implications for Cognitive Behavior Psychotherapies[1]

In two experiments, using a modified version of the Artificial Grammar Learning Paradigm, we found that individuals are able to learn complex cognitive structures (artificial grammars) that associate neutral and emotional contents. Moreover, they learn and use these cognitive structures, at least in part, unconsciously. These results suggest that, in principle, unconscious contents and processes may be involved in the generation of negative affective responses. Implications of these findings for theory, research, and practice of CBT are illustrated. However, we identified several limitations that we attempted to overcome in subsequent studies.

3.1.1. Experiment 1

The **objective** of the study was to determine whether it is possible to learn unconsciously complex structures that associate neutral stimuli with a negative emotion. However, in order to be able to pursue this objective, we first had to design an experimental task that would ensure that the emotion is learnt as an integral part of the structure. An important aspect was that we need to make sure that learning did not occur at the level of specific stimuli that instantiate the emotion, but rather at the level of emotion itself.

Accordingly, we used an adapted version of the Artificial Grammar Learning task (AGL; Reber, 1967). While in the original version participants are exposed to meaningless strings of letters that follow an artificial grammar, we replaced the letters with images, but we kept the same grammars (see Figure 1). Some of the images were neutral, some of them were emotionally-negative (i.e., instantiated sadness). Importantly, whenever the grammar requested a sadness-inducing image, the participant was presented, randomly, with one of the ten possible sadness-inducing image. In this

[1] Published as: Jurchiş, R., & Opre, A. (2016). Unconscious learning of cognitive structures with emotional components: Implications for cognitive behavior psychotherapies. Cognitive Therapy and Research, 40(2), 230-244. https://doi.org/10.1007/s10608-015-9743-z

way, we ensured that if learning occurs, it is at the abstract level of emotion, not at the level of surface stimuli that instantiate the emotion.

We hypothesized that **(1)** participants would learn the grammar, **(2)** they would learn the emotion as an essential part of the grammar, and **(3)** some of their grammar knowledge would be unconscious.

Method

Twenty-one participants (m_{age} = 20.61, SD = 1.85) were exposed, in the ***acquisition phase***, to 30 strings of images. Images were taken from the International Affective Picture System (Lang et al., 1998). Each image composing the string was presented for 2 seconds. Then the next image appeared automatically. Before each string, a blank screen appeared, followed by the text "The next string". We did not mention anything about the existence of a rule, of a grammar, that structured the order of the strings.

In a ***first test phase*** (Test 1), we exposed participants to new strings: 20 strings that followed the same grammar as in the acquisition, 20 strings that followed a different grammar. At this point, we informed participants that that there was a rule, but we did not mention anything about the rule's configuration. Participants' task was to respond for each string, whether it followed the grammar or not. We expected that if participants have learned the grammar, they will be able to classify strings into grammatical and nongrammatical with above-chance accuracy.

In order to determine whether participants have learned the grammar consciously or unconsciously, we used two different methods, widely used in implicit learning research. First, we used *response attributions*, asking participants to report after, each classified string, which was the basis of the classification, choosing from Guess, Intuition, Familiarity (which indicate unconscious knowledge, e.g., participant has a feeling of intuition that the string follows the grammar, but has no idea about the grammar's structure), Rules, and Recollection (which indicate conscious knowledge). Participants were provided with precise definitions and instructions regarding the meaning of each response attribution (Dienes & Scott, 2005; Mealor & Dienes, 2013). *The second method* consisted in asking participants, after each classification, to report how *confident* they are that their response was correct. Responses ranged from 50% confidence (which indicated that being correct or incorrect were equally likely) to 100% confidence (which indicated certainty that

the response was correct). If participants have as much confidence when they are wrong as when they are right, it means that their conscious confidence is not predictive of accuracy, thus their accuracy is sustained by unconscious knowledge (e.g., Dienes et al., 1995; Wierzchon et al., 2012).

For determining more accurately whether participants have learnt the emotion as an integrative part of the grammar, we designed two additional test phases. In Test 2, we exposed ten strings that followed the acquisition grammar in all respects, except that we removed the emotion from its componence. Thus, whenever the grammar would have requested an emotional image, now we simply omitted it, and exposed the next image from the grammar. Also, we exposed ten strings that were nongrammatical. If participants have learned the emotion as an essential part of the grammar, we expected them to be unable to discriminate grammatical from nongrammatical strings, when the emotion was missing.

Finally, Test 3 had the same objective like Test 2 – to determine whether the emotion was part of the grammar learned by participants. We reasoned that participants might not learn that the emotion was part of the grammar, but rather that at certain points in the grammar (where the sadness-inducing images appeared) there were some *randomly appearing images*, not necessarily emotional images. Thus, we replaced the emotional images with randomly-appearing neutral images.

Results

We found that participants' accuracy in classifying strings into grammatical and non-grammatical was above the chance level, m = 54.7, $t(20) = 2.68$, $d = 0.54$, **which indicate that they have learned the grammar.**

Also, we analyzed participants' accuracy split on response bases, in order to determine whether participants had accurate conscious and unconscious knowledge. Participants had above-chance accuracy when they relied on Intuition and Familiarity (taken together, as both map the same types of knowledge, cf Dienes & Scott, 2005), $m = 56.6\%$ (SD = 10.05), $t(20) = 2.87$, $p = .009$, $d = 0.6$. Therefore, **they acquired accurate unconscious knowledge.**

Also, when participants relied on conscious Rules and Recollection, their accuracy was above chance, $m = 55.47\%$ (SD = 11.11), $t(18) = 2.14$, $p = .04$, $d = 0.49$.

When participants attributed their classifications to Guessing, accuracy did not exceed chance, $t(18) = -0.60$, $p = .55$, $d = -0.15$.

The finding that participants acquired accurate unconscious knowledge was also sustained by the analyses conducted on confidence ratings, which indicated that participants were as confident in accurate responses (m = 70.28%, *SD* = 7.39) as in inaccurate ones (m = 69.96%, *SD* = 8.48), $t(20) = 0.34$, $p = .73$, $d = 0.09$.

Moreover, as standard null hypothesis significance test cannot prove the null, which was important in order to argue for a lack of difference in confidence, we conducted a Bayesian analysis. Bayes factors, which result from Bayesian analyses, provide continuous measures of evidence for specific hypotheses, with values greater than 1 indicating support for the alternative hypothesis, and values lower than 1 providing support for the null. By convention, *values less than 0.33 indicate substantial support for the null, values greater than 3 indicate substantial support for the alternative hypothesis*, and values between 0.33 and 3 indicate data insensitivity. We stress that**,** *unlike p-values, Bayes factors are able to provide support for the null hypothesis, relative to the specific alternative hypothesis considered* (e.g., Dienes, 2008, 2011, 2014)**.** Following Dienes (2008, 2011), we modeled the alternative hypothesis as a half-normal distribution, with the *mean* estimated from similar past research, and the *SD* calculated as mean/2. Dienes and Scott (2005) identified a mean difference in confidence between correct and incorrect responses of 3%. Accordingly, the alternative hypothesis was modeled as a normal distribution, with mean = 3 and SD = 1.5. The resulting Bayes factor, $B_{N(3;\ 1.5)} = 0.18$, provides substantial evidence for the null. That means, participants were equally confident in their correct and in their incorrect classifications, indicating that their conscious, subjective knowledge of the strings' grammatical status was irrelevant for the accuracy.

For determining whether participants learned the emotion as an essential part of the grammar, we compared whether their accuracy in Tests 2 and 3 exceeded chance level. It did not: For test 2, m = 45.8% (*SD* = 13.4), $t(17) = -1.31$, $p = .20$, $d = -0.31$, $B_{N(5;\ 2.5)} = 0.13$; for test 3, m = 45.8% (*SD* = 13.6), $t(17) = -1.31$, $p = .20$, $d = -0.31$, $B_{N(5;\ 2.5)} = 0.15$. Note that it was again important to check whether data support the null, thus we had to conduct Bayesian analyses. Now the prior distribution was modeled based on accuracy from Test 1. Therefore, we conclude that **participants learned that the emotion was an essential component of the grammar.**

Discussion

In this experiment, we found evidence that participants have learned the grammar, and that they have learned it, at least in part, unconsciously. Moreover, they have learned the emotion as an essential part of the grammar.

However, we identified several limitations of the present experiment. The most important one was that the neutral images from the grammar depicted inanimate objects, while the emotional ones depicted animate beings (human and non-human animals). Consequently, participants might have learned that the grammar associated stimuli depicting inanimate objects with stimuli depicting animate beings, not emotional and neutral stimuli. Also, we used only one grammar, thus the phenomenon might be restricted to the particular grammar we used.

Accordingly, we conducted a second experiment, which attempted to replicate the results of the first one, while overcoming the mentioned limitations.

3.1.2. Experiment 2

This second experiment had the same objective, hypotheses, and used the same task as the first one. We modified only the stimuli: if in the first one the neutral stimuli depicted objects, and the negative ones depicted animate beings, in the present experiment both emotional and neutral stimuli depicted animate beings. Therefore, the only difference between these categories of stimuli was the emotional load; they were now equivalent at the semantic level.

Second, we used a two-grammar design, widely employed in AGL studies, in which one group of participants learn one artificial grammar (e.g., Grammar A), the other group learns another grammar (e.g., Grammar B). Then, in the test phase, both groups of participants receive strings from Grammar A and strings from Grammar B. However, for participants that were exposed in acquisition to Grammar A, the new Grammar A strings are “grammatical”, while the Grammar B strings are “nongrammatical”; vice-versa for participants that have been exposed in acquisition to Grammar B (e.g., Dienes & Altmann, 1997; Dienes & Scott, 2015; Norman & Price, 2012). In this way, it is ensured that, if participants from both groups, together, exhibit above-chance accuracy, the learning effect cannot be explained by the fact that strings from one grammar are, a

priori, more likely to be chosen, even if participants do not learn the grammar. Also, this design ensures that the learning effect is not tied to a particular grammar, but is a more general one (i.e., appears at least for two different grammars).

Method

The methodology was almost identical to experiment 1, except that half of participants were now exposed to strings from grammar A in the acquisition, while the rest of participants were exposed to grammar B.

Thus, 36 participants (10 men; m_{age}= 22.55, SD = 3.95) underwent first the acquisition phase, then the three test phases: Test 1, Test 2, Test 3.

Results

The classification accuracy in Test 1 was 54.91% (*SD* = 10.31), significantly above the chance level, t(31) = 2.69, *p* = .011, *d* = 0.49. This indicates that ***participants have learned the grammars.***

For responses based on Guessing, Intuition and Familiarity, participant's performance did not exceed chance level: all *t*s <0.6, *p*s > .55, *d*s < 0.1. For responses based on conscious Rules and Recollection, the accuracy was significantly greater than the chance level, *m* = 59.51% (*SD* = 23.27), *t*(29) = 2.77, *p* = .009, *d* = 0.50. Therefore, ***according to this measure, participants acquired only accurate conscious knowledge.***

However, when we compared their confidence in correct responses with confidence in incorrect responses, we obtained no difference *t*(31) = 0.65, *p* = .51, *d* = 0.13, $B_{N(3;\ 1.5)}$ = 0.23. This suggests that participants' conscious evaluations regarding the accuracy of responses were unrelated to their actual accuracy. Accordingly, ***accuracy was mostly sustained by unconscious knowledge*.** The fact that the two awareness measures yield contradicting results might be explained by the fact that conscious responses represented only 35.95% out of the total number of responses; therefore, they were not enough to make a difference in the overall analysis of confidence in correct and incorrect answers.

Similar to experiment 1, participants' accuracy in test 2 and test 3 did not exceed chance (For test 2: *t*(30) = 0.38, *p* = 0.70, *d* = 0.06, $B_{N(5;\ 2.5)}$ = 0.33; For test 3: *t*(30) = 0.06, *p* = .94, *d* = 0.01, $B_{N(5;}$

$_{2.5)}$ = 0.22), which indicates that ***participants have learned the emotion as an essential part of the grammar.***

Discussion

This experiment replicated the findings of Experiment 1, using an improved design and stimuli. It showed that participants learned the grammars that associated neutral stimuli with a negative emotion; also, there is evidence that participants were largely unaware of what they have learned, although the two measures of awareness yielded contradicting results.

General discussion

The study reported here is *the first attempt to include an emotion in a complex cognitive structure and to investigate whether that cognitive structure can be learned unconsciously*. Both experiments included in the article revealed that, as hypothesized, participants learned the cognitive structure and were largely unaware of it. As it is typical for implicit learning studies, some amount of conscious knowledge was also developed.

Although this study provides evidence that might shed light on the link between unconscious cognition and emotional responses, the validity of its conclusions are limited by the employed paradigm. First, we operated on the assumption, supported by some implicit learning theorists, that participants, in AGL, acquire an abstract representation of the grammar (e.g., Reber, 1990; Matthews, 1990). On the other hand, other researchers have shown that participants are able, in particular variants of the AGL task, to discriminate grammatical form non-grammatical strings based on specific micro-rules, such as bigrams and trigrams (e.g., Perruchet & Pacteau, 1990). Therefore, it is difficult to exclude, in the present design, that participants based their classifications on knowledge that did not include the emotional contents. We attempted to verify if this is the case through Tests 2 and 3. They showed that participants had lower accuracy when the emotion was eliminated from the string. However, these differences might have been inflated by the fact that Tests 2 and 3 always followed after Test 1, so their order was not counterbalanced. The reason for not counterbalancing the test phases was that we expected a relatively small learning effect (based on Norman & Price, 2012), therefore interposing other phases between acquisition and Test 1 would have, likely, reduced the learning effect even more. On the other hand, studies have found that a learned grammar can be retained even two years after it has been

acquired (Allen & Reber, 1980), thus interposing Test 1 between acquisition and Tests 2 and 3 should not erase any knowledge of the grammar.

This limitation might be overcome in future studies by testing whether participants are able to discriminate grammatical and non-grammatical bigrams (i.e., two successive elements) that contain the emotion. Bigrams in which one of the element is the emotion can be classified accurately only based on knowledge that involves the emotion.

In conclusion, the present study provides evidence that it is possible to learn complex structures that include emotional components, although future improvements of the paradigm are needed.

3.2. Study 2. The Emotional Sequence Learning Task[2]

The objective of the present study was to determine whether it is possible to learn implicitly an abstract structure of emotions, irrespective of the specific stimuli that instantiate these emotions. To this end, we devised a novel experimental paradigm, called the *Emotional Sequence Learning Task.* In the first phase (acquisition), participants were exposed to a succession of negative, positive, and neutral words. Unknown to participants, words appeared upon a predetermined regularity (e.g., neutral–positive–negative–positive–neutral-negative). On each iteration of the sequence, the specific word that instantiates the emotion was extracted randomly, but the abstract sequence of emotions remained constant. Thus, we expected participants to learn the abstract sequence of emotions, regardless of the specific words. In the second phase (test phase), we presented participants with sequences of three words (triplets). For each triplet, they had to respond whether it appeared in the acquisition phase or not. Importantly, all triplets were new, but some of them were incompatible with the acquisition sequence (e.g., neutral-positive–neutral), while others were compatible with the acquisition sequence (e.g., neutral–positive–negative). If participants learned the regularity from the acquisition phase, we expected them to consider a larger proportion of compatible triplets as seen in the previous phase, compared to incompatible triplets. The results showed that overall, participants did not consider as having seen more often the triplets that conform to the sequence. However, they considered so when they relied on feelings of familiarity. Therefore, the present study partially supports the possibility of learning implicitly an abstract sequence of emotions.

[2] Presented as: Jurchiș, R., Costea, A., and Opre, A. – Implicit learning of emotional sequences. 6th Seminar on Implicit Learning, 25-26 Mai 2017, Budapest, Hungary
Also presented at 9th International Congress on Cognitive Psychotherapy, June, 2017, Cluj-Napoca, Romania

Introduction

Starting from the limitations of the previous study, we intended to pursue further the **objective** of determining the possibility of learning an abstract structure of emotions, irrespective of the specific stimuli that instantiate these emotions. Different from the previous study, we intended to create a task in which it would be ensured that if learning is detected, the learned regularity is (a) abstract, (b) relatively complex, and (c) that the detected learning effect cannot be the explained by the acquisition of non-emotional regularities.

