Katholieke Universiteit Leuven
Faculteit Psychologie en Pedagogische Wetenschappen

Centrum voor Leerpsychologie en Gedragstherapie

Beyond Evaluative Conditioning?

Searching for Associative Transfer of Non-Evaluative Stimulus Properties

Proefschrift aangeboden tot
het verkrijgen van de graad van
Doctor in de Psychologische
Wetenschappen
door

Tom Meersmans

Prof. Dr. Frank Baeyens, promotor
Prof. Dr. Jan De Houwer, copromotor

2004

Tom Meersmans Beyond Evaluative Conditioning? Searching for Associative
Transfer of Non-Evaluative Stimulus Properties

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologische Wetenschappen, 2004.

Leiding: Prof. Dr. Frank Baeyens, promotor, en Prof. Dr. Jan De Houwer, copromotor.

Evaluatieve conditionering verwijst naar de veranderingen in aangenaamheid van een oorspronkelijk neutrale prikkel (de VP of Voorwaardelijke Prikkel) als gevolg van het gepaard aanbieden van die prikkel met een aangename of onaangename prikkel (de OP of Onvoorwaardelijk Prikkel). Dit type van leren kan men zien als een vorm van (louter) referentieel leren: door het herhaaldelijk paren van een VP met een OP, zal het aanbieden van de VP op zichzelf genoeg zijn om de representatie van de (valentie van de) OP te activeren, zonder dat daardoor noodzakelijk de OP ook reëel verwacht wordt. Deze geactiveerde OP-valentie wordt vervolgens ('ten onrechte') toegewezen/toegeschreven aan de VP waardoor de gepercipieerde valentie van de VP verandert. Wij onderzochten of ook andere, niet-evaluatieve prikkeleigenschappen van een OP, kunnen worden overgedragen naar de VP. We vonden hiervoor aanwijzingen in de experimenten van Stevenson en collega's (vb, Stevenson, Boakes, & Wilson, 2000). In een reeks van experimenten boden de auteurs een geur samen aan met ofwel een zoete ofwel een zure smaak. Nadien beoordeelden de deelnemers deze geur als respectievelijk zoeter of zuurder. Dit resultaat kan men zien als evidentie voor niet-evaluatief referentieel leren. Meer onderzoek is echter nodig om na te gaan of deze associatieve overdracht van niet-evaluatieve prikkeleigenschappen een algemeen fenomeen is. Het doel van dit doctoraat was dan ook het onderzoeken van associatieve overdracht van niet-evaluatieve prikkeleigenschappen met niet-olfactorische en niet-gustatorische stimuli.

Een eerste reeks van experimenten was gebaseerd op het beeld-beeld paradigma dat Baeyens en collega's gebruikten om evaluatieve conditionering te onderzoeken (vb, Baeyens, Eelen, van den Bergh, & Crombez, 1989). Met een gelijkaardige procedure probeerden we de leeftijd van mensen (één experiment) en het geslacht van kinderen (vijf experimenten) over te dragen op foto's van andere mensen. In tegenstelling tot onze verwachting was het niet mogelijk om deze niet-evaluatieve eigenschappen betrouwbaar over te dragen. Een tweede lijn van experimenten was gebaseerd op een symbool-woord paradigma dat Diaz, Baeyens en Ruiz (2002) succesvol gebruikten om evaluatief leren te demonstreren. Met een gelijkaardig paradigma probeerden we het geslacht van voornamen over te dragen op Japanse karakters. In twee experimenten vonden we evidentie voor geslachtsoverdracht, maar de data suggereerden dat meer dan eenvoudige associaties nodig zijn om het effect te bekomen. In de derde lijn van experimenten gebruikten we een vreesconditioneringsparadigma om na te gaan of niet-evaluatieve eigenschappen van een aversieve OP ook geactiveerd worden door de VP. In vijf experimenten probeerden we een verband te creëren tussen de richting van waaruit een aversieve OP kwam en een kleur. Alhoewel we in twee experimenten een activatie-effect vonden, konden we het niet repliceren in de drie daaropvolgende studies. Daarenboven zouden de activatie-effecten die we vonden in de eerste twee experimenten een vorm van verwachtingsleren kunnen zijn eerder dan van louter referentieel leren.

Samengevat kunnen we zeggen dat we geen sterke evidentie hebben voor het bestaan van puur associatieve overdracht van niet-evaluatieve prikkeleigenschappen. Desondanks kunnen we niet beweren dat dit fenomeen niet bestaat omdat een aantal (onbekende) grenscondities voor het bekomen van deze associatieve overdracht misschien niet vervuld waren. Toch kunnen we, op basis van onze studies en verschillende niet gepubliceerde studies van anderen, minstens besluiten dat de grenscondities die gelden voor deze vorm van leren veelvuldig zijn.

Voorwoord

Eind vorige eeuw gaf Paul Eelen me de kans en het vertrouwen om aan een onderzoeksproject te beginnen, dat zou moeten uitmonden in een doctoraat. Vandaag, ongeveer 1160 dagen, 19 experimenten en 620 ‘vrijwillige deelnemers’ later, ligt het er eindelijk. Het had er echter nooit gelegen zonder de hulp en steun van een hele reeks mensen. Via deze weg wil ik hun van harte bedanken...

In de eerste plaats mijn promotor en copromotor, Frank Baeyens en Jan De Houwer. Bedankt voor alle discussies, goeie suggesties, naleeswerk en interesse. Frank, uit je altijd zeer heldere uitleg op mijn kleine en grote vragen, heb ik veel bijgeleerd. Jan, uw pen/vingers hebben een magische invloed op alle teksten die ze tegenkomen. Jullie zijn allebei onderzoekers met topklasse!

Ik wil ook alle collega's en ex-collega's van het centrum bedanken voor de toffe momenten en aangename sfeer: Bram (du hast?), Geert, Adriaan, Loes, Filip, Netje, Tom, Agnes, Carlos, Deb, Dirk, Paul, Sabine, Jeroen, Corry en Eva. Zonder jullie zou het onderzoek een stuk eenzamer geweest zijn. Ook Jan Vanrie, Wim De Neys en Jan Vos wil ik bedanken: ook al zitten jullie niet in ons centrum, jullie waren goeie collega's.

De voorzitter en leden van mijn doctoraatscommissie wil ik bedanken voor hun kritische inbreng: Prof. Dr. Willy Lens, Prof. Dr. Eddy Van Avermaet, Prof. Dr. Guido Peeters, Prof. Dr. Gert Storms en Prof. Dr. Karl Verfaillie. Ook volgende mensen verdienen en dankuwel: Tom Randell, voor het uitvoeren van een experiment in Southampton; Hilde Hendrickx en Estrella Diaz, voor het ter beschikking stellen van hun stimulusmateriaal; en alle ouders die me toelieten een foto te nemen van hun kleine spruit.

Niet te vergeten de mensen die me buiten de werkuren hebben bijgestaan. Zonder de kansen en steun van mijn moeder en grootouders was ik nooit zo ver geraakt. En niet te vergeten de ontspanning en bemoediging van alle vrienden.

Last, but not least wil ik Leen bedanken. Tijdens de zware periodes heb je me altijd bijgestaan en gesteund. Voor dat en al de rest: dank je, Leen!

Table of Contents

Table of Contents.....	i
List of Tables.....	v
List of Figures	vii
Introduction	1
Chapter 1 Evaluative Conditioning	3
Generality	4
Visual stimuli	4
Gustatory stimuli	5
Haptic stimuli	5
Biologically significant stimuli	5
Cross-modal stimuli.....	7
Observational EC.....	7
Failures to replicate EC	7
Functional characteristics	8
Extinction	8
Contingency-awareness.....	9
Statistical contingency.....	10
Occasion setting.....	11
Counter-conditioning.....	12
Post-acquisition US-revaluation.....	12
Sensory preconditioning.....	13
US pre- and post-exposure	13
Models of EC.....	14
Intrinsic change account.....	14
Holistic representation account	15
Conceptual categorization account.....	16
Artefactual account.....	17
Emotional-affective versus Sensory-descriptive model	19
Referential versus Expectancy system	20
Chapter 2 Beyond Evaluative Conditioning?	23
Existing evidence for ATNESP.....	24
Research objectives	26
Chapter 3 Picture-Picture Studies	29
Experiment P1	30
Method.....	30
Results	32
Discussion	34
Experiment P2	35
Method.....	35
Results	37
Discussion	39

Table of Contents

Experiment P3	40
Method.....	41
Results	45
Discussion.....	47
Experiment P4	48
Method.....	48
Results	49
Discussion.....	50
Experiment P5	50
Method.....	51
Results	52
Discussion.....	54
Experiment P6	54
Method.....	55
Results	56
Discussion.....	57
Experiment P7	58
Method.....	58
Results	59
Discussion.....	61
Experiment P8	62
Method.....	62
Results	63
Discussion.....	64
Overall discussion of picture-picture studies	65
Chapter 4 Symbol-Word Studies	69
Experiment S1	69
Method.....	70
Results	72
Discussion.....	76
Experiment S2	76
Method.....	77
Results	78
Discussion.....	80
Overall discussion of the symbol-word studies.....	80
Chapter 5 Fear Conditioning Studies.....	85
Experiment F1	85
Method.....	86
Results	91
Discussion.....	94
Experiment F2	95
Method.....	96
Results	96
Discussion.....	98
Experiment F3	99
Method.....	100
Results	100
Discussion.....	101

Experiment F4	102
Method.....	103
Results	103
Discussion	105
Experiment F5	105
Method.....	105
Results	106
Discussion	107
Overall discussion of the fear conditioning studies.....	108
Chapter 6 Other Studies	111
Other failures	111
Related studies.....	113
Conditional after effects	113
Implicit covariation detection.....	115
Chapter 7 General Discussion and Conclusion.....	121
General discussion.....	121
Is there evidence for general associative transfer of non-evaluative stimulus properties?.....	121
Is there evidence for associative <i>activation</i> of non-evaluative stimulus properties?.....	122
Other forms of associative- <i>like</i> transfer of non-evaluative properties?.....	123
Why did we not find clear evidence for ATNESP?.....	124
Conclusion.....	126
Samenvatting [Dutch Summary].....	129
Inleiding.....	129
Hoofdstuk 1: Evaluatieve conditionering.....	130
Hoofdstuk 2: Meer dan evaluatieve conditionering?.....	133
Hoofdstuk 3: Beeld-beeld studies.....	134
Hoofdstuk 4: Symbool-woord studies	135
Hoofdstuk 5: Vreesconditionering studies	136
Hoofdstuk 6: Andere studies	137
Hoofdstuk 7: Algemene discussie en besluit.....	139
References.....	143
Appendix	153

List of Tables

Table 1: <i>Average Age-Ratings Before and After Acquisition for the CS- and US-Categories from Experiment P1 (standard errors in parentheses)</i>	33
Table 2: <i>Average Gender-Ratings before and/or after acquisition for the CS-Categories from Experiments P2, P3, P4, P5, P6 and P8 (standard errors in parentheses)</i>	38
Table 3: <i>Most Important Procedural Differences Between the Transfer of Gender Experiments</i>	43
Table 4: <i>Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Gender-Priming Measure from Experiments P3 and P4 (standard errors in parentheses)</i>	45
Table 5: <i>Average Percentage of Male Names (%Male) Given in the Naming Tasks of Experiments P5 and P6 (standard errors in parentheses)</i>	53
Table 6: <i>Average Valence-Ratings for the CS-Categories from Experiment P7 (standard errors in parentheses)</i>	60
Table 7: <i>Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Affective Priming Measure from Experiment P7 (standard errors in parentheses)</i> ...	60
Table 8: <i>Average Gender-Ratings for the CS-Categories from the Symbol-Word Experiments (standard errors in parentheses)</i>	72
Table 9: <i>Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Gender Priming Tasks from the Symbol-Word Experiments (standard errors in parentheses)</i>	74
Table 10: <i>Average Reaction Times (RT) for the Gender Priming Task, for each Order Condition and Awareness Group from Experiments S1 and S2 (standard errors in parentheses)</i>	82
Table 11: <i>Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Affective Simon Tasks from the Fear Conditioning Experiments (standard errors in parentheses)</i>	91
Table 12: <i>Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Spatial Simon Tasks from the Fear Conditioning Experiments (standard errors in parentheses)</i>	92
Table 13: <i>Average Valence-Ratings for the CS-Categories from the Fear Conditioning Experiments (standard errors in parentheses)</i>	94

List of Figures

Figure 1: <i>Schematic Example of Successful Associative Spatial Simon.</i>	90
Figure 2: <i>Schematic Example of Successful Associative Affective Simon.</i>	90