We designed the acquisition phase of the present task starting from the Serial Reaction Time Task (SRTT; Nissen & Bullemer, 1989). In this task, participants respond to stimuli that appear successively on the screen. Unknown to them, the order in which they appear follows a complex repeating pattern. Participants' response times decrease as they respond to the sequence, and decrease when the repeating pattern is changed. Thus, participants learn the repeating pattern, so they are able to anticipate the stimuli when they follow the pattern. However, when the pattern is suddenly changed, their response time increase, because they can no longer anticipate the stimuli.

The present study

We designed a task in which participants are exposed to an abstract emotional sequence. That is, specific stimuli are unpredictable, but the emotion they instantiate is fully predictable. In order to determine whether structures that are more complex than those typically used in evaluative or classical conditioning can be learned implicitly, we used a second-order conditional sequence, in which one always needs to learn combinations between two elements in order to predict a third one. After the learning phase, participants' knowledge of the sequence is probed with a false-memory based test phase (adapted from Paciorek & Williams, 2015). Specifically, participants are exposed to sequences of three emotional stimuli. Their task will be to respond whether the three stimuli have appeared together in the exposure phase. However, unknown to participants, none of the exposed sequences have appeared in the acquisition, but some of them follow the abstract pattern of emotions, while other violate it. If learning occurs, participants should falsely-remember having seen the triplets that follow the pattern more often than the triplets that did not follow it. Therefore, we hypothesized that participants will learn the sequence, even without being aware of it.

Method

We first validated a list of words, in order to ensure that they have the desired emotional valence, and also to match them regarding arousal, familiarity, imaginability, length (i.e., number of letters). This was an important step, because we needed to eliminate the possibility that words differ systematically in order respects than the emotional valence. For example, if positive, neutral, and negative words would have differed regarding familiarity, participants could have learned a rule that concerns familiarity, not emotional valence.

Participants

Participants were 66 undergraduate students (m_{age}= 21.10, *SD* = 1.88, 88% women) that enrolled voluntarily or for course credit.

Procedure

Acquisition phase. The words were exposed on the screen individually. Each word remained on the screen for 2,000 ms, then the next word appeared automatically. Participants were instructed to read, understand, and process the meaning of the word to the best of their ability.

Unknown to participants, the order in which the words appeared was determined by a sequence of emotions: Neutral – Negative – Positive – Negative – Neutral – Positive. This sequence is a second-order conditional one, because every emotion can be fully predicted based on the previous two emotions. For example, after Neutral, both Negative and Positive appear in the sequence, but the pair Neutral – Negative always predicts Positive, or Neutral – Positive always predicts Neutral and so on. As already mentioned, the specific word that instantiated the emotion was always chosen randomly, out of the 15 words that were available for each emotion.

The emotional sequence was exposed 15 times in each block. There were four acquisition blocks, thus the sequence was exposed 60 times, summing 360 acquisition trials. After each block, participants had a 45 seconds break. The entire acquisition phase lasted approximately 15 minutes.

Test phase (assessment of learning and awareness). The acquisition sequence can be split in six different sequences of three emotions (triplets), presented in Table 2. As explained above, because the sequence was a second-order conditional one, these triplets were the simplest predictive units.

In the test phase, participants were exposed to triplets. More specifically, three words appeared on the screen successively, each word being exposed for two seconds. After all three words appeared on the screen, participants were prompted to respond whether they had seen the words appearing in that specific order in the previous phase. Unknown to participants none of word triplets had appeared in the acquisition phase. All were new, but half of them conformed to the abstract pattern of emotions, while the other half violated the pattern of emotions. We assumed that the learned structure will increase the familiarity with triplets that follow the structure, and, as a consequence, participants will falsely-remember having seen more triplets that follow the structure. The triplets that violated the pattern of emotions were constructed by changing the last emotion (see Table 2).

In the test phase, we exposed participants to four instances from each of the 12 triplets. That is, there were four different word triplets with the structure Neutral – Negative – Positive, four triplets with the structure Neutral – Negative – Neutral, and so on. In total, there were 24 specific triplets that followed the acquisition pattern and 24 that were inconsistent with it.

After the test phase, participants filled a **post-experimental questionnaire**, which had two purposes. *The first one* was to probe whether participants were aware of the sequence. *Second,* we found in a pilot study that some participants used idiosyncratic rules and criteria in the test phase in order to determine which triplet appeared and which did not. However, most of these rules had nothing to do with the sequence, therefore they could decrease the proportion of responses consistent with the sequence. For example, some participants tried to see whether the words from the triplet were somehow related, or whether they followed an alphabetical order etc. Consequently, the second purpose of the questionnaire was to determine whether participants used such irrelevant decision criteria.

Proportion of responses consistent with the pattern

The main dependent variable was the proportion of responses according to the pattern. A response consistent with the pattern was either responding with *Yes, I saw it in the previous phase* to a triplet that conforms to the pattern of emotions or responding with *No, I did not see it in the previous phase* to a triplet that violates the pattern.

When we compared whether the proportion of responses according to the pattern exceeded the chance level for participants that used relevant strategies, the difference was not significant $t(50)$

= 1.58, p = .06, d = 0.22. Therefore, we found no convinging evidence that participants' responses were significantly influenced by the pattern, hence that participants learned the pattern.

Table 3. One-sample t tests comparing accuracy with the chance level for different response attributions

		t	df	p	Mean accuracy	d_z
Guess	Student's t	0.11	39.00	0.456	0.50	0.02
Intuition	Student's t	0.38	39.00	0.353	0.51	0.06
Familiarity	Student's t	2.68	38.00	0.005	0.56	0.43
Memory	Student's t	0.25	32.00	0.401	0.51	0.04

Note. H_a population mean > 0.5

However, it is possible that, overall, participants did not respond significantly according to the pattern, but they did so only when they relied on a specific response attribution. Therefore, we compared to chance level participants' responses, split on response attributions. Results are presented in table 3. ***Participants' accuracy was above chance only when they relied on Familiarity, which indicates implicit structural knowledge.***

In sum, when taking into account all responses, irrespective of the response attribution, we failed to find evidence that the pattern of emotion influenced participant's recognition judgments. However, when we took into account responses based on feelings of familiarity, we did find evidence for responding according to the pattern.

Discussion

We consider that the present study offers some evidence for the implicit learning of an abstract pattern of emotion, but this evidence is relatively weak, for two reasons. First, we expected that the learning effect will be detectable when taking into account all responses, but we did not find evidence for this hypothesis. Second, we found a learning effect when participants relied on familiarity, which indicates implicit knowledge, but we did not find the same effect for the other response attributions that indicate implicit knowledge (guess, intuition), hence this result is also inconsistent with our predictions. Nonetheless, there may be a reason why the emotional pattern produced effects only when participants relied on feelings of familiarity. Dienes and Scott (2008)

demonstrated that in the Artificial Grammar Learning task, one of the main mechanisms that help participants discriminate grammatical from non-grammatical strings is *familiarity calibration*. Specifically, participants compare the familiarity of the current string with the average familiarity of the previously classified strings. Thus, if the familiarity of the current string is higher than the average familiarity of the previously classified ones, the string will be more likely classified as grammatical; contrarily, it will more likely be classified as non-grammatical. It is possible that this mechanism of familiarity calibration also played a role in the present study, leading to classifications consistent with the pattern only when participants reported to rely on Familiarity.

Therefore, although there might be some learning effect, we consider that the present task, in its current form, is not very practical to use in order to study implicit emotional learning.

It is possible that, with some modifications, the present task could induce a stronger learning effect. Specifically, instead of words, emotional and neutral images could be used as stimuli, as they might be stronger emotional inducers. Additionally, instead of asking participants to give dichotomous responses (*Yes, it appeared* or *No, it didn't appear*), participants could respond by giving continuous familiarity ratings. In this way, they could rely more on their feelings of familiarity towards the strings, that could be influenced more by the implicitly learned structure.

In conclusion, the present study provides preliminary results to support the possibility of implicit learning of second-order emotional sequences, but additional investigations with an improved versions of the task are recommended in order to produce clearer results.

3.3. Study 3. Evaluative conditioning of artificial grammars: Evidence that non-conscious structures bias affective evaluations of novel stimuli[3]

The present study proposes a new paradigm for investigations of implicit learning of emotional information, which is a combination between an implicit learning task and an evaluative conditioning (EC) one. EC is the process through which an initially-neutral stimulus (conditioned stimulus; CS) acquires emotional valence, after being paired with an emotional stimulus (unconditioned stimulus; US). An important issue in EC research regards whether, when participants are unaware of the CS-US contingency, the affective valence can generalize to new stimuli that share similarities with the CS. Previous studies have shown that generalization of EC effects appears only when participants are aware of the contingencies, but we suggest that this is because the contingencies typically used in EC research are salient and easy to detect consciously. We report a preregistered study in which participants (N = 217) were exposed to letter strings generated from two complex artificial grammars that are difficult to decipher consciously. Stimuli from one grammar were paired with positive USs, while those from the other grammar were paired with negative USs. Subsequently, participants evaluated new, previously-unseen, stimuli from the positively-conditioned grammar more positively than new stimuli from the negatively-conditioned grammar. Importantly, this effect appeared even when trial-by-trial subjective measures indicated lack of grammar awareness. We provide evidence for the generalization of EC effects even in the absence of awareness, *providing, thus, evidence that non-conscious structures can bias affective evaluations of novel stimuli.*

Introduction

The present study used a different approach from the previous ones: while in the first two studies we used rules that structured positive, neutral, and negative stimuli, in this study we exposed

[3] Submitted for publication: Jurchiș. R, Costea, A., Dienes, Z., Miclea, M. & Opre, A,. (under review). Evaluative conditioning of artificial grammars: Evidence that non-conscious structures bias affective evaluations of novel stimuli. Journal of Experimental Psychology: General.
The study is also available as a preprint: https://psyarxiv.com/jkfu5

participants to regularities that structure initially-neutral stimuli. However, some of these structures are associated with negative and some with positive stimuli. Thus, we expected participants to attribute positive or negative valence to stimuli that follow the structures, even when they are unaware of the structures. *Also different from the previous studies, we will directly evaluate whether the learned emotional structures influence participants' affective responses.*

We propose a task in which each participant learns two different grammars. One of the two grammars is associated with a negative emotion, while the other grammar is associated with a positive emotion. More precisely, in an acquisition phase, letter strings generated according to one grammar are paired with negative stimuli (i.e., images), while strings generated according to the other grammar are paired with positive stimuli. For brevity, we call the grammar associated with negative stimuli "the negative grammar", and the grammar paired with positive stimuli, "the positive grammar". In a subsequent test phase, we expect that participants will evaluate *new* strings that follow the "positive grammar" more positively, compared to new strings that follow the "negative grammar". In this case, it means that we have obtained a generalizable conditioning effect, because participants evaluate previously unseen strings. Importantly, we will also assess, for each evaluated string, whether participants are aware of what make the string more or less pleasant, because our main objective was to test whether generalization can appear without awareness. We hypothesized that (i) there will be an overall conditioning effect, (ii) a conditioning effect when participants use only unconscious knowledge and, also, (iii) a conditioning effect based on conscious knowledge. Furthermore, we compared the unconscious with the conscious effects, in order to determine (iv) whether awareness enhances the generalization of the EC.

Method

The hypotheses, procedure, data collection, and statistical analyses were preregistered, before the data collection began, here: https://osf.io/cbdu6/

Participants

A number of 240 undergraduate students underwent the task, in exchange for course credit; 23 failed the attention/engagement checks (see Procedure), therefore the final sample is composed of 217 participants (176 women, m_{age}= 19.94 years, *SD* = 4.14).

Materials

In the acquisition/conditioning phase, we presented 32 strings from each grammar. In the test phase, participants had to evaluate 20 strings from each grammar. In the acquisition phase, the strings from one grammar were always presented together with negative images, while the strings from the other grammar were always presented with positive images. For counterbalancing, for some randomly-determined participants, grammar A was the positive grammar (and grammar B the negative one); conversely for the rest of participants. We used 23 positive and 23 negative images, taken from IAPS (Lang et al., 1997), NAPS (Marchewka et al., 2014), and OASIS (Kurdi et al., 2017). It was randomly determined which particular image appeared with which particular string.

Procedure

Acquisition/conditioning phase. In this phase, participants were expected to learn both grammars and to associate them with emotions. On each trial, the emotional image appeared first. Then, after 1.5 seconds, the string appeared just below the image. The string remained on the screen, together with the image, for 7.5 seconds, then the next trial began automatically. Participants were instructed to memorize which string appears together with which image. They were not informed that there are regularities in how the strings were constructed.

The acquisition phase was divided in four blocks: two for grammar A and two for grammar B. After each block, participants had a 30 seconds break. In each block, participants were exposed to all 32 strings from one grammar, each string being paired with an emotional image. Thus, in each block, the participant saw strings that followed only one grammar and images that had only one valence (positive or negative), in order to facilitate learning the grammar, by reducing interference with the other grammar and to benefit from emotional carry-over effects.

Test phase (valence rating and awareness measure). Participants were presented, unexpectedly, with new letter strings, without images. They had to evaluate, for every string, *how much they like or dislike it*, using a Likert scale with values ranging from -5 (strongly dislike) to +5 (strongly like).

After rating the string's valence, participants had to respond to an *awareness scale*, while the string was still on the screen. This was adapted from awareness scales that are widely-used in AGL studies (e.g., Dienes & Scott, 2005; Norman et al., 2016; Wierzchon et al., 2012):

To what extent are you aware of what makes the string likeable/dislikeable for you?

1- *I don't have any clue about what makes it likeable/dislikeable*
2- *I am more or less guessing what makes it likeable/dislikeable, but I could not describe what it is that makes it likeable/dislikeable*
3- *I think I know what makes it likeable/dislikeable*
4- *I know what makes it likeable/dislikeable*

Responses 1 and 2 denote that participants are not aware of the structures that were associated with the positive or the negative affect, while responses 3 and 4 indicate that participants have some conscious knowledge of the structures (Dienes & Scott, 2005; Ramsøy & Overgaard, 2004; Wierzchon et al., 2012).

Results

In order to determine whether there was an overall conditioning effect, we compared the ratings received by all strings that followed the positive grammar with those received by the strings that followed the negative one, irrespective of the conscious/unconscious basis of the evaluation. A repeated-measures *t*-test revealed, as expected, that strings from the positive grammar were evaluated more positively than those from the negative grammar, $t(216) = 7.69$, $p < .00001$, Cohen's $d_z = 0.52$, $m_{diff} = 0.538$, 95%CI [0.40; 0.67], $B_{H(0;\,0.66)} = 10^{12}$.

When testing whether there was a conditioning effect based on unconscious knowledge (Figure 2a), we found, again, that the strings from the positive grammar were evaluated more positively than those from the negative grammar, $t(208) = 4.38$, $p = .00001$, $d_z = 0.30$, $m_{diff} = 0.317$, 95%CI [0.17; 0.46], $B_{H(0;\,0.66)} = 3{,}133.45$.

The effect was also present when participants relied on conscious knowledge (responses 3 and 4 from the awareness scale; Figure 2b), $t(181) = 6.35$, $p < .00001$, $d_z = 0.47$, $m_{diff} = 1.06$, 95%CI [0.73; 1.39], $B_{H(0;\,0.66)} = 8 \times 10^{7}$.

Finally, we found that the conditioning effect based on conscious knowledge was higher than that based on unconscious knowledge, $t(173) = 4.60$, $p < .00001$, $d_z = 0.35$, $m_{diff} = 0.822$, 95%CI [0.47; 1.18], $B_{U(0;\ 1.59)} = 11{,}993$.

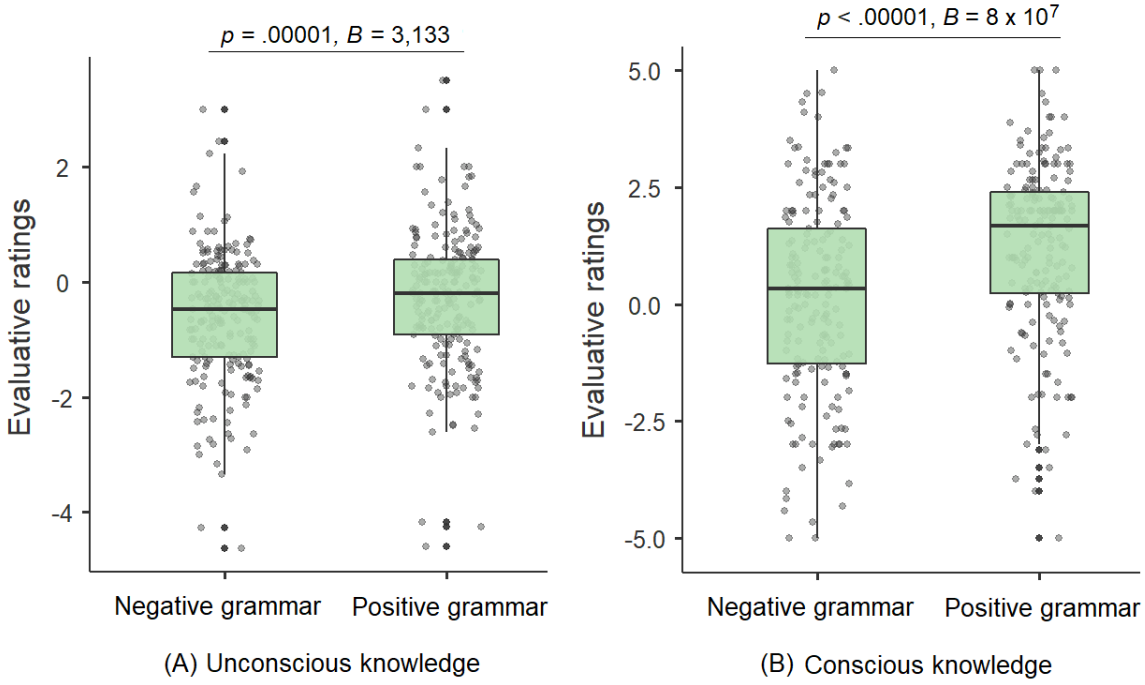

Figure 2. Evaluative ratings received by strings from the positive and the negative grammars when participants used (A) unconscious knowledge and (B) conscious knowledge.

An unexpected, but very robust, result was that when participants did not have conscious knowledge of the structures they evaluated strings more negatively than when they had conscious knowledge. This difference appeared both for strings from the positive grammar $t(188) = 11.34$, $p < .00001$, $d_z = 0.83$, $m_{diff} = 1.70$, 95% CI [1.41; 2.00], $B_{H(0;\ 0.66)} = 14 \times 10^{25}$ and for those from the negative grammar, $t(179) = 4.79$, $p < .00001$, $d_z = 0.36$, $m_{diff} = 0.76$, 95% CI [0.45; 1.08], $B_{H(0;\ 0.66)} = 19{,}382$.

This result could illustrate that, during the test phase, participants were consciously attempting to figure out some structures in the strings, and not succeeding was an aversive experience. Additionally, it might be the case that conscious knowledge regarding the string's structures makes it more predictable, familiar, and hence more positive (e.g., Ivanchei & Asvarich, 2018). Regardless of its explanation, this unanticipated difference provides evidence that the unconscious and conscious responses from the awareness scale reflect different types of knowledge and are used in different situations by participants.

In sum, we found strong support for (i) an overall conditioning effect, (ii) an effect based on unconscious knowledge, (iii) an effect based on conscious knowledge, and (iv) that the conscious effect was higher than the unconscious one.

Discussion

In the context of the present thesis, this study offers the most robust evidence for the unconscious learning of emotional structures: the study was pre-registered, the sample size was relatively large (n = 217 in a within-subjects design), and the Bayesian and frequentist statistical indicators both showed robust differences.

In conclusion, the present study proposes a novel, robust, method for investigating unconscious emotional knowledge, showing that non-conscious structures can influence affective responses.

3.4. Study 4. Implicit and Explicit Learning of Socio-Emotional Information in Depression[4]

This study had a twofold objective: First we aimed to determine whether it is possible to learn without awareness regularities followed by socio-emotional stimuli. Second we tested whether and how depression affects this learning process. The socio-cognitive deficits associated with depression might impair the implicit learning of socio-emotional regularities, a process that is essential for adaptive social functioning. In contrast, research on adaptive functions of depression suggests that depressed individuals are, in some circumstances, more accurate than non-depressed controls in detecting regularities. However, an assessment of implicit learning of socio-emotional regularities in depression has not been conducted yet. In the present study, we modified the Artificial Grammar Learning task, adapting it to induce implicit and explicit learning of regularities that structure strings of emotional facial expressions. Sixty-two depressed and 134 non-depressed participants, classified by their Beck Depression Inventory scores, underwent the task. We found, first, *that participants acquired the regularities implicitly and explicitly*. Also, we found that *implicit social learning was intact in depressed participants, and explicit social learning was enhanced in depressed participants*. Therefore, the present study is the first to show implicit learning of social regularities with emotional relevance, and the first to test how depression affects this process. We discuss the implications of these results for the study of implicit learning in depression and for theories of adaptive functions of depression.

Introduction

The present study follows a particular line of investigation regarding implicit learning of emotional structures. Specifically, it focuses on the possible involvement of this process in *extracting emotional regularities that govern the social domain*. Past research has advanced implicit learning as one of the central processes involved in the development and functioning of social cognition

[4] Submitted for publication: Jurchiș, R., Costea, A., & Opre, A. (under review). Implicit and explicit learning of socio-emotional information in depression. Frontiers in Psychology
Presented as: Jurchiș, R., Costea, A., & Opre, A – Implicit and explicit learning of socio-emotional information in depression. 7th Seminar on Implicit Learning, 24-26 May 2018, Cluj-Napoca, Romania

and social skills (Dowd, 2006; Jurchiş & Opre, 2016; Lieberman, 2007; Norman & Price, 2012). One's behavior in social situations, or one's inferences about others' behavior and mental states, seem to be partially governed by associations and regularities that have been learned non-consciously (e.g., Hudson, Nijboer, & Jellema, 2012; Lewicki, 1986). However, there are very few studies that investigated to what extent people indeed learn implicitly such regularities, because there are relatively few tasks that are can investigate implicit social learning. Therefore, *the first aim of the study was to design a task that would afford investigations on the implicit learning of socio-emotional information.*

In addition to bringing a contribution towards clarifying the role of implicit learning in the social domain, *such a task would enable investigations on the role of implicit social learning in various disorders.* For instance, people with depressive symptoms have difficulties in inferring others' mental or emotional states (Air, Weightman, & Baune, 2015; Bora & Berk, 2016; Ladegaard, Larsen, Videbech, & Lysaker, 2014), and have lower levels of social skills (Hames, Hagan, & Joiner, 2013). Therefore, it is possible that, since they have difficulties in perceiving accurately the social world, they also have difficulties in learning its regularities.

On the other hand, the *analytical rumination hypothesis of depression* (Andrews & Thomson, 2009) states that, despite the costs it incurs on patients, depression might serve some adaptive functions. More specifically, given the increased levels of rumination associated with depression, depressed individuals can analyze more thoroughly the situations that have led to the depressive episode, situations that tend to be interpersonal and emotionally-relevant in nature (Andrews & Thomson, 2009; Badcock et al., 2017; Beck & Bredemeier, 2016; Durisko et al., 2015). Consequently, given their predisposition to the conscious analysis of information, depressed individuals might be, in some circumstances, more accurate in detecting regularities [other theories that make similar predictions are the *depressive realism hypothesis* (Alloy and Abramson, 1988), and the *affect-as-information theory* (Schwarz and Clore, 2003)]. In order to disentangle these predictions, the second aim was *to examine the impact of depression on implicit emotional learning.*

The Present Study

The present study **had two main aims**: The first one is to develop and test a task that is capable to produce implicit learning of socio-emotional regularities. The second one is to investigate learning

of regularities followed by socio-emotional stimuli, particularly by emotional facial expressions, in depression.

To these ends, we have modified a "standard" AGL task, by creating strings of emotional facial expressions that follow an artificial grammar. Based on the literature review presented above, **we tested two alternative hypotheses regarding the influence of depression, as follows**: (a) that *depressed participants will have lower performance* in learning the socio-emotional regularities, compared to non-depressed ones and (b) *that depressed participants will have higher performance* in learning the socio-emotional regularities.

Method

In the present study were included 196 students (172 women, mean age = 19.96 years, *SD* = 3.43).

Participants with Beck Depression Inventory-II (BDI – II) scores from 0 to 13 constituted the *non-depressed* group. Participants with scores ranging from 14 (the cut-off for *mild depression*) to maximum constituted the *depressed* group.

Stimuli

For the AGL task we deployed two widely-used artificial grammars. From Scott and Dienes (2008), we took the letter strings built upon these grammars, and replaced each letter with a different facial expression (X was replaced with fear, M – happiness, R – disgust, T – calm, V – anger). Thus, we changed the surface stimuli that compose the strings, but preserved their underlying grammars. Importantly, we intended to rule out the possibility that participants learn the grammars based on non-social perceptual features (e.g., shape, dimensions, configuration of the faces) but based on the socio-emotional features of the stimuli (i.e., facial emotional expressions). To this end, a single person depicted all facial expressions, thus only the emotional expression varied between stimuli and all other perceptual features were kept constant (see Norman and Price, 2012, for a similar approach). Each string was composed of 5 to 9 expression, which appear simultaneously on the screen.

Procedure

As in the previous experiments, participants underwent, first, an acquisition phase, in which they were exposed to the strings of facial expression. Unknown to them, the strings followed an

artificial grammar. After acquisition, there was a tes phase, in which participants were exposed with new strings, and they had to respond, for each string, whether or not it follows the previous grammar. As in the previous studies, after classifying each string as grammatical or non-grammatical, they had to report the basis for their classification, choosing from: Guess, Intuition, Familiarity, Rules, Memory (remembering/recollection)

Results

Did learning occur?

In order to determine whether participants learned the grammars, we compared their accuracy from the test phase (i.e., proportion of strings classified correctly) with the chance level of .50.

*Table 1***.** One Sample T-Tests comparing classification accuracy for different response attributions against the chance level of .50

					95% CI for Mean Difference		
	t	**df**	**p**	**Mean Difference**	**Lower**	**Upper**	**Cohen's d**
Overall	11.95	195	< .001	0.10	0.09	0.12	0.85
Guess	0.88	163	0.38	0.02	-0.02	0.06	0.07
Intuition	5.57	183	< .001	0.08	0.05	0.11	0.41
Familiarity	7.11	185	< .001	0.11	0.08	0.14	0.52
Rules	3.56	110	< .001	0.09	0.04	0.14	0.34
Memory	8.42	161	< .001	0.18	0.14	0.22	0.66

Note. Mean difference is calculated by subtracting the chance level of .50 from the proportion of correct responses given by participants. Thus, for example, a mean difference of 0.10 means that participants classified correctly a proportion of .6 (or 60 %) of the total strings.

Table 1 shows that participants learned both implicit and explicit knowledge regarding the grammars. Specifically, they had above-chance performance for all response attributions, except for *guessing*.

Between-Groups Comparisons

Analytical approach. For assessing learning performance of depressed and non-depressed participants, we used both Null Hypothesis Significance Testing (NHST) and Bayesian analyses.

In the context of the present study, the Bayesian analyses provide two important advantages. *First, they are able to provide evidence for the null hypothesis* (e.g., Dienes, 2008, 2011, 2014). This is an important aspect, because, if there are no differences between depressed and nondepressed participants, we will be able to test whether the nonsignificant result is due to insufficient statistical power or due to the fact that the two groups have equivalent performance. For all Bayesian between-groups comparisons, following Ziori and Dienes (2015), we used an informed prior distribution, modelled as a half-normal with the mode of 0 and the *SD* equal to the overall learning effect obtained in our study. The overall classification accuracy of participants was .603, thus the *SD* was set at .603 -.50 = .103 (where .50 is the chance level).

Second, the Bayesian approach gives the possibility to test multiple alternative hypotheses on the same data[5]. Hence, for each between-groups comparison, we calculated two Bayes factors: The first one, noted as B_{ADV}, indicates the extent to which data are consistent with *an advantage for the depressed group* (i.e., higher accuracy for depressed participants), relative to the null. The other one, noted as B_{DEF}, indicates the extent to which data are consistent with a *deficit in the depressed group*, relative to the null.

Overall learning. A between groups t-test did not detect differences between the overall accuracy of non-depressed (m = 0.601, SD= 0.11) and depressed participants (m = 0.608, SD = 0.14), $t(194)$= 0.36, p = .723, B_{ADV} = 0.25, B_{DEF} = 0.14, M_{diff} = 0.007, 95% CI [-0.03; 0.04], d = 0.06. Both Bayes factors indicate that *data are consistent with the null hypothesis of no difference between depressed and non-depressed participants.*

Implicit knowledge. We found *strong evidence against the hypothesis that depressed participants (n = 62, m = 0.597, SD = 0.151) have a deficit in implicit knowledge*, relative to non-depressed participants (n = 133, m = 0.575, SD = 0.118), $t(193)$ = 1.10, p = .272, B_{ADV} = 0.56, **B_{DEF} = 0.09**, M_{diff} = 0.02, 95% CI [-0.02; 0.06], d = 0.17.

[5] Of course, as long as the alternative hypotheses make different predictions regarding how the data will look like. For a rationale, see the section "The answer to the question should depend on the question" in Dienes & Mclatchie, 2018; see Ziori & Dienes, 2015, for a recent example of AGL study that tested two alternative hypotheses on the same data.

Explicit knowledge. We conducted a mixed ANOVA, with *response attribution* (Rules and Memory) as a within-subjects factor, and *depression level* (depressed vs non-depressed) as a between-subjects factor. This revealed a significant *response attribution by depression level interaction* [$F(1, 93) = 5.73$, $p = .019$, $\eta^2_p = 0.058$]. The main effect of depression level was non-significant [$F(1, 93) = 1.94$, $p = .167$, $\eta^2_p = 0.020$], and so was that of *response basis* [$F(1, 93) = 2.37$, $p = .127$, $\eta^2_p = 0.025$].

Planned comparisons revealed that *depressed participants had more accurate conscious knowledge of the Rules* (n = 35, $m = 0.69$, $SD = 0.24$), compared to non-depressed participants (n = 60, $m = 0.55$, $SD = 0.27$), $t(93) = 2.55$, $p = .012$, $\mathbf{B_{ADV} = 11.64}$, $B_{DEF} = 0.14$, $M_{diff} = 0.139$, 95% CI [0.03; 0.25], $d = 0.54$.

For classifications based on conscious memory (*Remembering*), there were no significant differences between depressed ($m = 0.66$, $SD = 0.26$) and non-depressed participants ($m = 0.68$, $SD = 0.24$), $t(93) = 0.39$, $p = .695$, $B_{ADV} = 0.35$, $B_{DEF} = 0.64$, $M_{diff} = 0.02$, 95% CI [-0.08; 0.13], $d = 0.08$. Here both Bayes factors are insensitive, but B_{ADV} is close to excluding an advantage for depressed participants.

Discussion

The present study has shown that *artificial grammars that structure sequences of emotional facial expressions of a single person can be learned both implicitly and explicitly*. Consequently, this is one of the first studies to show implicit learning of social regularities with emotional valences.

Also, the present study has found consistent evidence that *implicit learning of socio-emotional stimuli is intact in depression*, as depressed and non-depressed participants had the same accuracy when they classified strings into grammatical and non-grammatical.