Introduction

People's preferences are not all innate. In fact, most of our likes and dislikes are acquired through individual experience (Rozin & Millman, 1987). Evaluative conditioning is one of the mechanisms through which preferences are acquired (Martin and Levey, 1978). Evaluative conditioning refers to the phenomenon that simply pairing an affectively neutral stimulus with a stimulus that is already liked or disliked, can change our liking of the originally neutral stimulus. For example, simply repeatedly pairing an affectively neutral picture (called the conditional stimulus; CS) with a liked/disliked picture (called the unconditional stimulus; US), will make that originally neutral picture more positive/negative. This form of conditioning has received a lot of attention the past 25 years (for a review, see De Houwer, Thomas, and Baeyens, 2001), especially because it seems to behave different than more traditional forms of Pavlovian conditioning. Evaluative conditioning seems to be insensitive to extinction-procedures: repeatedly presenting the CS by itself after acquisition, does not abolish the acquired valence. Another difference is that evaluative conditioning seems to be independent of awareness of the contingencies between the conditional and unconditional stimulus. Different theories have been developed to account for these functional differences between evaluative conditioning and more traditional forms of Pavlovian conditioning. One of the most promising ones claims that evaluative conditioning is a form of merely referential learning (Baeyens & De Houwer, 1995; Baeyens, Eelen & Crombez, 1995; Baeyens, Eelen, Crombez, & Van den Bergh, 1992): by repeatedly pairing a conditional stimulus with a clearly liked/disliked unconditional stimulus, a merely referential association will be formed between those two stimuli. If the conditional stimulus is subsequently presented by itself, it will activate the representation of the liked/disliked unconditional stimulus, resulting in the unconditional stimulus valence being 'misattributed' to the conditional stimulus.

Up to now the idea of referential learning has only been applied in the context of transfer of valence. But if the judgement of the valence of a conditional stimulus is influenced by the valence of stimuli with which it has been repeatedly paired, then maybe judgements of other properties of that conditional stimulus will also be influenced by these associations. In other words, maybe not only the valence of stimuli can be associatively 'transferred', but also other stimulus-properties. There have been a few experiments

conducted by Stevenson and colleagues (e.g., Stevenson, Boakes, & Prescott, 1998; Stevenson, Boakes, & Wilson, 2000a, 2000b; Stevenson & Case, 2003; Stevenson, Prescott, & Boakes, 1995) that suggest that this could be possible indeed. In a series of studies, the authors repeatedly paired a flavour with a sweet or sour taste. Afterwards, participants experienced the smell of the flavour as being sweeter or sourer respectively. The shift seemed to be independent of valence and this 'sweetness/sourness'-transfer seemed to behave similarly as evaluative conditioning (e.g., resistant to extinction, independent of contingency-awareness). However, the authors do not explain this non-evaluative transfer as referential learning. Stevenson et al. (2000a, 2000b) suggest that flavour-taste transfer might be something unique: flavours and tastes are sensory very close to one another and might be stored holistically. This non-evaluative transfer might therefore be unique for olfactory-gustatory stimuli. More research is necessary to support the claim that associative transfer of non-evaluative stimulus properties is a general phenomenon.

The main aim of this dissertation was to search for evidence of associative transfer of non-evaluative stimulus properties, especially with non-gustatory stimuli. If such evidence would be found, we planned to investigate whether this transfer of (non-evaluative) properties behaves in the same way as evaluative conditioning does (e.g., resistant to extinction, independent of contingency-awareness,...). In Chapter 1 we will give an overview of evaluative conditioning: discussing the generality, the functional characteristics and the models for evaluative conditioning. In Chapter 2, we go a step further and look at associative transfer of non-evaluative properties. We will discuss the sparse empirical evidence and end with the research objectives. In the next three chapters we will discuss our experiments in which we searched for associative transfer of non-evaluative properties. In each chapter we used a different paradigm: In Chapter 3 we describe the picture-picture experiments, in Chapter 4 the symbol-word experiments and in Chapter 5 the fear-conditioning experiments. In Chapter 6 we take a look at the experiments of other researchers that relate to associative transfer of non-evaluative stimulus properties. Finally, in Chapter 7 we provide a summary of our findings and the findings of other researchers and attempt to draw conclusions.

Chapter 1

Evaluative Conditioning

At a procedural level, evaluative conditioning (EC) can be seen as a form of Pavlovian conditioning (PC). In traditional PC studies, a biologically neutral stimulus (Conditional Stimulus or CS) is repeatedly paired with a biologically significant stimulus (Unconditional Stimulus or US), after which the originally neutral stimulus elicits preparatory behaviour (Conditioned Response or CR). For example, if the presentation of food (US) to a dog is repeatedly preceded by the ringing of a bell (CS), the dog will start salivating when it hears the bell (Pavlov, 1927). In EC, an affectively neutral stimulus is repeatedly paired with an affectively relevant (positive or negative) stimulus, after which the perceived valence of the originally neutral stimulus changes (e.g., Martin and Levey, 1978). The affectively neutral stimulus is equivalent to the CS and the affectively loaded stimulus to the US. The changed evaluative rating can be seen as a CR.

Modern EC research is based on the work of Martin and Levey (1978). Inspired by earlier demonstrations of EC (Razran, 1954; Staats and Staats, 1957), Martin and Levey suggested that the transfer of valence is the core mechanism that drives classical conditioning. Since the early demonstrations of EC, a variety of researchers from different backgrounds have studied the phenomenon (for reviews, see Baeyens et al., 2001a; De Houwer, Thomas, & Baeyens, 2001; Field, in press). In the following paragraphs we will give a short overview of EC. In the first section we will briefly look at the wide range of stimuli and procedures with which EC has been demonstrated. The second section focuses on the most prominent functional characteristics of this type of conditioning. In the third section we will describe a few models that try to explain EC. As such, this chapter is structured along the same lines as and is in a large part based on the review of De Houwer, Thomas, and Baeyens (2001).

Generality

Visual stimuli

One of the first preparations in which EC was systematically investigated is called the ‘picture-picture’ paradigm. In the picture-picture paradigm, visual stimuli are used as CSs and USs. The most well known variant is probably the human faces picture-picture procedure as implemented by Baeyens and colleagues (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Baeyens, Eelen, & Van den Bergh, 1990; Baeyens, Eelen, Van den Bergh, & Crombez, 1989a, 1989b, 1992; Baeyens, Hermans, & Eelen, 1993). The basic procedure is as follows: In a first phase (baseline rating) participants were asked to express their liking for a series of pictures of human faces on a liked/disliked rating scale. On the basis of these ratings, three types of pictures were selected: the most liked, the most disliked and the most neutral faces. The selected pictures were used to make three types of CS-US pairs: neutral-liked, neutral-disliked and neutral-neutral pairs. After repeated presentations of these pairs (e.g., 1 sec CS, 4 sec blank screen, 1 sec US, 8 second until the next trial), participants were invited to re-rate the CS-pictures. The prototypical result of this kind of experiment is that the valence of the previously neutral CS-picture had shifted in the direction of the valence of the US with which it had been paired: so an originally neutral face that had been repeatedly paired with a liked face would have become more positive, while an originally neutral face that had been repeatedly paired with a disliked face would have become more negative. Other experiments have replicated EC with other visual stimuli, like pictures of sculptures (Hammerl & Grabitz, 1993) or paintings (Baeyens et al., 1989b). It is also possible to use positive/negative words as USs. Staats and Staats (1957), for example, found EC when repeatedly pairing a non-sense word with clearly positive or negative words. Diaz, Baeyens and Ruiz (2002) obtained EC when pairing Japanese characters with clearly positive or negative words. There are also EC-studies in which an invariant feature of stimuli is used as a CS. Baeyens, Heremans, Eelen, and Crombez (1993), for example, arranged a correlation between the line thickness (CS) of different geometrical figures with the presentation of a liked or disliked picture. After acquisition, participants liked new figures better if they were presented in the line thickness that had always been paired with positive pictures, compared to new figures that were presented in the line thickness that had always been paired with negative pictures.