Depressed (and highly-depressed) participants were more accurate in classifying strings when they used conscious knowledge of the Rules, which indicates that *depression is associated with better explicit detection of the grammars*. However, given the novelty of our task, and the relatively small sample of participants that have used the Rules attribution, a dose of caution is warranted in interpreting these results, until they are replicated in future studies. From a theoretical standpoint, this result is consistent with the predictions of the *analytical rumination hypothesis* (Andrews & Thompson, 2008), that depressed persons allocate more cognitive resources to social stimuli and

situations, in order to disentangle their complexities. A complementary explanation for this enhanced performance is based on the *affect-as-information* theory, which posits that a negative mood signals a possible problem in the environment, switching the processing style from a global one towards a more analytical, local, one (e.g., Gasper and Clore, 2002; Schwarz and Clore, 2003). As a result, depressed participants in our study could preferentially process local regularities, for example bigrams or trigrams (i.e., pairs of two or three successive emotional expressions allowed by the grammar; Pothos, 2007; see Shang et al., 2010, for a discussion). Previous studies have shown that conscious knowledge of bigrams and trigrams can sustain a high level of performance in AGL (e.g., Perruchet and Pacteau, 1990; Pothos, 2007). In conclusion, it is plausible that depressed participants employed more local processing, this helped them learn local but adaptive micro-rules, which, in turn, sustained their increased performance when they responded based on conscious rules.

Limitations

A first limitation of the study is that we classified participants as depressed and non-depressed based only on their BDI score, in the absence of a clinical interview, which limits the validity of our classification. Also, our depressed group had lower mean BDI scores (22.02) compared to other studies that investigated implicit learning in depression (e.g., 33.70 in Janacsek et al., 2018). Thus, it is possible that the results would have been different in a more severely depressed sample.

Future directions

In order to provide a more fine-grained assessment of the analytical rumination hypothesis, one could test whether rumination mediates the relationship between depression and classification performance, especially when using Rules. Also, we propose that, although it is still difficult, these hypotheses should be tested in even more ecologically-relevant paradigms, such as virtual-reality environments that expose participants to interactions with avatars, maybe over extended periods of time (e.g., several days/weeks). In this way, the observed implicit social learning would be even closer to what happens in real life interactions.

Conclusion

Using a modified experimental paradigm that is suitable to assess implicit learning of socio-emotional information, the present study has found (a) that it is possible to learn implicitly sequences of emotional facial expressions, (b) that the overall performance of depressed participants is not impaired, (c) that implicit (unconscious) acquisition is also not impaired, and (d) that explicit social learning can be enhanced by depression.

3.5. Study 5. Implicit Learning of Regularities from Scenarios that Describe Real-Life Situations[6]

The present study tested the possibility of learning implicitly emotional structures from events that can occur in real life. Specifically, we exposed participants to scenarios that described situations regarding a fictitious target-character. Unknown to participants, whenever the action was caused by the character, the valence of the situation was positive; whenever the situation was caused by another person, and the character was affected by it, the valence was negative. In a test phase, participants evaluated as more familiar new situations that followed the mentioned regularity, compared to new scenarios that violated the regularity. This result appeared even for participants that did not become aware of the regularity. In conclusion, it appears that *it is possible to learn implicitly the structure of ecologically-relevant scenarios.*

Introduction

The relevance of the results from implicit learning tasks in explaining real-life psychological phenomena is limited by the artificial, sterile, nature of the tasks they are obtained in. Thus, a demonstration of the phenomenon using more ecologically-valid tasks is necessary. Using a novel experimental task, we aim, in this study, to investigate if it is possible to learn implicitly the structure of scenarios that describe real-life situations.

Previous studies that have investigated learning the structure of scenarios that describe real-life situations have actually been conducted not in implicit learning research, but in the study of interpretation biases. For example, in one of the tasks used for this purpose, participants are presented with the description of scenarios that are ambiguous regarding the valence of the

[6] Presented as: Jurchiș, R., Costea, A., & Opre, A. –Implicit learning from scenarios that describe real-life situations. The *22nd Congress of the Association for the Scientific Study of Consciousness,* 26-29 June, 2018, Krakow, Poland

outcome. Scenarios are disambiguated by the last word from their composition, either in a positive manner or in a negative manner (Matthews & Mackintosh, 2000). The results typically show that participants exposed to positively-disambiguated scenarios are more likely to interpret, subsequently, *new* ambiguous scenarios positively (while those exposed to the negatively-disambiguated scenarios interpret new scenarios negatively; e.g., Matthews & Mackintosh, 2000; Hallion & Ruscio, 2011). Therefore, they are able to induce some form of learning. Several researchers in the field have argued that this learning is mainly implicit (e.g., Peters, 2009). Indeed, this form of learning is implicit in the sense that it is non-intentional and it takes place through direct exposure. But, to the best of our knowledge, there are no studies to investigate whether participants are conscious or not of the learned structures.

In the present study, we intend to test whether learning the structure of scenarios can take place without awareness. Importantly, following the objectives of the present thesis, one dimension of this regularity will be the emotional valence of the scenarios. Furthermore, in order to decrease the chances that the regularity will be detected consciously, we increased the complexity of the rule, by adding another variable in the structure of the scenarios. Specifically, we varied the agent that caused the situation described in the scenario. We choose this dimension because we reasoned that it is adaptive to register and to anticipate the causes of emotionally-relevant events, thus this dimension would have high chances to be learned implicitly (see Leung & Williams, 2011).

In sum, we designed a task in which we expected that participants would learn a regularity consisting of a covariance between the identity of the agent/patient and the emotional valence of the scenario. We tested this hypothesis in two experiments that had the same design, with the second experiment bringing slight modifications regarding the experimental task.

3.5.1. Experiment 1

In this experiment, we exposed participants to scenarios that described situations concerning a character, called Mihai (which is a typical Romanian men name, translated as *Michael*). The regularity was that, whenever Mihai caused the action (i.e., he was the agent), the valence of the situation was positive; whenever somebody else caused the situation, and Mihai was affected by it (i.e., he was the patient), the situation had a negative valence.

For determining whether participants have learned the structure, we included, after acquisition, a phase in which participants were exposed to scenarios that had an ambiguous valence. Therefore, they neither violated, nor followed the regularity. In a subsequent test phase, these scenarios were modified so that half of them followed the structure from the acquisition (Mihai caused – Positive; Somebody else caused – Negative), while the other half violated the structure (Mihai caused – Negative; Somebody else caused – Positive). For example, the ambiguous scenario *Mihai is seen explaining something during the course break* is changed in the rule-following *During the break, Michael patiently clarifies learners' queries.* The test phase was a false-memory task, adapted from Paciorek and Williams (2015). Specifically, we exposed rule-conforming and rule-violating scenarios, and asked participants, first, to rate how familiar they are for them and, second, to respond whether they think the scenario appeared in the acquisition or not. Even though none of the scenarios appeared previously, we reasoned that, if participants learn the structure, it will bias participants' recollections, make them more likely to respond that they have seen the rule-conforming scenarios than the rule-violating ones. Also, we expected them to give higher familiarity ratings to the rule-conforming ones.

Method

Participants were, in total, 56 students (mean age = 20.34 years; *SD* = 2.22, 87% females), out of which 33 participated in the first experiment and 23 in the second one.

Before conducting the experiment, we validated all scenarios, making sure that the emotional scenarios have the intended emotional valence. Also, it was crucial to make sure that grammatical and non-grammatical scenarios presented in the acquisition phase do not differ systematically in other respect, except for the fact that some are grammatical and the others are nongrammatical. Our attempt to ensure this equivalence consisted in controlling for comprehensibility, emotion, semantic relatedness with acquisition scenarios, semantic relatedness to ambiguous scenarios, length, voice (active vs passive), and position of the target character (i.e., whether *Mihai* was situated at the beginning or at the end of the phrase describing the scenario).

Procedure. ***Acquisition phase.*** Participants were exposed to the 12 acquisition scenarios and to the 12 neutral scenarios, in random order. They were not informed about the existence of regularities. The cover task asked participants to read, understand, and evaluate the emotional valence of presented situations. We gave participants this task in order to increase the chances that

they process the meaning and the emotion of the scenarios. Note that detecting explicitly the emotional valence of scenarios should not lead to explicit learning of the regularity, as the regularity concerned the *covariation between the emotion and the agent/patient dimension* not the emotional valence per se. Each scenario was presented for 5 seconds, that, according to our pre-tests, were sufficient for reading and understanding it, but that limited the possibility to search for patterns or regularities.

After this phase, we included a distractor task, in which participants had to solve 10 multiplications, with numbers composed of two to five figures. Then they had a five-minute break.

Test phase. Participants were informed that they are going to see scenarios, some of which have been presented in the preceding phase, some of which did not. Unknown to them, all scenarios were new, but kept some elements from the neutral scenarios. Crucially, half of them followed the acquisition structures, the other half did not. Participants' tasks were (1) to respond whether they think the scenario had appeared in the previous phase or not (Yes/No) and (2) to rate the familiarity of each scenario on a 1 to 99 scale.

Finally, participants had to respond to an awareness questionnaire that asked them whether they detected *any pattern, rule, or regularity regarding the scenarios presented in the first phase of the experiment*. Also, they were explicitly instructed to *report even regularities they are unsure about, or, if they did not detect any regularity, to try to guess which was the regularity.*

Results

The post-experimental questionnaire revealed that only two participants were aware of the pattern. Both reported that they became aware retrospectively, when they were requested by the awareness questionnaire to think about the regularity; nevertheless, we excluded them from subsequent analyses.

For determining whether participants learned the pattern, we compared, first, participants' responses regarding which scenarios appeared in the previous phases were consistent with the rule above the chance level. Responses consistent with the rule were *Yes* to a rule-conforming scenario, or *No* to a rule-violating one. Indeed 57% (*SD* = 7.78) of the total responses were consistent, which was above-chance: $t(31)= 4.9$, $p < .001$, $d = 0.8$.

Second, we compared the familiarity ratings obtained by the rule-conforming scenarios with the rule-violating ones. As expected, the rule-conforming ones were evaluated as more familiar than the rule-violating ones, m_{diff}= 8.50, SE_{diff} = 2.29, $t(32) = 3.70$, $p < .001$, $d = 0.64$.

Therefore, *both continuous familiarity ratings and binary decisions regarding previous apparitions of the scenarios sustained our hypothesis, that participants can learn the structure of scenarios.* Importantly, we included in these analyses only participants that did not detect consciously the pattern, therefore, we conclude that *they learned the structures implicitly.*

3.5.2. Experiment 2

We conducted a replication experiment, in which we slightly modified the procedure. The objective, hypotheses, and the design were identical to those from Experiment 1. The acquisition and test scenarios were also the same, but we eliminated the ambiguous scenarios from which, in Experiment 1, we derived the test scenarios. This was in order to reduce the possibility of having imbalances between rule-consistent and rule-inconsistent scenarios regarding their level of relatedness with acquisition scenarios; hence, we intended to eliminate one possible confounding variable. Instead, we just presented, individually, the words from the test scenarios, in a random order. In this way, we intended to induce semantic priming, which can facilitate false memories, but at a more granular level (i.e., at word/concept level), thus not at the level of situation.

Method

In the acquisition phase, we exposed the only the 12 acquisition scenarios (as mentioned, the 12 ambiguous scenarios were eliminated). After this phase, we exposed all words that composed the test scenarios, individually, in a random order. Of course, we did not mention that they will appear as part of scenarios in the test phase (or that there will be a test phase). Participants' cover task was to rate the level of familiarity for each word. Subsequently, they underwent the distractor task, identical to Experiment 1. The test phase was also identical to that from Experiment 1: we exposed six rule-conforming and six rule-violating scenarios, and asked participants to respond whether they think they saw previously the scenario and to rate their familiarity with the scenario. Finally, participants filled the same awareness questionnaire as in Experiment 1.

Results

None of the participants detected consciously the rule. As expected, participants responded according to the rule when asked whether they think they saw the scenario in the previous phase: m = 55% (SD = 10), $t(21) = 2.3$, $p = .017$, $d = 0.49$. Also, they evaluated rule-conforming scenarios as more familiar than rule-violating ones, m_{diff}= 9.04, $SE_{diff} = 2.62$, $t(20) = 3.44$, $p = .001$, $d = 0.75$.

Thus, *replicating experiment 1, results showed that participants learned the rule implicitly*, which led to falsely-remembering having seen the rule-conforming scenarios more often than the rule-inconsistent ones and, also, to rating as more familiar than the rule-consistent scenarios.

Discussion

The present study has found that participants can learn implicitly the structure of scenarios that describe real-life situations. Specifically, they learned the relationship between the agent/patient dimension and the emotional valence of scenarios.

This study seems to be the first to show implicit learning from scenarios, being, therefore, a possibly important step towards demonstrating the often suggested, but not yet convincingly proven, involvement of implicit learning in acquiring the regularities that govern daily life. Furthermore, this study is also relevant for the involvement of implicit learning in the social realm, as all scenarios involved interactions between Mihai and other persons; hence, the regularities governed Mihai's social interactions. In conclusion, this study demonstrates implicit learning of emotionally-loaded structures that govern real-life, social, situations

Limitations

Our attempt to ensure that grammatical and nongramatical scenarios from the test phase were equivalent, consisted in controlling for comprehensibility, emotion, semantic relatedness with acquisition scenarios, semantic relatedness to ambiguous scenarios, length, voice (active vs passive), position of the target character (i.e., whether *Mihai* was situated at the beginning or at the end of the phrase describing the scenario). However, it is very difficult to make sure that we have identified all relevant variables that can influence participants' familiarity evaluations.

Another limitation was that we needed to keep a relatively small number of scenarios, so that we could manage to have balanced scenarios in the two categories (consistent and inconsistent with the rule); as the number of variables that need to be taken into account when generating scenarios

was relatively high, it was increasingly difficult to generate new scenarios that satisfy all the constrains. Therefore, the number of scenarios was relatively low (12 acquisition, 12 test scenarios), which leaves open the question regarding how generalizable is the learning effect.

In conclusion, this first investigation of implicit learning from scenarios that describe real-life situations provides promising results for a more ecologically-relevant approach to implicit learning. Clearly, we also appreciate that replications of the phenomenon are needed, using various other regularities and scenarios, and using more detailed assessments of the acquired knowledge and of the processes through with it exerts influence on participants' responses.

3.6. Study 6. Fundamental Properties of Implicit Knowledge Structures: Implications for Cognitive Behavior Therapies[7]

The present chapter overviews properties of implicit knowledge discovered in cognitive psychology and describes their relevance for cognitive behavior therapies. It shows, first, that implicit structures are learned by direct exposure to environments or stimuli governed by a structure or by acting/responding according to a structure. Second, one can develop both conscious and unconscious knowledge regarding a structure, and, third, implicit structures themselves can generate conscious knowledge. Hence, these conscious contents might mask sometimes the presence of unconscious structures. Fourth, implicit structures can be counteracted by alternative structures, but they cannot be erased. Consequently, they can be sub-activated, but they can re-activate in some circumstances. Fifth, implicit structures are learned better in childhood and adolescence. Moreover, the paper suggests several directions of research that could clarify the role of implicit structures in CBTs and, more generally, in psychopathology and human adaptation.

Introduction

A claim often made by researchers of CBTs is that the efficiency of this family of therapies could increase if, in addition to targeting conscious thoughts and beliefs, the therapy would also target effectively implicit cognitive structures. Their central argument is that, often in the process of therapy, restructuring the conscious cognitions succeeds, but the emotional and behavioral symptoms do not disappear. Thus, if the conscious cognitions have been restructured, the residual symptoms must be maintained by structures that are non-conscious, non-verbalizable, hence are non-amenable to cognitive restructuring (e.g., Brewin, 1996; David & Lynn, 2010; Dowd, 2006; Dowd & Courchaine, 2002; Williams, Watts, MacLeod, & Mathews, 1997). These unconscious structures have been studied in cognitive psychology, specifically in the field of implicit learning, a phenomenon defined as the acquisition of information that takes place without the awareness of

[7] Published as: Jurchiș, R. (2018). Fundamental properties of implicit knowledge structures: Implications for cognitive-behavior therapies. *Journal of Evidence-Based Psychoterapies,* 18(2), 117-129. Doi: 10.24193/jebp.2018.2.17

what has been learned (e.g., Reber, 1967, 1993). In this sense, the mentioned CBT researchers have encouraged the CBT community to borrow from data on implicit learning, and to transfer knowledge generated in this area into CBT's theory and practice. However, this transfer has been, with very few exceptions (e.g., Dowd, 2006), almost non-existent.

The present chapter attempts to stimulate the infusion of knowledge from implicit learning research into CBT; more specifically, it synthesizes some properties of unconscious knowledge, analyses their relevance for psychotherapy, and provides suggestions for how these implicit cognitions should be conceptualized and targeted in CBT.

Properties of Implicit Structures

(1) Implicit Knowledge Structures Are Learned Through Direct Exposure to Stimuli that Follow a Structure and/or by Behaving According to a Structure

Implicit knowledge is acquired by exposing the person directly to the stimuli that follow an implicit structure (as in AGL), and/or by acting according to a structure (as in SRT). Making inferences about the structure, testing hypotheses about it, or having it described to you, results in explicit, not implicit, knowledge. For example, a person that memorizes the dictionary and the grammar rules of a language previously unknown to her, would, most likely, have more difficulties in speaking the language, compared to a native speaker that has no idea what a grammar rule is. Compared to memorizing explicitly the grammar rules, being exposed to the language and practicing it gives one the implicit knowledge necessary to have a good mastery of the language.

While this characteristic may seem obvious, it has some important, less obvious, implications. A first implication is that, because implicit knowledge is acquired by direct exposure to stimuli, or by acting directly according to a structure, implicit knowledge should often have a higher level of activation, hence more influence on emotional or behavioral processes, compared to knowledge acquired purely intellectually. Being present in an environment affords a more direct and accurate assessment of the respective environment, thus it might be usually adaptive to give this type of knowledge preeminence in the human cognitive system. Furthermore, being directly exposed to a stimulus has more motivational (and sometimes emotional) relevance than thinking or reasoning about it, and knowledge with motivational and emotional relevance tends to be more activated in memory (see Jurchis & Opre, 2018, for a detailed discussion). Therefore, *trying to counteract an implicit structure with cognitions derived purely intellectually cannot be very effective.* This leads

to a second implication: The best method for creating adaptive implicit structures is through direct exposure to stimuli that follow a structure and/or by acting according to a structure. These are mainly behavioral strategies already used in CBTs: various types of exposure, behavioral activation, behavioral experiments, acting *as-if* (e.g., Dowd, 2006).

Let's take the example of a child that tries to befriend several classmates, but is rejected by them. Consequently, he develops a tendency to isolate from his peers. If his parents persuade him that not all his classmates would reject him, he might believe them, but he could still be fearful in approaching other peers. Thus, the implicit association between *trying to befriend somebody* and *being rejected*, which was learned through direct experience, will have more impact on the child's behavior and emotions, than the explicit belief learned from his parents. A more efficient strategy would be to help the child befriend peers that are likely to respond positively. Thus, the child would be directly exposed to the contingency "approaching others – being accepted", he could acquire it, and his emotions and behavior might be influenced more by this newly acquired association.

(2) One Can Acquire Both Implicit and Explicit Knowledge Regarding a Structure

Often in implicit learning studies, participants develop both implicit and explicit knowledge. For instance, Haider, Eichler, and Lange (2011) have shown that in the SRTT, participants develop, first, implicit knowledge, which sustains the decreases in reaction time. Then, they might consciously detect that their responding is more fluent, and this detection triggers an explicit search for patterns, which, in turn, leads to the development of explicit knowledge (see also Esser & Haider, 2017). As a result, participants would have now both implicit and explicit knowledge. Also, in AGL, it has been repeatedly demonstrated that participants acquire both implicit and explicit knowledge (e.g., Dienes & Scott, 2005; Norman & Price, 2012). This double coding brings important difficulties for implicit learning researchers, because it is very difficult to demonstrate convincingly that participants' performance is sustained by implicit knowledge, when explicit knowledge is also present (e.g., Shanks, 2010; Shanks & St John, 1994). This double coding can also set a trap for cognitive psychotherapists. Specifically, given that CBT models emphasize the role of verbalizable cognitions and that implicit cognitions are more difficult to detect (as they are unreportable by the clients), therapists might be tempted to focus only on restructuring the

conscious cognitions, and ignore the fact that there might also be some implicit structures that generate reactions regarding the same activating event.

In the example with the child rejected by his peers, the child, besides learning implicitly the covariation between trying to befriend colleagues and rejection, might also develop negative conscious expectancies regarding related to approach other peers. Thus, a therapist might restructure these conscious expectancies, but fail to establish a new tacit association between approaching others and being accepted.

(3) Implicit Knowledge Structures Can Generate Conscious Contents

Another interesting phenomenon is that implicit knowledge structures can produce conscious knowledge. Dienes and Scott (2005) proposed a model that explains how different knowledge types inter-relate in implicit learning tasks. They showed that participants, first, learn information about the structure followed by the stimuli (e.g., about the artificial grammar), called *structural knowledge*. Then, when they have to judge, for example, whether a new string follows the structure or not, they exhibit judgment knowledge (e.g., *Yes, it follows* or *No, it doesn't follow the grammar*). The most common situation in AGL and SRT is that participants are able to express, consciously, whether a string or sequence follows the learned structure, but they are unable to articulate any knowledge regarding the structure. Thus, they have conscious judgment knowledge based on unconscious structural knowledge. This conscious judgment based on unconscious structures is typically experienced as a feeling of intuition or of familiarity (e.g., *I have a feeling that the string follows the structure, but I have no idea what this feeling is based on*). Indeed, the majority of knowledge expressed in AGL and SRT tasks is of this type (Dienes & Scott, 2005; Fu et al., 2010; Jurchis & Opre, 2016; Norman & Price, 2012).

A similar pattern could to be at work in the generation of some emotional responses: Often when experiencing an emotion, the person can express consciously, relatively easily, the automatic thought(s) that contain different judgments (e.g., appraisals, inferences) regarding the activating event. However, according to the CBT theory, these automatic judgments stem from "deeper" structures, that is, from core and intermediate beliefs. Often in therapy, these structures are not directly accessible, but are inferred indirectly, from themes or patterns of automatic thoughts, emotional responses, or behaviors (J Beck, 2011). This means that the client judged the activating

event consciously (conscious judgment knowledge), but this conscious judgment is sustained, biased, by an implicit structure (unconscious structural knowledge).

CBT manuals seem to assume that when the therapist identifies the client's beliefs, and expresses it in a verbalizable form (e.g., *I am incapable* or *Social contexts are dangerous*), core or the intermediate beliefs become conscious for the client and ready for rational restructuring. However, this is only a representation of an *inference* about the implicit structure. More specifically, the client will have now a conscious, verbal, description of the structure, but, crucially, there is no reason to assume that the implicit structure itself has been "moved" into awareness. The implicit structure is still implicit. The description of the structure is explicit and restructurable. Restructuring this explicit representation is useful, because it constitutes the starting point for building the alternative, adaptive cognition, but it is important to keep in mind that the implicit structures are still there and might still produce effects.

(4) Implicit Knowledge Structures Are Learned More Easily in Childhood and Early Adolescence and Are Very Robust in Time

Janacsek, Fiser, and Nemeth (2012) investigated implicit learning with a version of SRTT, in participants with ages ranging from 4 to 85 years. They found that learning was the most efficient in childhood and in the early adolescence. This means that implicit structures are learned more easily in childhood and adolescence than during adulthood, a finding consistent with the CBT tenet that most schemas are acquired in this period (e.g., Beck, 1979). Also, this implies that acquiring new, adaptive structures, should be more effective in this period.

Furthermore, implicit structures are very resistant in time. Two years after they attended an AGL study, participants were asked by Allen and Reber (1980) to undergo a new test phase. Specifically, they had to classify strings as grammatical or non-grammatical relative to a grammar they had learned two years earlier. While there was a slight decrease in performance between the two time points, participants were able to classify the strings with above chance accuracy. Thus, the grammar learned two years earlier was mostly intact, and influenced participants' classifications. In a study that used the SRTT, participants retained an advantage in reaction time when the target-stimulus followed a sequence they had learned one year earlier (Kobor, Janacsek, Takacs, & Nemeth, 2017). Both studies show that these artificial structures, which lack any ecological relevance for participants, are stored and retained years after they are learned. It is, most likely,

safe to assume that implicit structures that have ecological and motivational relevance are stored at least as long as these. A more relatable example is that it is impossible, or at least very difficult, to forget how to ride a bicycle, a behavior that is executed mostly implicitly (it is close to impossible to describe how one maintains the equilibrium on a bicycle).

This robustness in time of implicit knowledge is consistent with the observation that the therapeutic process typically deals with structures that have been learned in childhood and adolescence, sometimes years or decades before they produce dysfunctional emotions.

(5) An Existing Implicit Knowledge Structure Is Not Erased by Learning a New One

Experimental evidence shows that once learned, implicit structures cannot be erased or restructured, at least not by the means available now. Dienes, Altmann, Kwan, and Goode (1995) presented participants, first, with letter strings that followed a specific artificial grammar. Subsequently, they presented participants with strings composed of the same letters, but that followed a different grammar. In the test phase, participants were presented with new strings that followed the first grammar, strings that followed the second one, and with strings that did not follow any of the two grammars. They had to respond whether the strings were grammatical or non-grammatical. However, on some trials, participants had to respond whether they are grammatical according to the first grammar, while on other trials, they had to respond according to the second grammar. Results showed that participants were able to make accurate classifications according to both grammars. Thus, after learning the first grammar, even though participants learned a new grammar that governed the same stimulus set (i.e., strings made of the same letters), the first grammar remained intact. Learning the second one did not interfere with the representation of the first one. This finding has been replicated several times in this two-grammar AGL design (e.g., Norman, Price, & Jones, 2011; Norman, Scott, Price, & Dienes, 2016) and also in SRTT, where Szegedi-Hallgato et al. (2017) found that two different non-conscious visuomotor sequences can coexist in participants' long-term memory.

While the representation of a learned structure seems to be impossible or difficult to alter, these studies also show that, in the same context and regarding the same stimulus set, one can hold multiple implicit structures. Even though learning a new structure (e.g., a positive/adaptive one) will not erase an old, initial, one (e.g., a maladaptive one), one's responses can be influenced not only by the initial structure, but also by the new one.

Implications and Conclusions

We could synthesize the characteristics presented above in three more general ones: First, implicit knowledge is acquired through experience, behavior, and exposure – thus, by direct contact with reality, by having some "skin in the game" (see Taleb, 2018). Second, they are very robust, rigid, and difficult to change. Third, their existence can be easily masked by the presence of conscious knowledge.

This last characteristic, seconded by the fact that we do not know precisely how to measure implicit structures, complicates their detection and study. Thus, if it is so difficult to know for sure whether, in addition to the conscious thoughts, there is also an implicit structure present, what should therapists do regarding implicit structures? Probably a good rule of thumb would be to help clients create or activate adaptive implicit structures whenever it is possible to do so. Even though there would be no dysfunctional implicit structures at work, creating adaptive ones would not harm, and might have preventive functions. But the reverse situation, where there are maladaptive structures present, but the therapist does not create adaptive alternatives, would be harmful. More specifically, therapists should insist on helping the clients engage in adaptive behaviors (e.g., Dowd, 2006; Dowd & Courchaine, 2002). CBTs already have a variety of strategies that involve the engagement in adaptive behaviors and that could be used in this sense. Importantly, in order to achieve this goal, therapists should help clients be aware that, initially, their emotions and behavioral tendencies will be influenced more by the dysfunctional structures, and should help them tolerate these influences (e.g., through rational disputation of low frustration tolerance or through acceptance techniques). Moreover, even though these behaviors are first practiced in less emotionally- and motivationally-charged contexts, practicing these behaviors should be moved towards more motivationally-relevant situations. As mentioned before, the human cognitive system is able to learn several alternative structures in very similar contexts (e.g., learning two artificial grammars that structure the same letters). Therefore, if adaptive behaviors are practiced only in contexts with low relevance, the cognitive system might activate these structures only in contexts with low relevance, and the contexts with high relevance would remained governed by the maladaptive structures. In conclusion, clients should be encouraged to use these adaptive behaviors in contexts that have motivational relevance for them.

3.7. Study 7: A Dual Process Approach to Current Issues from Cognitive Behavior Therapies: An Integrative Perspective[8]

The present chapter analyzes cognitive mechanisms advanced by CBT models as sustaining psychopathology, through the lens of developments from the dual process approach to judgment, decision, and cognitive appraisal. We show that augmenting CBT theories with data from dual process models reveals new mechanisms (e.g., automatic appraisals, heuristics) that are shaped, sometimes, by implicit learning and that can contribute to psychopathology and that can be targeted in therapy. Furthermore, we propose that different concepts from different CBT approaches – Cognitive Therapy (CT), Rational Emotive Behavior Therapy (REBT), and "third wave" therapies – can be translated into common underlying constructs from the dual process theories. This translation facilitates a deeper understanding of the commonalities shared by these different CBT approaches and enables concrete steps for their integration. Finally, we propose new testable hypotheses and research directions derived from these analyses, which can enhance our knowledge of cognitive processing in CBTs.

A significant number of authors have proposed that, in the process of refining CBT's theory and efficiency, a natural source of information can and should be the field of cognitive psychology, which studies the basic mechanisms involved in information processing (David, Miclea, & Opre, 2004; Ingram & Siegle, 2010).

The *present chapter's main objective is to augment CBT theories with knowledge derived from the dual-process approach to learning, judgment, decision-making, and emotional appraisal*. Briefly, this approach posits that adaptive and maladaptive cognition depend on the interplay between two types of processing: type 1 processing, which is relatively automatic, effortless, and fast, and type

[8] Published as: Jurchiș, R., & Opre, A. (2018). A Dual Process Approach to Current Issues from Cognitive Behavior Therapies: An Integrative Perspective. International Journal of Cognitive Therapy, 11(4), 374-409. https://doi.org/10.1007/s41811-018-0023-z

2 processing, which is controlled, effortful, and slow (e.g., Evans, 2008; Kahneman, 2011; Tversky & Kahneman, 1983).

Precisely, we will approach (a) the role of automatic appraisal processes and their implications for CBTs; (b) the relationship between different constructs from CT, REBT, and "third wave" therapies that can be translated into common, more basic, concepts from the dual process approach; (c) the dual process mechanisms through which cognitive and behavioral strategies from CBTs elicit therapeutic change. As the present work attempts to be, at least in part, a call to research, we will also propose specific research hypotheses and broader research directions on each of these subjects. We expect that the steps taken in this paper will bring (a) a deeper clarification of mechanisms involved in psychopathology and therapeutic change and (b) a better understanding of commonalities and compatibilities between different CBT approaches, which will enable new, concrete, steps for their integration at a theoretical and practical level.

Type 1 Processes

Type 1 processes, also called intuitive, implicit, or automatic, refer to a family of cognitive mechanisms that enable *rapid adaptation to a complex, dynamic, environment*. In this sense, these mechanisms sustain *fast disambiguation of, and fast responding to, environmental stimuli, with relatively little effort* (e.g., Kahneman, 2011; Kahneman & Frederick, 2005). Type 1 processes rely on fast activation of implicit and explicit memory representations that are relevant for the encountered stimuli. Also, these processes provide appraisals of stimuli on emotionally relevant dimensions (e.g., Moors, 2010; Moors, Ellsworth, Scherer, & Frijda, 2013). Once a stimulus has been evaluated, *type 1 processes remain over-focused on the initial evaluation and ignore relevant information that is not immediately available.* Consequently, the results of this automatic evaluation can achieve an *overly-intense influence on subsequent processes or can produce effects in inappropriate situations* (e.g., Kahneman, 2011; Kahneman, Krueger, Schkade, Schwarz, and Stone, 2006; Schkade & Kahneman, 1998). The influences of type 1 processes can be overcome by activating type 2 processes, which are able to go beyond the immediately available information and to provide more nuanced evaluations.