Gustatory stimuli

EC has also been repeatedly demonstrated in flavour-taste paradigms in which a relatively neutral flavour is repeatedly paired with either a positive or a negative taste (e.g., Baeyens, Crombez, Hendrickx, & Eelen, 1995; Baeyens, Eelen, Van den Bergh, & Crombez, 1990; Baeyens, Vanhouche, Crombez, & Eelen, 1998; Zellner, Rozin, Aron, & Kulish, 1983). In a prototypical flavour-taste experiment, participants are exposed to a relatively neutral CS-flavour (e.g., apricot) that is paired with a negative US-taste (e.g., Tween20, a soaplike taste). After repeated consumptions of the compound, participants are presented with the CS-flavour in plain water (without the US-taste), and asked to indicate how much they like or disliked the flavour. In general, the flavour that had previously been paired with the negative taste will be rated more negatively than a control flavour (e.g., pear) that had never been paired with a negative taste. For a review of human evaluative flavour-taste conditioning, see Baeyens, Vansteenwegen, Hermans, and Eelen (2001b).

Haptic stimuli

Hammerl and Grabitz (2000) found evidence for EC using haptic stimuli. During the experiment, participants were asked to touch objects that were hidden from view inside a box. Before acquisition, participants indicated how much they liked the ‘feel’ of each of a set of items with different textures (e.g., wood, silk, sandpaper,...). Based on these ratings the CSs and USs were selected. During acquisition two types of pairs were repeatedly presented: neutral-liked pairs (i.e., first touching something that they found neutral, then touching something that felt nice) and neutral-neutral pairs (i.e., sequentially touching two neutral objects). In a subsequent re-rating phase, previously neutral textures (CSs) that had been paired with a liked texture (USs) were rated more positive than neutral textures that had been paired with other neutral textures.

Biologically significant stimuli

One of the largest procedural differences between EC and traditional PC is that in traditional PC biologically significant stimuli (e.g., food, electric shock) are used as USs, while in EC the USs are not or at least less biologically relevant. There are however, some studies that use biologically significant USs to investigate EC. Johnsrude, Owen, Zhao, and White (1999), for example, obtained EC when pairing visual patterns (CS) with food

rewards (US). Other researchers have also observed EC after repeatedly pairing pictures (CS) with an electrical shock (US) (e.g., Hermans, Spruyt, & Eelen, 2003; Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002; Francken, 2003). This type of fear-conditioning paradigm has received much attention in recent years, because it provides a procedure allowing to directly compare (the functional characteristics of) traditional PC and EC within the same experiment. One of the disadvantages of this type of experiment is that the CS-US contingencies are very apparent and thus sensitive to demand effects. To avoid these problems, researchers started using more indirect measures for EC. The most promising measure thus far appears to be affective priming. In a standard affective priming task (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Hermans, De Houwer, & Eelen, 1994), on each trial a positive or negative target-stimulus is presented that has to be evaluated as quickly as possible as either positive or negative. Each target stimulus is preceded by the brief presentation of another neutral, positive or negative stimulus. Although participants are instructed to ignore the prime stimulus, the response time to evaluate the target is mediated by the valence of the prime: Participants categorize the target faster if the prime has the same valence as the target (i.e., congruent trial) than when the prime valence differs from the target valence (i.e., incongruent trial). This so called ‘affective priming’ effect seems to be caused by the automatic processing of the valence of the prime (Hermans, Van den Broeck, & Eelen, 1998). As such, the affective priming procedure provides an unobtrusive, and thus less demand sensitive, way to assess the valence of the prime picture. If participants respond faster to a negative target when it is preceded by CS1 than when the target is preceded by CS2, then we have evidence that the CS1 has acquired a (more) negative valence than the CS2 (an analogous argument can be made for positive targets). Take for example the experiments of Hermans, Vansteenwegen, et al. (2002), who investigated EC by pairing pictures of human faces (CS) with an electrical shock (US), and used affective priming as indirect measure of valence. During acquisition, one neutral face (CS+) was always paired with an aversive shock, while another neutral face was never paired with a shock (CS-). After acquisition and the subsequent re-rating phase, both CSs were introduced as primes in an affective priming measure that included positive and negative target words. Participants were faster in evaluating the valence of a negative target if it was preceded by the CS+ or of a positive target if it was preceded by the CS-, than the other way around. These results provide clear evidence for EC with biologically significant USs. Note that this affective priming procedure has also been used in EC-studies without biologically significant stimuli (e.g., Diaz et al., 2002).

Cross-modal stimuli

EC can also occur when the CS and the US belong to different sensory modalities. For example, Todrank, Byrnes, Wrzesniewski, and Rozin (1995) demonstrated EC by pairing pictures of human faces (CS) with liked or disliked odour stimuli (US). Eifert, Craill, Carey, and O'Connor (1988) changed the valence of neutral Greek letters (CS) by pairing them with liked or disliked music (US). There have also been a few field studies, in which researchers have tried to demonstrate EC in a real-life context, that fit into the cross-modal category. Baeyens, Wrzesniewski, De Houwer, and Eelen (1996) placed odorants (CS) in the toilet rooms at the work place for some weeks. After this acquisition period, people who frequented those toilets were asked to rate these odours. Participants who subjectively enjoyed their sanitary breaks (US) subsequently rated the CS-odour more positively than a control odour, while the opposite effect was found for participants who disliked these bathroom activities. In a second experiment similar results were obtained by pairing an odorant (CS) with an agreeable or painful massage (US).

Observational EC

In observational EC the participant is not her/himself exposed to the CS-US contingencies, but observes another individual, the model, really or purportedly being exposed to the CS-US contingencies and expressing his/her evaluation of the US. Experiments by Baeyens and colleagues (Baeyens, Eelen, Crombez, & De Houwer, 2001; Baeyens, Vansteenwegen, De Houwer, & Crombez, 1996) indicated that for the observer the CS can acquire a valence congruent with the valence of the CS for the model. In these experiments children consumed a series of neutral coloured and flavoured drinks while simultaneously watching an actor drinking identical drinks and facially expressing his liking for the drink (i.e., clear dislike or a neutral expression). Results showed that the observers acquired a clear dislike for the CS-drink that the model disliked, compared to the CS-drink that the model found neutral.

Failures to replicate EC

The above descriptions of the wide variety of procedures in which EC has been found, suggests that EC is a pervasive and ecologically valid phenomenon. This however, does not mean that valence transfer will appear in every instance in which an affectively neutral

stimulus is repeatedly paired with a liked/disliked stimulus. There have been failures to observe EC in conceptual and even exact replications of the paradigms described above (see Field, 2002; Field & Davey, 1999; Rozin, Wrzesniewski, & Byrnes, 1998). This suggests that there are clear boundary conditions to evaluative conditioning, which are still unknown.

Functional characteristics

Although at a procedural level EC looks very similar to other forms of PC, there appear to be clear functional differences between traditional PC and EC. In this section we will discuss these functional characteristics of EC.

Extinction

Extinction refers to the phenomenon whereby post-acquisition presentations of the CS alone (i.e., without the US), gradually diminish the previously acquired conditioned response. But unlike most forms of traditional PC, EC appears to be resistant to extinction. Once the CS has acquired the valence of the US with which it had been paired, CS-only presentations do not seem to abolish the acquired valence of the CS. Take for example the picture-picture procedure with human faces of Baeyens and colleagues (Baeyens et al., 1988, 1989a): the valence transfer that had occurred after 10 CS-US pairings, remained intact after an extinction phase in which 5 or 10 CS-only trials were presented. These results were later replicated in a methodically superior experiment of De Houwer, Baeyens, Vansteenwegen, and Eelen (2000). Resistance to extinction of EC has also been demonstrated after having paired Japanese characters with positive/negative words (Diaz et al., 2002), having paired flavours with a neutral/negative taste (Baeyens, Crombez, et al., 1995; Baeyens, Vanhouche, et al., 1998), and having paired human faces with/without an electric shock (Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2002; Francken, 2003). Overall there is considerable evidence that EC is resistant to extinction (but also see Lipp, Oughton, & LeLievre, 2003).

Contingency-awareness

There is evidence that in traditional PC, conditioning only occurs when participants are aware of the CS-US contingencies (for a review, see Lovibond, & Shanks, 2002). In EC on the other hand, a number of studies have provided evidence for learning without contingency awareness (for reviews, see De Houwer, Baeyens, & Hendrickx, 1997; Field, 2000). Baeyens, Eelen, and Van den Bergh (1990) tried to manipulate contingency awareness in a human faces picture-picture experiment: before acquisition participants were either given explicit instructions to search for CS-US contingencies, or were told a cover story that made no reference to the contingencies. To measure contingency awareness at the end of the experiment, participants were asked to reconstruct the pairs presented to them during acquisition: they were presented with each CS picture and asked to select the US picture that had been paired with it. Participants were classified as aware if they a) selected the US that had been paired with that CS, b) selected a US that had the same valence as the US that had been paired with that CS, or c) were unable to select a specific US, but could indicate the valence of the picture that had been paired with it. In addition, the participants had to be either 'sure' or 'rather sure' of their answer. The contingency-awareness manipulation had a clear effect on contingency awareness but did not influence EC.

Other experiments revealed additional dissociations between contingency awareness and the size of the EC-effect. Baeyens, Eelen, Crombez, and Van den Bergh (1992) found that increasing the number of acquisition trials did increase contingency-awareness, but reduced the EC-effect. Baeyens et al. (1989b) observed that increasing the similarity within the CS-US pairs did enhance contingency awareness, but did not influence the transfer-effect. In the flavour-taste paradigm, Baeyens, Eelen, Van den Bergh, and Crombez (1990) observed EC even though none of the participants were aware of the critical CS-US combination. Likewise, in the experiments that associated the line thickness of geometrical figures with liked/disliked pictures, none of the participants were aware of the CS-US contingencies, but still EC occurred (Baeyens, Heremans, et al., 1993). In most other studies no correlation was found between contingency awareness and the size of EC (e.g., Baeyens et al., 1988; Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Baeyens, Hermans, & Eelen, 1993). A few studies even suggest a negative correlation between EC and contingency awareness (e.g., Hammerl & Grabitz, 2000).

Finally, some experiments have successfully used subliminally presented positive/negative USs in an attempt to find EC without awareness (e.g., De Houwer, Baeyens, & Eelen, 1994; De Houwer, Hendrickx, & Baeyens, 1997).

Although the experiments presented here seem to indicate that EC is more or less independent of contingency awareness, this conclusion might be too general. In some experiments contingency awareness seems to be a pre-requisite for obtaining evaluative conditioning (e.g., Purkins & Lipp, 2001). In a critical review of the literature, Field (2000) argued that in many studies the measurement of contingency awareness suffers from methodological and statistical shortcomings. As opinions differ about what measures are sufficiently sensitive, valid and reliable to measure CS-US contingency awareness, it is hard to draw definitive conclusions. Nevertheless, even researchers who are sceptical with regard to the idea of learning without awareness (e.g., Lovibond & Shanks, 2002) acknowledge that there are strong indications that EC at least sometimes does not depend on contingency awareness (also see Baeyens, Vansteenwegen, et al., 2001a).

Statistical contingency

In traditional PC, the degree of statistical contingency between the CS and the US seems crucial (e.g. Rescorla, 1968). For EC, however, there is evidence that not the contingency but the simple contiguity of CS and US is important. Baeyens, Hermans, and Eelen (1993) manipulated the degree of statistical congruency while keeping the number of contiguous CS-US pairings constant in the human faces picture-picture paradigm. In the first condition, there were 10 CS-US presentations (perfect contingency); in the second condition there were an additional 10 CS-only presentations (partial reinforcement); and in the final condition not only 10 CS-only presentations were added but also 10 US only presentations (zero contingency). In all three conditions a clear EC effect was found, and there was no significant difference between the three contingency groups. Note, however, that the crucial result is a null finding that might have been due to a lack of statistical power. In the flavour-taste paradigm additional evidence has been found that EC is insensitive to statistical contingency: Baeyens, Crombez, et al. (1996) compared the effect size of an experiment where only CS-US trials were presented with the effect sizes of experiments in which, in addition to the 6 CS-US trials, also 6 CS-only trials were presented. The effect sizes were very similar if not identical. In other experiments in the flavour-taste paradigm

the experimenters observed that, when using two CSs of which one was more contingent but less spatiotemporally contiguous with the US than the other, it was the more contiguous flavour that became the most disliked, and not the more contingent flavour (Baeyens, Crombez, et al., 1996; Baeyens, Hendrickx, Crombez, & Hermans, 1998).