Type 2 Processes

Type 2 processes, also called reflective, explicit, analytic, or rational (e.g., Evans, 2008), process information in a more conscious, effortful, and controlled manner. They are dependent on working memory and executive functions, and deal with data in a sequential, step-by-step manner (e.g., Darlow & Sloman, 2010). Relatedly, they rely on rules (e.g., of syllogistic reasoning), and draw conclusions based on logical and empirical proofs (Epstein, 1994; Evans, 2003, 2008). Hence, *they can correct sometimes the outputs provided by type 1 processing.* However, deployment of type 2 processes does not necessarily mean that they will correct biases generated by type 1 processes. If the person does not have access to corrective information, type 2 processes can sometimes strengthen or amplify the distorted outputs of type 1 processes (e.g., Beevers, 2005). Moreover, they are significantly slower and highly dependent on resource availability (attention, working memory), thus they are not always available for providing alternatives to type 1 processing.

Cognitive Therapy (CT) and Rational Emotive Behavior Therapy (REBT)

The general assumption shared by both CT and REBT is that dysfunctional emotions and psychopathology emerges when negative life events are appraised by distorted, irrational, cognitions. Therefore, the main purpose of the therapeutic process is to replace these cognitions with more functional, rational, ones.

The two main general types of therapeutic procedures employed by CT and REBT are verbal restructuring procedures and behavioral strategies. Verbal restructuring procedures consist in gathering evidence against dysfunctional cognitions and in formulating alternative, functional, ones. The therapy also relies heavily on behavioral strategies, which are considered for some disorders (e.g., anxiety) as being more efficient than purely cognitive, talk-based, ones (e.g., Longmore & Worrell, 2007). The main strategies/techniques in this category are behavioral experiments, behavioral activation, and exposure therapy. Even though the infrastructure of these strategies is mainly behavioral (i.e., they imply engaging the patients in specific behaviors), the mechanisms through which they lead to reduction of dysfunction are assumed in CT to be cognitive (e.g., Beck, 1976).

One important disagreement between CT and REBT is that the latter operates a distinction between two types of cognitions that can contribute to psychopathology: inferences and evaluations. Specifically, Ellis (2005) sustained that most distorted thoughts proposed by Beck (1976) are inferences, which cannot produce dysfunctional emotions, unless they are further evaluated through irrational evaluations. For example, after a fight with his wife, one client makes the negative inference "She will leave me". According to the REBT model, this inference could be further evaluated rationally or irrationally. If it is evaluated irrationally, for example, "She should not leave me! I couldn't stand the situation", the consequence will be dysfunctional (e.g., anxiety, depression). If the same negative inference is evaluated rationally, for example, "I strongly want her to stay with me, but if she leaves me, I could endure the situation", the consequences will be negative, but functional (e.g., concern, sadness).

The specific irrational evaluations involved in psychological disturbance are evaluations in terms of *must, awfulizing, low frustration tolerance, self-* and *other-downing* (e.g., Dryden, Neenan, & Yankura, 2001; Ellis et al., 2010). Evaluations in terms of *must* (or *should, ought, have to,* etc.), called *demandingness,* consist in transforming one's desires into absolutist, rigid, demands regarding certain aspects of reality (e.g., "I absolutely must succeed in this task"). *Awfulizing* occurs when one considers a negative event (actual or inferred) as the worst thing that could happen (e.g., "It would be awful, the worst thing ever, to fail this exam"). *Low frustration tolerance* consists in evaluating oneself as unable to tolerate an unpleasant aspect of internal or external reality (e.g., "I cannot tolerate feeling anxiety"*;* or "I could not stand failing this exam"). *Self-* and *other-downing* implies attributing to one's own or to other's person a global rating, a general label [e.g., "I am (or you are) a stupid person"]. REBT posits that whenever there is a dysfunctional emotion, there is also present an irrational evaluation, even though often it is not expressed by patients. In other words, the evaluation is there, but it remains "implicit".

However, it is very difficult to demonstrate convincingly that, if one has a dysfunctional emotion, there has to be an implicit irrational evaluation that causes it, even though it is not expressed and it is not consciously accessible. Moreover, REBT does not provide detailed cognitive mechanisms that could explain how these evaluations emerge. As a result, these proposals regarding evaluations have not penetrated other mainstream CBT approaches, such as CT.

Irrational Evaluations from a Dual-Process Perspective

According to REBT theory, *musts* result by transforming one's desires into absolutistic demands (e.g., Dryden, 2014). We will argue that type 1 processing can explain how this transformation occurs. First, type 1 processes evaluate automatically the motivational relevance and congruence of an event (e.g., "I want/don't want event X to happen"). Further, as type 1 processing takes into account only the immediately available information, in this case the personal motivation regarding the activating event, other aspects that matter in the situation will be ignored. As a result, if *what I want is all there is,* it would be easy to conclude that the wanted event *must* happen – there would be nothing to interfere with the wish, nothing to prevent it from transposing to reality.

These mechanisms are not limited to irrational *musts*, but are also able to account for the rest of irrational evaluations. In the case of awfulizing, first, the person appraises the situation as motivationally incongruent, and the situation is perceived as negative, bad. If the person over-focuses on the badness of the situation, the situation can be easily evaluated as awful, that is, as the worst thing that could happen; for type 1 processes, is the only situation that exists in that moment, and it is entirely bad, thus it is the worst. A similar course can lead to low frustration tolerance: concentrating totally on the negativity and potential harmfulness of the activating event does not allow to take simultaneously into account the personal resources to face the event. Therefore, the person concludes that he/she cannot stand it. Finally, in self- and other-downing, over-focusing on an undesirable behavior (or thought, emotion, etc.), leads to the conclusion that the behavior is the single relevant feature in judging the person. Hence, if the single feature of a person is bad, and there is no other relevant aspect regarding the person, the person is bad.

In sum, irrational evaluations are sustained by automatic appraisals, followed by over-focusing on the results of these appraisals (e.g., on one's own personal motivation, on the negativity of motivationally-incongruent events), while failing to consider other relevant information (e.g., personal resources to face the event).

Similar to irrational evaluations from REBT, distorted automatic thoughts proposed by CT result from type 1 processes. Specifically, automatic thoughts arise spontaneously, effortlessly, from processes that are not consciously accessible, in response to different environmental stimuli (e.g., J Beck, 2011), and are automatically biased by cognitive structures, especially by core and intermediate beliefs. All these characteristics of automatic thoughts are also marks of type 1

processing. Furthermore, the cognitive distortions found in automatic thoughts are characterized by over-focusing. For example, *overgeneralization* consists in focusing on one element or situation in a category and making conclusions regarding the entire category, by ignoring elements that do not fit the generalization. *Personalization* consists in focusing on one's – real or imagined – contribution to a negative outcome, ignoring other factors' contribution, and assuming the entire responsibility for the outcome (e.g., *It's my fault that my parents broke-up*). If inferential and evaluative automatic thoughts share these similarities, and correlate so strongly (e.g., r = .85, in Visla et al., 2015), what differentiates them? Briefly, the main difference resides in their domain of reference: while evaluative thoughts reflect mainly appraisals of activating events on motivational dimensions, inferential thoughts reflect disambiguation of activating events on other dimensions. For example, they refer to the causes of activating events (*personalization*), to their consequences (*fortune telling* and *jumping to conclusions*), to other persons' thoughts or intentions (*mind reading*), etc.

Predictions. Next, we will offer several testable predictions derived from our analysis, which can contribute to a better understanding of the dual process perspective's utility for CT and REBT.

First, we hypothesize that, because distorted inferences, irrational evaluation, and biases in judgment/decision making are favored by the same mechanism (i.e., over-focusing), all these constructs should covariate, especially in persons with psychological disorders. Second, the hypothesis that distorted inferences and irrational evaluations are caused by type 1 processes can be tested by experimentally manipulating conditions that favor type 1 processes (e.g., imposing a response deadline, engaging in multiple tasks). Thus, if this hypothesis is valid, conditions that favor type 1 processing should increase the frequency and/or credibility of distorted inferences and irrational evaluations. There are preliminary results sustaining this prediction: Goldin et al. (2013) show that under a response deadline, patients suffering from OCD infer higher probabilities for harm-related events, compared to a no-deadline condition (see also Freeman, Lister, & Evans, 2014).

Relatedly, a third hypothesis would be that techniques used in judgment/decision for defocusing and reduction of impact bias, should also promote more rational cognitions. For instance, research in this field shows that asking people to think concretely about their daily activities, reduces the tendency to predict exaggerated influence of specific events on their emotions (Pedersen,

Kristensson, & Friman, 2012; Wilson et al., 2000). This technique promotes defocusing because it helps people to see that, besides the particular event they are focusing on, there are many other factors that influence their emotions. If, as we have proposed, over-focusing is at the core of irrational evaluations, concrete thinking should weaken them. More precisely, concrete thinking should enable a complex and nuanced assessment of the relevance and valence of activating events, reducing, as a result, the tendency to remain over-focused on, for example, their "badness".

CT and REBT promote the necessity of increasing the availability of rational cognitions, as they strongly emphasize the necessity of repeating the rational cognitions. However, we propose, taking into account existing results on *availability heuristic* (e.g., Schwarz et al., 1991), that the believability of an adaptive cognition could be enhanced by increasing the availability of arguments, memories, etc. that sustain it.

Cognitive Defusion and Cognitive Restructuring

A central principle of the third-wave psychotherapies (e.g., Acceptance and Commitment Therapy; ACT) is that psychological disturbance depends, to a large extent, on the way people relate to their thoughts. Specifically, psychopathology is favored when people confound the symbols used by the human mind to represent aspects of reality, with their referents; that is, thoughts about realities are taken as the realities themselves. In consequence, the thoughts will achieve an exaggerated influence on other psychological processes and on behavior (e.g., Bernstein et al., 2015; Hayes, Levin, Plumb-Vilardaga, Villatte, & Pistorello, 2013). For example, a client's thought "My spouse will leave me", is experienced as a direct and accurate reflection of the reality. The client doesn't consider "I have the thought that *my spouse will leave me*, but I know it is just a thought that may or may not be accurate". As a consequence, the client's cognitive, emotional, and behavioral reactions, will be close to those that would appear if the thought would be real. In ACT, this phenomenon is called *cognitive fusion* (e.g., Gillanders et al., 2014; Hayes et al., 1999).

The Defusion-Restructuring Relationship from a Dual Process Perspective

From the dual process perspective, cognitive fusion seems to be a manifestation of over-focusing. Specifically, as the ACT model states, during fusion the thought becomes, for the person having it, the entire reality, and achieves an exaggerated influence on behavioral responses. This is

precisely the landmark of type 1 processing: considering, by default, that *what you see is all there is*, in order to be able to give fast and/or sustained behavioral responses (e.g., Kahneman, 2011). A direct implication of this proposal is that cognitive fusion is sustained, at least partially, by the same phenomenon that sustains irrational inferences and evaluations, that is, by over-focusing. Correlational data provide preliminary evidence for this possibility: although they are different constructs with different theoretical underpinnings, they are strongly associated [e.g., r = .61 for fusion and automatic thoughts (inferential and evaluative, taken together), Gillanders et al., 2014].

If cognitive fusion and dysfunctional cognitions are based on similar mechanisms, then their counterparts – cognitive restructuring and cognitive defusion – target a common mechanism: That, is, they both try to reduce overfocusing on the available information. However, we do not consider that cognitive defusion and restructuring are identical or equivalent. Cognitive restructuring involves, in addition to defusing from the dysfunctional cognition, a rational assessment of its accuracy and generation of incompatible information (counterarguments and the alternative, adaptive, cognition). Thus, cognitive restructuring consists in defusion plus generation of alternative information. This generation of alternative information leads to the development of adaptive cognitive structures (e.g., beliefs), which is considered essential for long term improvement (J Beck, 2011). [9]

If cognitive restructuring involves defusion, *training patients explicitly to defuse from their thoughts, might make them more efficient in restructuring these thoughts* (see Troy et al., 2013, for preliminary evidence). Also, after restructuring a dysfunctional cognition, and generating an adaptive alternative cognition, when the patient encounters again the dysfunctional cognition, it may not be necessary to question it again and to generate arguments against it. It might be enough to defuse from that cognition and to retrieve the alternative one. Because cognitive restructuring typically consists in several sequential steps, it likely needs substantial amounts of cognitive resources (e.g., working memory capacity, inhibition). On the other hand, cognitive defusion can be achieved, for example, simply by stating "I have the thought that…" before stating the dysfunctional thought. Therefore, such *defusion strategies likely need fewer cognitive resources*

[9] Note that restructuring is not the only way to develop adaptive cognitive structures. As we will show, engaging in adaptive behaviors can sometimes change cognitions better than cognitive restructuring does (see **Behavioral strategies and cognitive change in CBTs**). Consequently, third-wave therapies that do not use cognitive restructuring can still generate cognitive change by engaging patients in value-congruent behaviors.

compared to restructuring. In situations where one's cognitive resources are unavailable, one could reduce the over-focusing on a negative thought, situation, or experience, and the consequences of this over-focusing, by defusion, without restructuring that thought.

In sum, according to the dual process account, cognitive fusion and dysfunctional cognitions are, partially, sustained by the focalism of type 1 processes. Moreover, we have identified compatibilities and differences between cognitive restructuring and defusion, which can enable practical steps for combining them in clinical practice.

Predictions. If cognitive fusion is the product of type 1 processing, then it should be more intense, should influence behavior more pronouncedly, under conditions that favor type 1 processing. Also, it should correlate with judgment biases, irrational evaluations, and distorted inferences, as we hypothesize that they are all sustained by over-focusing.

If cognitive fusion/defusion superpose, at least partially, with over-focusing/defocusing, then defusion strategies should reduce the consequences associated with over-focusing, such as impact bias, focusing illusion, or the influence of availability/fluency, or of other heuristics.

If cognitive defusion and cognitive restructuring both reduce over-focusing, then combining the two, compared to using them in isolation, should better counteract negative cognitions and their consequences, in terms of magnitude, speed, and/or stability of change (see Troy et al,. 2013).

Behavioral Strategies and Cognitive Change in CBTs

Empirical evidence suggests that behavioral strategies are among the best or, in some circumstances, actually the best methods for reducing symptomatology and changing cognitions (e.g., Dimidjian et al., 2006; McManus, Van Doorn, & Yiend, 2012). Even though research has been able to identify some cognitive mechanisms that are changed by behavioral strategies (e.g., Jacobson et al., 1996; Dimidijan et al., 2006), a thorough account of why behavioral strategies change cognitions better than cognitive strategies do, is still lacking. We propose that these results are explainable taking into account the interplay of type 1 and type 2 processing. More specifically, we argue that, while verbal restructuring strategies use mostly type 2 processing, behavioral strategies influence both type 1 and type 2 processing.

A central tenet of the dual process approach is that type 1 processing is influenced by the cognitive structures that are available in a given context (e.g., Kahneman, 2011). Based on some well researched properties of human learning and memory, we will explain that behavioral strategies are more efficient in creating cognitive structures that can sustain adaptive functioning or in increasing their level of activation if these structures already exist.

For example, engaging in a behavioral experiment tends to elicit more intense emotion than verbal restructuring (Bennett-Levy, 2003). *As emotional memories tend to have a higher level of availability than neutral ones (e.g., Petrican et al., 2008), the emotionally-charged data encoded in a behavioral experiment will be more accessible than, for example, those encoded in a verbal restructuring situation.* As a result, the outcome of the behavioral experiment will constitute an argument which will be easier to retrieve and easier to believe (cf. availability heuristic), than a logical argument with little or no emotional component.

Research *on mood-congruent memory* reveals the automatic dependence of memory activation on the emotional context. Specifically, the probability of retrieving an information is more likely if the emotional context at retrieval matches the emotional context from the moment of encoding (Barry, Naus, & Rehm, 2004; Bower, 1981). Very likely, this phenomenon explains, to a significant extent, the efficiency of behavioral activation. There is ample evidence that during a depressive episode, the depressed mood and negative ideation are associated and this association is stored in memory. Consequently, the presence of a negative mood maintains activated the association, and, implicitly, the negative ideation (Scher, Ingram, & Segal, 2005). When the patient is engaged in activities that promote positive emotions, one source of activation of negative ideation (i.e., the negative mood) will be interrupted, and the accessibility of positively-toned memories and ideation will increase. Hence, *type 1 processes might now be influenced more by cognitive structures with positive content than by those with negative content. Also, the successful execution of this type of activities can be used as argument in rational, type 2 reasoning about the patient's personal competence and quality of life.*

In conclusion, we consider that the efficiency of behavioral strategies is sustained by their capacity to influence directly heuristics (e.g., availability) and other type 1 processes, being thus able to produce changes at a more implicit, emotional, intuitive level, rather than only at a rational, intellectual one (e.g., Samoilov & Goldfried, 2000).