Occasion setting

Occasion setting, or modulation, refers to a phenomenon observed in some traditional PC preparations in which a stimulus (called the feature) does not elicit or inhibit a conditioned response directly, but modulates responding to another CS (called the target). Take for example a feature-positive schedule, in which the target stimulus A is reinforced if and only if it is accompanied by the feature stimulus X. Holland (1983) demonstrated that this procedure can lead to an A-US association that is modulated by the feature X, that is, the feature X becomes a facilitator or occasion setter that controls the behavioural expression of the A-US association.

Baeyens and colleagues (Baeyens, Crombez, De Houwer, and Eelen, 1996; Baeyens, Hendrickx, et al., 1998) were unable to find modulation in a flavour-taste EC-preparation. In one experiment, the colour of a fluid was used as a feature and a fruit flavour was used as a target CS in a sequential feature-positive schedule with Tween20 (US): the CS-flavour was only paired with the negative Tween-taste if the liquid was in a certain feature-colour. The experimenters found no evidence for modulation, instead participants acquired an un-modulated dislike for the CS-flavour. Similar results were found in conceptual replications of this experiment (Baeyens, Crombez, et al., 1996; Baeyens, Hendrickx, et al., 1998). On the other hand, Hardwick and Lipp (2000), did find evidence in a sequential feature-positive design that might imply occasion setting. However, Hardwick and Lipp (2000) relied solely on blink startle modulation to measure EC, which is a disputed measure for valence (see Vansteenwegen, Crombez, Baeyens, & Eelen, 1998). Moreover, no tests were performed to ascertain that it was in fact occasion setting that was observed (i.e., no effect of feature extinction and specificity of transfer to other targets).

Counter-conditioning

Although EC seems to be resistant to extinction, the associatively induced valence-shift is not irreversible. Just like traditional PC, EC seems to be sensitive to counter-conditioning. Baeyens et al. (1989a) demonstrated this in their human faces picture-picture design. After an acquisition-phase with 10 presentations of the CS-US pairs, another phase followed: one third of the CSs were subsequently paired 10 times with a US of the opposite valence as the one with which they had been paired during acquisition (e.g., a CS that was paired with a positive US during acquisition was now paired with a negative US) (counter-conditioning), another third of the CSs were presented without any US (extinction), and a final third of the CSs were not presented again (control). The results showed that only in the control and the extinction condition an EC effect was observed. In the counter-conditioning condition no net valence shift was observed.

Post-acquisition US-revaluation

Just like traditional PC, EC seems to be sensitive to post-acquisition revaluation of the US. In post-acquisition US-revaluation studies the valence of the US is changed after acquisition and in the absence of the CS (e.g., a positive US is made neutral or negative). After this revaluation, the CS is tested by itself, and is generally found to track the changes in the valence of the US. Baeyens, Eelen, Van den Bergh, and Crombez (1992) investigated US-revaluation in their human faces picture-picture paradigm. After the acquisition phase, half of the USs were revaluated. This revaluation was successfully implemented by presenting the US-pictures together with verbally presented negative or positive adjectives, that purportedly described the personality of the person presented in the US-photo. Half of the USs were presented with congruent adjectives (e.g., a positive US was paired with positive adjectives) and the other half of the USs with incongruent adjectives (e.g., a positive US was paired with negative adjectives). Results of the subsequent re-rating phase (and a follow up one month later) showed that the ratings for the CSs paired with a US that had undergone successful revaluation, shifted in the same direction as the revaluated US.

However, not all experiments show that EC is sensitive to revaluation. Baeyens, Vanhouche, et al. (1998) failed to find post-acquisition US-inflation effects in a flavour-

taste paradigm. US-inflation is a variant of US-revaluation in which the degree of liking is changed (e.g., from moderately negative to strongly negative) rather than the evaluative sign of the US (e.g., from negative to neutral or positive). The authors hypothesized that the discrepancy between revaluation in the flavour-taste and picture-picture paradigms was caused by the differential sensory complexity of the USs in both paradigms. Baeyens and colleagues proposed that for the US-pictures more sensory-descriptive information is encoded than for the US-flavours, and that this sensory-descriptive information is necessary for revaluation to occur.

Sensory preconditioning

EC seems sensitive to sensory preconditioning, mirroring traditional PC. Sensory preconditioning refers to a procedure in which first two neutral stimuli, CS1-CS2, are associated with one another, after which one of the CSs is paired with a US, CS2-US. Typically, not only the CS2 elicits a conditioned response, but also the CS1 which had never been paired directly with the US. Hammerl and Grabitz (1996), demonstrated sensory preconditioning in EC using a picture-picture paradigm with pictures of sculptures and fountains. In the first phase two neutral pictures were paired with one another (CS1-CS2). Next, CS1 was repeatedly paired with a positive US. Results showed that both CS1 and CS2 had become more positive. This effect was later replicated and extended to negative USs by Walther (2002).

US pre- and post-exposure

US pre-exposure refers to a procedure in which a US is repeatedly presented *before* acquisition. In traditional PC this leads to a diminished conditioned reactions after acquisition. Hammerl, Bloch, and Silverthorne (1997) demonstrated that EC is also sensitive to US pre-exposure. In a picture-picture paradigm using pictures of sculptures and fountains, Hammerl and colleagues observed that repeated presentations of the US before acquisition reduced the evaluative shift of the CS after a subsequent conditioning. The authors observed the same reduction when the US-only presentations came *after*, as opposed to before, acquisition. Hammerl and colleagues hypothesized that the repeated US-only presentations resulted from a reduction of the evaluative value of the US. As such,

the post-acquisition US-only presentations would result in an implicit US-revaluation procedure (in which the US was made less extreme after acquisition).

Models of EC

The apparent functional differences between EC and traditional PC suggest that they depend on different processes. Over the years, a number of different models have been proposed to account for EC¹. Most of these models, however, were constructed to explain a specific set of data. As a consequence they are frequently not detailed enough to make predictions about other functional differences. In this section we will discuss these models and the extent to which they can explain the functional characteristics of EC.

Intrinsic change account

A first possible explanation of EC and its apparent resistance to extinction is that EC is an intrinsic form of learning (e.g., Rozin & Zellner, 1985; Zellner et al., 1983). Note that this account is not really a specific theory, but more a general notion of what happens in preference learning or EC. In intrinsic learning, acquisition results in the representation of the CS itself being endowed with the positive or negative valence. This approach comes very close to the S-R learning: after acquisition the CS directly activates the conditional response (i.e., the valence of the US), bypassing the activation of the US itself. Baeyens, Eelen, and Crombez, (1995) argued that if the CS acquired the evaluative meaning of the US without reference to the US, it can be expected that extinction leaves the acquired valence of the US intact. This would happen because after acquisition each presentation of the CS is in fact the presentation of a positive/negative CS, and as such the absence of the US does not disconfirm the acquired valence of the US.