Predictions. Starting from the possibility that behavioral strategies influence type 1 processes more directly than verbal restructuring strategies, we predict that changes related to memory processes – implicit and explicit – or related to automatic associations (e.g., Frank, De Raedt, & De Houwer, 2007) should be achieved faster and/or should be stronger when using behavioral strategies. More broadly, *after successfully engaging in behavioral experiments, behavioral activation, or exposure, it should be easier to retrieve functional cognitive contents (e.g., beliefs, memories) related to the specific situation one engages in, than after a cognitive restructuring session.*

Conclusions

The present work makes several contributions for a deeper understanding of CBT theory and practice, using a perspective strongly grounded in cognitive science research. More specifically, the paper has tried to reconceptualize through the dual process perspective some of the mechanisms that contribute to dysfunctional psychological phenomena and some mechanisms of change targeted by CBTs. Also, it indicates starting points for overcoming some limitations of current therapeutic strategies. Significantly, this perspective is able to provide insights into the commonalities shared by major CBT schools and to connect theories and strategies that seem incompatible. Thus, we appreciate that a dual process approach is able to promote an in-depth analysis of clinical cognitive phenomena while keeping a significant integrative power.

4. IMPLICATIONS AND CONCLUSIONS

The thesis presented seven studies that targeted three aspects, dimensions, of implicit learning of emotional structures: (a) fundamental research studies that would prove the existence of the phenomenon, (b) studies employing stimuli that are closer to real-life circumstances, in order to examine the ecological relevance of the phenomenon, and (c) studies that would facilitate the transfer of knowledge regarding implicit structures and processes, from cognitive psychology, towards the theory and practice of CBTs.

The first three studies investigated the existence of implicit learning of emotional structures, and found evidence that non-conscious structures can acquire emotional valence. Importantly, they can influence affective evaluations of stimuli that follow the structures.

Studies 4 and 5 dealt with more ecologically-relevant dimensions of implicit emotional learning. Specifically, study 4 found that regularities followed by facial emotional expressions can be learned implicitly. It also found that this process is not undermined by depression, contradicting the studies that have shown implicit learning impairments in this disorder. Also, it provided evidence that the explicit acquisition component might even be enhanced in depression. Study 5 went even further in addressing the ecological relevance of implicit emotional learning, and proposed a task in which participants learned the structure of scenarios that describe situations that may occur in real life.

Studies 6 and 7 explored the underexploited knowledge base generated by fundamental research on human implicit cognition, and showed its implications for CBT theories and practice. Study 6 synthesized several properties of implicit knowledge structures, derived from empirical research, and described their relevance for the theory and practice of CBT. Finally, study 7 decomposed CBT processes into deeper implicit, type 1, processes, and generated directions for integration of various CBT approaches and for their development in new directions.

4.1. Empirical and Theoretical Contributions
(What do we know now that we did not know before?)

First, *we have learned that acquisition of complex and abstract structures with emotional valence is possible even in the absence of awareness*. The first three studies *provide evidence that might*

fill all the gaps from previous research: Different from the existing studies on the topic, especially in study 3, (a) the structures were complex, (b) the evidence for unconscious learning was strong, (c) the influences of the emotionally-loaded structures generalized to new stimuli that were not directly involved in the acquisition/conditioning phase, and (d) the learned structures influenced participants' affective evaluations of these new stimuli.

Second, *we have found that regularities followed by emotional facial expressions can be learned implicitly, which brings support for the tenet that implicit learning is involved in the acquisition of socio-emotional regularities* (e.g., Lieberman, 2007). Study 4 from the present thesis is, to my knowledge, the first study to show implicit learning for social stimuli with emotional valence.

Third, *we have learned that this implicit acquisition of socio-emotional regularities is not impaired by depression and that the explicit acquisition component might be enhanced in depressive individuals.* Mapping the circumstances in which depression could have adaptive roles would be, beyond doubt, an important step in understanding its etiopathogenesis and in determining the most appropriate ways of treating it. As the empirical investigations on the topic are relatively sparse, the task we used in study 4 might be useful for exploring further the mechanisms that associated with adaptive functions of depression.

Fourth, *we have found preliminary evidence that, even in the absence of awareness, participants are able to extract regularities followed by scenarios that can occur in real life.* This could represent an important step in increasing our knowledge of the ecological relevance of implicit learning.

Fifth, from reviewing the properties of implicitly learned knowledge, it resulted that, *in the context of therapy, implicit structures are difficult to detect, to change, are robust in time and to interference with other structures, are learned by direct exposure to stimuli that follow the regularity.* Specifically, our analysis revealed that (a) implicit structures are very difficult or impossible to remove and (b) the only method for counteracting them would be through creating (or activating) an alternative structure, definitely not through rational counterarguments. This is not to say that CBT never changes implicit structures. As mentioned, behavioral strategies used in CBT can trigger the process of creating/activating these implicit structures. Still, behavioral strategies (e.g., behavioral experiments) are sometimes used just as another tool for cognitive change, which can replace or assist verbal restructuring strategies (only behavioral strategies incur

more costs, can trigger stronger negative emotions, thus are less likely to be used). From our perspective, the direct exposure to adaptive patterns in the environment, which can be achieved through behavioral strategies, should lead to a *qualitatively different* type of knowledge, compared to engaging in rational disputation.[10]

Sixth, this analysis of the roles that implicit knowledge can play in real life, and of how the effects of implicit structures can be counteracted, invites to a careful assessment of the very definition and conceptualization of implicit learning and of the resulting knowledge. Specifically, all knowledge derived in implicit learning tasks is derived by *direct exposure to stimuli that follow the regularities and/or by responding to stimuli that follow regularities*; not by reasoning about regularities, or by merely having regularities described to participants. Hence, when we propose that in CBT clients should be assisted in creating adaptive implicit structures, we do *not* mean that therapists should create complex learning situations in which clients would acquire knowledge without being aware of it; we mean that the client should be exposed repeatedly to real life contexts governed by adaptive regularities and should be taught to act, to behave, following adaptive regularities.

Seventh, *we have proposed that most cognitive mechanisms that are considered as maladaptive or dysfunctional by CBT models could have important, yet neglected, adaptive functions*. As in the case of biases revealed by judgment and decision-making, they could be sustained by properties of the cognitive system that are essential for adaptive functioning. That is, even though they violate sometimes the rules of logical inference, they might still produce adaptive responses. For example, over-focusing on the motivational relevance of an event, which can sustain *demandingness* from REBT, can also sustain persistence in difficult tasks (Morewedge & Buechel, 2013); *self-downing* can sustain pro-social behaviors (Bryan et al., 2013), and so does *cognitive fusion* (Carnes & Winer, 2017). Accordingly, focusing only on the fact that they co-occur with psychopathology, and ignoring that they also sustain adaptive functioning, impede our understanding of the roles they can play in psychopathology and of the difficulties in changing them.

[10] To avoid possible misunderstandings: The emphasis we put on direct exposure, behavior, does not imply that conscious, rational, verbal strategies are unnecessary. In specific situations, a reappraisal of the situation might be enough for generating an adaptive emotional or behavioral response. Also, engaging in behavioral strategies might require, first, the reappraisal of one's thoughts about the difficulty of the situation, about one's own capacity to endure it etc.

Eighth, *we have shown how different CBT constructs from different CBT approaches (or "schools") can be translated into common underlying constructs from dual process theories*. Accordingly, we proposed *concrete steps for integrating these CBT approaches both at a theoretical and at a practical level*. For instance, we have shown that over-focusing on the results of automatic appraisals is a process that could contribute to distorted automatic thoughts from CT, irrational evaluations from REBT, and cognitive fusion from ACT. From these theoretical claims *we advanced specific, testable, hypotheses that could verify the plausibility of our conjectures* (e.g., automatic thoughts, irrational evaluations, and fusion should covary especially under conditions that favor type 1 processing). Furthermore, *we illustrated several practical applications of the advanced theoretical claims* (e.g., therapy should create alternative structures by helping clients to act adaptively not only in training, artificial, contexts, but in contexts with motivational relevance, because structures with motivational relevance become more available, hence they are able to influence type 1 processes).

Ninth, *from integrating CBT models on the dual-process platform, and from infusing phenomena from the dual-process approach into CBT, we obtained two general cognitive factors or processes that contribute to therapy.* First, effects of type 1 processing (e.g., cognitive fusion, automatic inferential and evaluative thoughts) can be counteracted by effortful type 2 processing (e.g., defusion, reappraisal). However, considering that type 2 processes are slower and resource-consuming, they cannot, by themselves, overcome the effects of strongly-activated cognitive structures, which influence type 1 processing. Therefore, the second factor consists in creating/activating alternative structures that are more adaptive in the context, so that these structures would influence type 1 processes.

4.2. Methodological contributions

(What can we study now that we could not study before?)

All experimental paradigms used in the empirical studies had various degrees of novelty, being specifically built for the objectives of this thesis, although they were based, more or less, on classical tasks in implicit learning and in other areas (evaluative conditioning, cognitive bias modification). Therefore, all these tasks have the potential to open new areas for investigations. For example, the mixt AGL–EC task developed in Study 3 could be used for further investigations on unconscious emotional learning or on unconscious evaluative conditioning. The AGL task in

which artificial grammars structured strings of emotional facial expressions might be used for investigating socio-emotional learning in various groups, disorders etc. The scenario-based paradigm from study 5 can be employed for investigating acquisition of various types of regularities followed by real-life situations.

4.3. Future directions

An important topic for further exploration regards *the properties of these emotionally-loaded structures*. In study 6 we described several properties of implicit structures; as we acknowledged, the structures have been discovered mostly in paradigms that used neutral regularities. Consequently, it would be important to determine to what extent these properties apply to emotionally-loaded ones.

Second, *the malleability or changeability of these structures is a topic that warrants further exploration*. As described in chapter 6, once an implicit structure has been learned, it is very robust in time, and to interference with other structures. For example, as several studies have shown, participants are able to learn two grammars that structure the same stimulus set; thus the second structure does not replace the first one, but the two structures coexist (e.g., Dienes et al., 1995; Norman et al., 2012, 2016). In study 3 we saw that this is also the case with emotionally-loaded structures: participants associated an artificial grammar with a negative emotion and another grammar that structured the same letters, with a positive emotion. It would be relevant to find out what happens when a grammar that has been conditioned, first, negatively, is then associated with positive stimuli. Would the negative valence be replaced by the positive valence? Or would there be two competing representations: one in which the grammar has a positive valence, and one in which the same grammar has a negative valence?

A fourth research avenue derived from the present thesis could investigate whether there are *inter-individual variables that influence the acquisition of implicit emotional structures*. For example, persons that are high in neuroticism could be more prone to learn structures with negative valence, as high levels of neuroticism are associated with allocation of cognitive resources to negative stimuli (e.g., Ormel et al., 2013) and with increased levels of fear conditioning (e.g., Hur, Ioardan, Berenbaum, & Dolcos, 2016). Moreover, it would be possible that persons with higher levels of neuroticism would manifest similar amounts of learning, only they would learn these structures more explicitly, as it was the case with depressed participants in study 4.

A fifth research direction would be *to produce paradigms that are even more ecologically-relevant than those we used in the present thesis*. For example, *virtual reality (VR) environments* could be created, in which some of the stimuli are predictable according to some regularities; or participants could interact, in VR, with avatars whose behavioral or emotional responses can be anticipated according to some rules (see Sense & van Rijn, 2018, for a VR-based, motor sequence learning paradigm).

5. REFERENCES

Air, T., Weightman, M. J., & Baune, B. T. (2015). Symptom severity of depressive symptoms impacts on social cognition performance in current but not remitted major depressive disorder. *Front. Psychol.*, *6*. doi: https://doi.org/10.3389/fpsyg.2015.01118

Alloy, L. B., & Abramson, L. Y. (1988). Depressive realism: Four theoretical perspectives. In L. B. Alloy (Ed.), *Cognitive processes in depression* (pp. 223–265). New York, NY: Guilford Press

Allen, R., & Reber, A. S. (1980). Very long term memory for tacit knowledge. *Cognition*, *8*(2), 175-185.

Andrews, P. W., & Thomson, A. J. (2009). The Bright Side of Being Blue: Depression as an Adaptation for Analyzing Complex Problems. *Psychol. Rev.*, *116*(3), 620–654. doi: https://doi.org/10.1037/a0016242

Atas, A., Faivre, N., Timmermans, B., Cleeremans, A., & Kouider, S. (2014). Nonconscious learning from crowded sequences. *Psychological science*, *25*(1), 113-119. https://doi.org/10.1177/0956797613499591

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The Depressed Brain: An Evolutionary Systems Theory. *Trends Cogn. Sci.*, *21*(3), 182–194. doi: https://doi.org/10.1016/j.tics.2017.01.005

Barry, E. S., Naus, M. J., & Rehm, L. P. (2004). Depression and implicit memory: Understanding mood congruent memory bias. *Cognitive Therapy and Research*, *28*(3), 387-414. doi:10.1023/B:COTR.0000031808.00502.2e

Beck, A.T. (1976). *Cognitive therapy of the emotional disorders*. New York, NY: New American Library.

Beck, A. T., & Bredemeier, K. (2016). A Unified Model of Depression: Integrating Clinical, Cognitive, Biological, and Evolutionary Perspectives. *Clin. Psychol. Sci.*, *4*(4), 596–619. doi: https://doi.org/10.1177/2167702616628523

Beck, A. T., Rush, A. J., Shaw, B. E., & Emery, G. (1979). *Cognitive therapy of depression*. New York, NY: Guilford Press.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond* (2nd ed). New York: Guilford Press.

Beevers, C. G. (2005). Cognitive vulnerability to depression: A dual process model. *Clinical Psychology Review*, *25*(7), 975-1002. doi: 10.1016/j.cpr.2005.03.003

Bennett-Levy, J. (2003). Mechanisms of change in cognitive therapy: The case of automatic thought records and behavioural experiments. *Behavioural and Cognitive Psychotherapy, 31*, 261-277. doi: 10.1017/S1352465803003035

Bernstein, A., Hadash, Y., Lichtash, Y., Tanay, G., Shepherd, K., & Fresco, D. M. (2015). Decentering and related constructs: a critical review and metacognitive processes model. *Perspectives on Psychological Science, 10*(5), 599-617. Doi: https://doi.org/10.1177/1745691615594577

Bora, E., & Berk, M. (2016). Theory of mind in major depressive disorder: A meta-analysis. *J. Affect. Disord.*, *191*, 49–55. doi: https://doi.org/10.1016/j.jad.2015.11.023

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129-148. doi: http://dx.doi.org/10.1037/0003-066X.36.2.129

Brewin, C. R. (1996). Theoretical foundations of cognitive-behavior therapy for anxiety and depression. *Annual Review of Psychology, 47*(1), 33-57. https://doi.org/10.1146/annurev.psych.47.1.33

Bryan, C. J., Adams, G. S., & Monin, B. (2013). When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General*, *142*(4), 1001-1005. doi: http://dx.doi.org/10.1037/a0030655

Carnes, N. C., & Winer, J. P. (2017). Do unto others: How cognitive fusion shapes the transmission of moral behavior. *Journal of Experimental Psychology: General*, *146*(4), 472-484. doi: 10.1037/xge0000294

Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 382-392. doi: 10.1002/wcs.34

David, D., & Lynn, S. J. (2010). A summary and a new research agenda for rational-emotive and cognitive-behavior therapy. In David, D., Lynn, S. J., & Ellis, A. (Eds.). *Rational and irrational beliefs: Research, theory, and clinical practice* (pp. 339-347). New York, NY: Oxford University Press.