The intrinsic account is incompatible with some of the results found in EC research. It cannot explain the sensitivity of EC to post-acquisition US-revaluation (Baeyens, Eelen, Van den Bergh, & Crombez, 1992). In this human faces picture-picture EC experiment it was demonstrated that if the valence of the US is somehow changed after acquisition, the

¹ Note that not all researchers believe that EC and traditional PC reflect qualitatively distinct learning mechanisms (see for example Lipp & Purkis, in press).

valence of the CS will track the changes of the US with which it had been paired. This cannot be explained by the intrinsic change account: if EC had resulted in an intrinsic change of the valence of the CS, then subsequently changing the valence of the US should not influence the CS. The intrinsic change model can also not account for sensory preconditioning effects. Hammerl and Grabitz (1996; see also Walther, 2002) demonstrated that when two neutral CS-pictures, CS1 and CS2, are associated after which the CS2 is subsequently associated with a positive US-picture, that not only the CS2 becomes more positive but also the CS1. As the CS1 had never been paired with a positive US, this cannot be explained by intrinsic change.

Holistic representation account

Martin and Levey, the parents of modern EC research, see the transfer of valence as the core mechanism of associative learning (Martin & Levey, 1978, 1987, 1994). These learning psychologist claim that the contiguous presentations of a CS and a (salient) US result in a representation of a CS/US stimulus complex in which both stimulus elements are integrated. The formation of this CS/US stimulus complex is a basic and automatic process. When either of the stimulus elements subsequently occurs, it activates the CS/US complex and thereby also the integrated stimulus-characteristics of the US, including the US-valence (Martin & Levey, 1978). Although this model seems somewhat similar to the intrinsic change account, there are differences. While the intrinsic account emphasises the integration of the US-valence, the CS/US complex of the holistic account includes also more sensory-descriptive properties of the US. Second, in the holistic model the presentation of the US will also activate the CS/US complex, whereas the intrinsic account would not expect the US to activate the (elements of the) CS.

The holistic account can explain resistance to extinction: After acquisition the presentation of the CS activates the CS/US complex and thus also the US-valence. Therefore, the omission of the US should not influence the CS-valence. It can also explain the results from post-acquisition US-revaluation: because after acquisition the US also activates the CS/US complex, the new information about the valence of the US is integrated into the CS/US stimulus complex. The CS will subsequently activate this modified CS/US complex, and thus its new valence (Martin & Levey, 1994).

However, the holistic account has difficulties explaining the results of sensory preconditioning studies. As in sensory preconditioning (i.e., CS1-CS2, then CS2-US) the CS1 has never been paired with the valenced US, there would be no CS/US complex for that stimulus and thus no transfer of valence. One could, however, expand the theory and claim that when two neutral CSs are paired, a CS1/CS2 complex is constructed. When the CS2 is subsequently paired with the US, the CS1/CS2 complex (activated by presentation of the CS2) will be expanded with the US, resulting in a CS1/CS2/US complex. This expansion would allow the CS1 to activate the valence of the US through the activated CS1/CS2/US-complex.

Conceptual categorization account

Davey (1994) has proposed a very different and more general explanation for EC. He suggests that EC does not rely on associative relations between stimuli, but might be based on some kind of conceptual transfer. Davey's model argues that stimuli consist of features which by themselves are not necessary or sufficient to make the stimulus as a whole liked or disliked. But when a stimulus has more (dis)liked components, it will be more (dis)liked. Affectively neutral stimuli contain a number of liked, but also a number of disliked features. Davey hypothesised that by pairing a neutral CS with a (valenced) US, the features that the CS and US have in common will be made more salient. Because a (dis)liked US contains more (dis)liked features, the features that consequentially become salient are especially (dis)liked features. This will result in a change in the balance of the CS-evaluation and the CS will be re-categorized as more (dis)liked. Take for example a neutral CS face containing the liked features blond hair and blue eyes, and disliked features large nose and short hair. This face is repeatedly paired with a liked US face containing only positive features, namely blond hair, blue eyes, a fine nose and long hair. As a consequence of the pairings, the shared features between CS and US become more salient (i.e., the blond hair and blue eyes), and the CS is re-categorized as liked based on these newly salient features. This model is not a pure associative model, because the shift in valence does not rely on an association between the CS and the US. The evaluative change of the CS occurs as a consequence of pairing the CS with a US, but not necessarily through an associative link between CS and US (Field & Davey, 1997).

Resistance to extinction of evaluative transfer is in line with this model: presenting the CS by itself will not differentially influence the salience of its features, thus leaving salient the features that are congruent with the previously paired US. However, the conceptual categorization model also has shortcomings. First, although it can explain some examples of post-acquisition US revaluation, it has difficulty explaining the selective nature of post-acquisition US-revaluation. In a US-revaluation experiment, Baeyens, Eelen, Van den Bergh, and Crombez, (1992) presented several pairs during acquisition, including two neutral-liked pairs, CS1-US1 and CS2-US2. Subsequently the valence of the US2 was revaluated towards neutral. This resulted in the CS2 becoming more neutral but did not influence the liking of CS1. While Davey's model could explain the shift of the CS2 by claiming that the (features of the) US2 became less liked, and that some of these features are shared with (and made salient in) the CS2, it cannot explain why the CS1 was unaffected. As the criteria for liked pictures have changed, the valence of the US1 and CS1 should also have changed to some extent, as they share some of the liked features that have been revalued. Second, the conceptual re-categorization account also has problems explaining sensory preconditioning. Even if one assumes that pairing the two neutral stimuli, CS1 and CS2, will cause the shared features to become salient, (on average) as many liked as disliked features will be shared and made salient. The subsequent pairing of the CS2 and US will not change the salience of the features of the CS1, and thus no evaluative shift of CS1 can be expected. Finally, it is not clear how the model would explain the cross-modal EC experiments: it is unlikely that there are enough shared features between stimuli of different modalities for a re-categorization to occur.

Artefactual account

In most of the early EC picture-picture experiments (e.g., Baeyens et al., 1988; Martin & Levey, 1978) the assignment of CSs to USs was not determined randomly but was based on perceptual similarity. It was believed that similarity between CS and US would enhance the valence transfer. Field and Davey (1997, 1998, 1999) argued that this pairing procedure could induce a non-associative, artefactual EC-effect. They suggested that combining the neutral CS with a valenced US based on similarity would result in the valence of the CS shifting in the direction of the US, even when these stimuli are *never* presented together. It is possible, for example, that during the baseline rating phase, participants rate each of the pictures based on a general liked/disliked framework, without knowing what is the valence

range of the pictures in the set. During this rating phase, participants start to create a context-specific liked/disliked framework for this set of stimuli, defining the most liked/disliked pictures as exemplars/anchors. When rating the pictures again (e.g., after acquisition), some of the pictures that were initially rated as neutral will be evaluated as more liked or disliked because the exemplars/anchors have changed. The initially neutral CSs that are similar to a positive US may be re-rated as more positively, while the CSs that are similar to a negative US may be re-rated more negatively (Field & Davey, 1999; see also De Houwer et al., 2000). Field and Davey (1999) provided some experimental support for this artificial account in a picture-picture EC-experiment in which they only found ‘valence transfer’ when the CS-US pairs were rated as perceptually similar (compared to when the CS-US pairs were rated dissimilar). Moreover, the size of the ‘transfer effect’ found for CS-US pairs (that were rated perceptually similar), was the same regardless of whether the CS-US pairs were actually presented.