David, D., Miclea, M., & Opre, A. (2004). The information-processing approach to the human mind: Basics and beyond. *Journal of Clinical Psychology*, *60*(4), 353-368. doi: 10.1002/jclp.10250

Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. New York, NY: Palgrave Macmillan.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on?. *Perspect. Psychol. Sci.*, *6*(3), 274-290. doi: 10.1177/1745691611406920

Dienes, Z., & Altmann, G. (1997). Transfer of implicit knowledge across domains? How implicit and how abstract? In D. Berry (Ed.). *How implicit is implicit learning?* (pp. 107–123). Oxford: Oxford University Press.

Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1322-1338. http://dx.doi.org/10.1037/0278-7393.21.5.1322

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychon. Bull. Rev.*, *25*(1), 207-218. doi: https://doi.org/10.3758/s13423-017-1266-z

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychol. Res.*, *69*(5–6), 338–351. doi: https://doi.org/10.1007/s00426-004-0208-3

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., ... & Atkins, D. C. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology, 74*(4), 658-670. doi: http://dx.doi.org/10.1037/0022-006X.74.4.658

Dowd, E. T. (2006). What changes in cognitive therapy? The role of tacit knowledge structures. *Journal of Cognitive and Behavioral Psychotherapies, 2*, 141-148.

Dowd, E. T., & Courchaine, K. E. (2002). Implicit learning, tacit knowledge, and implications for stasis and change in cognitive psychotherapy. In R. L. Leahy, & E. T. Dowd (Eds.). *Clinical advances in cognitive psychotherapy: Theory and application* (pp. 325-344). New York, NY: Springer.

Dryden, W. (2014). *Rational emotive behaviour therapy: Distinctive features* (2nd ed). New York, NY: Routledge.

Dryden, W., Neenan, M., & Yankura, J. (2001). *Counselling Individuals: A Rational Emotive Behavioural Handbook* (3rd ed). London, UK: Whurr.

Durisko, Z., Mulsant, B. H., & Andrews, P. W. (2015). An adaptationist perspective on the etiology of depression. *J Affect Disord, 172*, 315–323. doi: https://doi.org/10.1016/j.jad.2014.09.032

Ellis, A. (2005). Discussion of Christine A. Padesky and Aaron T. Beck," Science and Philosophy: Comparison of Cognitive Therapy and Rational Emotive Behavior Therapy". *Journal of Cognitive Psychotherapy*, *19*(2), 181-185. doi: 10.1891/jcop.19.2.181.66789

Ellis, A., David, D., & Lynn, S. J. (2010). Rational and irrational beliefs: A historical and conceptual perspective. In D. David, S. J. Lynn, & A. Ellis (Eds.), *Rational and irrational beliefs: Research, theory, and clinical practice* (pp. 3-22). New York, NY: Oxford University Press.

Esser, S., & Haider, H. (2017). The emergence of explicit knowledge in a serial reaction time task: The role of experienced fluency and strength of representation. *Frontiers in Psychology*, *8*, 502. Doi: https://doi.org/10.3389/fpsyg.2017.00502

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278. doi: 10.1146/annurev.psych.59.103006.093629

Freeman, D., Lister, R., & Evans, N. (2014). The use of intuitive and analytic reasoning styles by patients with persecutory delusions. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(4), 454-458. doi: http://dx.doi.org/10.1016/j.jbtep.2014.06.005

Frensch, P. A., & Rünger, D. (2003). Implicit Learning. *Curr. Dir. Psychol. Sci.*, *12*(1), 13–18. doi: https://doi.org/10.1111/1467-8721.01213

Fu, Q., Dienes, Z., & Fu, X. (2010). Can unconscious knowledge allow control in sequence learning?. *Consciousness and Cognition*, *19*(1), 462-474.

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychol. Sci.*, *13*(1), 34-40. doi: https://doi.org/10.1111/1467-9280.00406

Gillanders, D. T., Bolderston, H., Bond, F. W., Dempster, M., Flaxman, P. E., Campbell, L., ... & Masley, S. (2014). The development and initial validation of the Cognitive Fusion Questionnaire. *Behavior Therapy*, *45*(1), 83-101. doi:10.1016/j.beth.2013.09.001

Goldin, G., van't Wout, M., Sloman, S. A., Evans, D. W., Greenberg, B. D., & Rasmussen, S. A. (2013). Risk judgment in Obsessive–Compulsive Disorder: Testing a dual-systems account. *Journal of Obsessive-Compulsive and Related Disorders*, *2*(4), 406-411. doi: http://dx.doi.org/10.1016/j.jocrd.2013.08.002

Haider, H., Eichler, A., & Lange, T. (2011). An old problem: How can we distinguish between conscious and unconscious knowledge acquired in an implicit learning task?. *Consciousness and Cognition*, *20*(3), 658-672. Doi: https://doi.org/10.1016/j.concog.2010.10.021

Hallion, L. S., & Ruscio, A. M. (2011). A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychological Bulletin*, *137*(6), 940-958.

Hames, J. L., Hagan, C. R., & Joiner, T. E. (2013). Interpersonal Processes in Depression. *Annu. Rev. Clin. Psychol.*, *9*(1), 355–377. doi: https://doi.org/10.1146/annurev-clinpsy-050212-185553

Hayes, S. C., Levin, M. E., Plumb-Vilardaga, J., Villatte, J. L., & Pistorello, J. (2013). Acceptance and commitment therapy and contextual behavioral science: Examining the progress of a distinctive model of behavioral and cognitive therapy. *Behavior Therapy, 44(2),* 180-198. doi:10.1016/j.beth.2009.08.002

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change.* New York, NY: Guilford Press.

Hudson, M., Nijboer, T. C. W., & Jellema, T. (2012). Implicit Social Learning in Relation to Autistic-Like Traits. *J. Autism Dev. Disord.*, *42*(12), 2534–2545. doi: https://doi.org/10.1007/s10803-012-1510-3

Hur, J., Iordan, A. D., Berenbaum, H., & Dolcos, F. (2016). Emotion–attention interactions in fear conditioning: Moderation by executive load, neuroticism, and awareness. *Biological psychology*, *121*, 213-220. Doi: https://doi.org/10.1016/j.biopsycho.2015.10.007

Ingram, R. E., & Siegle, G. J. (2010). Cognitive science and the conceptual foundations of cognitive-behavioral therapy: Viva la evolution!. In K. S. Dobson (Ed.), *Handbook of cognitive-behavioral therapies* (3rd ed., pp. 74-93). New York: Guilford Press.

Ivanchei, I. I., & Asvarisch, A. (2018). The Nature Of Affect In The Structural Mere Exposure Effect. *National Research University Higher School of Economics, Working Papers, Series: Psychology* (No. WP BRP 99/PSY/2018).

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E., & Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, *64*(2), 295-304. doi: 10.1037/0022-006X.64.2.295

Janacsek, K., Borbély-Ipkovich, E., Nemeth, D., & Gonda, X. (2018). How can the depressed mind extract and remember predictive relationships of the environment? Evidence from implicit probabilistic sequence learning. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 81,* 17-24

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, *15*(4), 496-505. Doi: https://doi.org/10.1111/j.1467-7687.2012.01150.x

Jurchiș, R., & Opre, A. (2018). A Dual Process Approach to Current Issues from Cognitive Behavior Therapies: An Integrative Perspective. *International Journal of Cognitive Therapy*, *11*(4), 374-409. https://doi.org/10.1007/s41811-018-0023-z

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak, & R.G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267-293). Cambridge, UK: Cambridge University Press.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? A focusing illusion. *Science*, *312*(5782), 1908-1910. doi: 10.1126/science.1129688

Kóbor, A., Janacsek, K., Takács, Á., & Nemeth, D. (2017). Statistical learning leads to persistent memory: Evidence for one-year consolidation. *Scientific Reports*, *7*(1), 760. Doi: 10.1038/s41598-017-00807-3

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior Research Methods*, *49*(2), 457-470. https://doi.org/10.3758/s13428-016-0715-3

Ladegaard, N., Larsen, E. R., Videbech, P., & Lysaker, P. H. (2014). Higher-order social cognition in first-episode major depression. *Psychiatry Res.*, *216*(1), 37–43. doi: https://doi.org/10.1016/j.psychres.2013.12.010

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International affective picture system (IAPS): Technical manual and affective ratings.* Gainesville, FL: University of Florida, Center for Research in Psychophysiology.

Leganes-Fonteneau, M., Scott, R., & Duka, T. (2018). Attentional responses to stimuli associated with a reward can occur in the absence of knowledge of their predictive values. *Behavioural Brain Research, 341,* 26-36. Doi: https://doi.org/10.1016/j.bbr.2017.12.015

Leung, J. H., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, *33*(1), 33-55. https://doi.org/10.1017/S0272263110000525

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 135 - 146.

Lieberman, M. D. (2007). Social Cognitive Neuroscience: A Review of Core Processes. *Annu. Rev. Psychol.*, *58*(1), 259–289. doi: https://doi.org/10.1146/annurev.psych.58.110405.085654

Longmore, R. J., & Worrell, M. (2007). Do we need to challenge thoughts in cognitive behavior therapy?. *Clinical Psychology Review*, *27*(2), 173-187. doi: 10.1016/j.cpr.2006.08.001

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior research methods*, *46*(2), 596-610. doi: 10.3758/s13428-013-0379-1

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of abnormal psychology*, *109*(4), 602-615

Mathews, R.C. (1990). Abstractness of implicit grammar knowledge: Comments on Perruchet and Pacteau's analysis of synthetic grammar learning. *Journal of Experimental Psychology: General, 119*, 412-416.

McManus, F., Van Doorn, K., & Yiend, J. (2012). Examining the effects of thought records and behavioral experiments in instigating belief change. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(1), 540-547. doi: http://dx.doi.org/10.1016/j.jbtep.2011.07.003

Mealor, A. D., & Dienes, Z. (2013). The speed of metacognition: Taking time to get to know one's structural knowledge. *Consciousness and Cognition*, *22*(1), 123-136. doi: 10.1016/j.concog.2012.11.009

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1-32. Doi: https://doi.org/10.1016/0010-0285(87)90002-8

Norman, E., & Price, M. C. (2012). Social intuition as a form of implicit learning: Sequences of body movements are learned less explicitly than letter sequences. *Advances in Cognitive Psychology*, 8(2), 121.

Norman, E., Price, M. C., & Jones, E. (2011). Measuring strategic control in artificial grammar learning. *Consciousness and Cognition, 20*, 1920-1929. https://doi.org/10.1016/j.concog.2011.07.008

Norman, E., Scott, R. B., Price, M. C., & Dienes, Z. (2016). The relationship between strategic control and conscious structural knowledge in artificial grammar learning. *Consciousness and Cognition*, *42*, 229-236. https://doi.org/10.1016/j.concog.2016.03.014

Ormel, J., Bastiaansen, A., Riese, H., Bos, E. H., Servaas, M., Ellenbogen, M., ... & Aleman, A. (2013). The biological and psychological basis of neuroticism: current status and future directions. *Neuroscience & Biobehavioral Reviews*, *37*(1), 59-72. Doi: https://doi.org/10.1016/j.neubiorev.2012.09.004

Paciorek, A., & Williams, J. N. (2015). Semantic generalization in implicit language learning. *Journal of Experimental Psychology: Learning, Memory, and cognition*, *41*(4), 989-1002. http://dx.doi.org/10.1037/xlm0000100

Pedersen, T., Kristensson, P., & Friman, M. (2012). Counteracting the focusing illusion: Effects of defocusing on car users' predicted satisfaction with public transport. *Journal of Environmental Psychology*, *32*(1), 30-36. doi:10.1016/j.jenvp.2011.10.004

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge?. *J. Exp. Psychol. Gen.*, *119*(3), 264 - 275.

Peters, K. (2009). Attenuation of Depressogenic Attributional Style to Improve Negative Mood. PhD Thesis, Tulane University, USA

Petrican, R., Moscovitch, M., & Schimmack, U. (2008). Cognitive resources, valence and memory retrieval of emotional events in older adults. *Psychology and Aging 23(3),* 585-594. Doi: 10.1037/a0013176

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*(2), 227-244. doi: 10.1037/0033-2909.133.2.227

Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verb. Learn. Verb. Behav*, *6*(6), 855–863. doi: https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1993). *Implicit learning: An essay on the cognitive unconscious.* New York, NY: Oxford University Press.

Reber, A.S. (1990). On the primacy of the implicit: Comment on Perruchet and Pacteau. *Journal of Experimental Psychology: General, 119, 340-342.*

Samoilov, A., & Goldfried, M. R. (2000). Role of emotion in Cognitive-Behavior Therapy. *Clinical Psychology: Science and Practice, 7*(4), 373-385. doi: 10.1093/clipsy.7.4.373

Scher, C. D., Ingram, R. E., & Segal, Z. V. (2005). Cognitive reactivity and vulnerability: Empirical evaluation of construct activation and cognitive diathesis in unipolar depression. *Clinical Psychology Review, 25*, 487-510. Doi: 10.1016/j.cpr.2005.01.005

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, *9*(5), 340-346. Doi: 10.1111/1467-9280.00066

Schwarz, N., & Clore, G. L. (1988). How do I feel about it? The informative function of affective states. In K. Fiedler and J. P. Forgas (Eds.), Affect, Cognition, and Social Behavior. Gottingen, Germany: Hogrefe, pp. 44-62.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*(2), 195-202. Doi: 10.1037/0022-3514.61.2.195

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1264-1288. http://dx.doi.org/10.1037/a0012943

Sense, F., & van Rijn, H. (2018). Probabilistic motor sequence learning in a virtual reality serial reaction time task. *PloS one*, *13*(6), e0198759.

Shang, J., Fu, Q., Dienes, Z., Shao, C., & Fu, X. (2013). Negative affect reduces performance in implicit sequence learning. *PLoS One, 8*(1), e54693. https://doi.org/10.1371/journal.pone.0054693

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology*, *61*, 273-301. doi: 10.1146/annurev.psych.093008.100519

Shanks, D. R., & St John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*(03), 367-395. doi:*10.1017/S0140525X00035032*

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The role of awareness in attitude formation through evaluative conditioning. *Personality and Social Psychology Review*, *18*(2), 187-209. https://doi.org/10.1177/1088868314527832

Szegedi-Hallgató, E., Janacsek, K., Vékony, T., Tasi, L. A., Kerepes, L., Hompoth, E. A., ... & Németh, D. (2017). Explicit instructions and consolidation promote rewiring of automatic behaviors in the human mind. *Scientific Reports*, *7*(1), 4365. Doi: 10.1038/s41598-017-04500-

Taleb NN. (2018). Skin in the game: Hidden asymmetries in daily life. New York, NY: Random House

Troy, A. S., Shallcross, A. J., Davis, T. S., & Mauss, I. B. (2013). History of mindfulness-based cognitive therapy is associated with increased cognitive reappraisal ability. *Mindfulness*, *4*(3), 213-222. Doi: 10.1007/s12671-012-0114-5

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293-315. doi: 10.1037/0033-295X.90.4.293

Vîslă, A., Grosse Holtforth, M., & David, D. (2015). Descriptive/inferential cognitive processes and evaluative cognitive processes: Relationships among each other and with emotional distress. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *33*(2), 148-159.

Wierzchoń, M., Asanowicz, D., Paulewicz, B., & Cleeremans, A. (2012). Subjective measures of consciousness in artificial grammar learning task. *Consciousness and Cognition*, *21*(3), 1141-1153. http://dx.doi.org/10.1037/a0012943

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews, A. (1997). *Cognitive psychology and emotional disorders* (2nd ed.). New York, NY: Wiley & Sons

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: a source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*(5), 821-836. Doi: 10.1037/0022-3514.78.5.821

Ziori, E., & Dienes, Z. (2015). Facial beauty affects implicit and explicit learning of men and women differently. *Frontiers in Psychology*, *6*, 1124. https://doi.org/10.3389/fpsyg.2015.01124