Field and Davey’s artefactual account however, can only explain effects in experiments where the CS-US pairs are constructed on the basis of perceptual similarity. It is thus only valid for a limited number of experiments, that is, most of Martin and Levey’s experiments (e.g., Martin & Levey, 1978) and some of the early picture-picture experiments of Baeyens and colleagues (Baeyens et al., 1988, 1989a; Baeyens, Eelen, & Van den Bergh, 1990). In most other experiments the CSs, are randomly assigned to the USs or the assignments were counterbalanced. This was the case in the more recent picture-picture experiments of Baeyens and colleagues (Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Baeyens et al., 1989b; Baeyens, Eelen, Van den Bergh, & Crombez, 1992; Baeyens, Hermans, & Eelen, 1993; De Houwer et al., 2000), the flavour-taste experiments, the haptic EC experiments and in observational EC. The artificial account can also not explain cross-modal EC experiments (including experiments with biologically significant stimuli), as it is unclear how CSs and USs can be paired based on similarity if they are from different modalities. Furthermore, although this account can explain resistance to extinction, it cannot explain all of the functional characteristics of EC (e.g., it can not explain sensory preconditioning, the selective nature of post-acquisition US-revaluation, or sensitivity to counter-conditioning).

Emotional-affective versus Sensory-descriptive model

Baeyens, Eelen, and Crombez (1995) formulated another theory to explain resistance to extinction. Several researchers claim that a US consists of multiple levels of representation, that can each form a separate association with a CS. Wagner and Brandon (1989), for example, distinguished between the emotional-affective and sensory-descriptive components of a US (see also Mackintosh, 1983). Both levels of representation differ in their specificity and in the kind of responses they produce. The emotional-affective component produces a more general and diffuse emotional state, and can be shared by different similar USs. The sensory-descriptive component produces more specific stimulus-elements and a more discrete response. Baeyens, Eelen and Crombez (1995) hypothesized that the conditional responses in EC reflect the association of the CS with the emotional-affective component of the US, and that associations with the emotional-affective component are resistant to extinction. This could explain the resistance to extinction of EC. In the same article, however, Baeyens, Eelen, and Crombez argue that this model has shortcomings. The authors argue that results from US-revaluation studies in EC demonstrate that it is untenable to claim that *only* the association between the CS and the emotional-affective component of the US is responsible for EC: As the CS tracks the valence of the US, there also must be an association with the sensory-descriptive characteristics of the US. If there was only an association between the CS and the emotional-affective component of the US, the organism could not unambiguously relate the changed valence of that particular US to the emotional-affective component.

However, if one assumes that during EC (sometimes) not only a link with the emotional-affective component is made, but also with the sensory-descriptive component of the US, then some forms of sensitivity to US-revaluation in EC can be explained. The revaluation of a US could, in this case, be linked to the valence (the emotional-affective component) of the US-representation that had been associated with the CS². With the same

² Baeyens and colleagues (Baeyens, Vanhouche et al., 1998; Baeyens, Vansteenwegen, et al., 2001a) suggest that sensory-descriptive information is sometimes, but not always, encoded in EC. While changing the valence of the US after acquisition seems to work in the picture-picture paradigm (Baeyens, Eelen, Van den Bergh, & Crombez, 1992), it was not found in the flavour-taste paradigm. Baeyens and colleagues hypothesized that whether or not the sensory-descriptive component is associated might depend on the specific nature of the US: if the US has a more elaborate sensory-descriptive representation (e.g., a picture of

assumption one could also explain the sensitivity of EC to sensory-preconditioning: if pairing two neutral stimuli, CS1-CS2, creates (also) an association on the sensory-descriptive level, subsequent pairings of CS2 with a valenced US could be tracked by the CS1.

Referential versus Expectancy system

Baeyens and colleagues (Baeyens, Eelen, & Crombez, 1995; Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Baeyens & De Houwer, 1995) proposed that at a phenomenological level there are two different types of CS-US relations. For some types of CS-US associations the CS will cause the active expectancy that the US will really occur in the near future. This type of association can be described as a genuine expectancy relation. For example, a bell might lead to the active expectancy that food is going to be presented in a few seconds. For other CS-US relations however, perceiving the CS will merely make one think –consciously or unconsciously- of the US, without additionally expecting the US to occur here-and-now. This type of relation can be described as a merely referential relation. For example, seeing a cigar of the brand your late grandfather always used to smoke, may make you think of your grandfather, without necessarily expecting him to appear in front of you. Most forms of traditional PC can be categorized as using expectancy relations, while cases of EC can be categorized as using merely referential relations. Note that both types of relations are not mutually exclusive, but rather hierarchical: an expectancy relation requires a referential relation, but a referential relation may exist without an expectancy relation.

According to the authors (e.g., Baeyens, Eelen, & Crombez, 1995; for more recent descriptions see Baeyens, Vansteenwegen, et al., 2001a; De Houwer, Thomas, & Baeyens, 2001) these two different types of relations are the reflection of two different learning systems: an Expectancy System (ES) and a merely Referential System (RS). The ES is thought to be responsible for the detection of reliable predictors of (biologically) significant events with the purpose of preparing the organism as well as possible for the impending US. In this system, the CS leads to the expectation that the US will actually

a boy with short hair, glasses, blue eyes, straight eyebrows,...) the association between the CS and US is more likely to include the sensory-descriptive component of the US than when the emotional-affective component dominates the US (e.g., a disgusting taste).

occur. It triggers an active preparatory conditional response that is either appetitive, defensive or orienting. As this response is relatively costly for the organism, the system normally only responds to stimuli that reliably and non-redundantly predict the appearance of the US. This is reflected in the sensitivity of the system to extinction, CS-US contingency and modulation. To be able to perform this task, the system must put a relatively high load on information-processing resources, typically requiring attention and resulting in awareness of the CS-US contingencies. The RS, on the other hand, is thought to be a less complex system, using only information about the co-occurrences of stimuli to shape general liking/disliking of the CS. In this system, the CS also activates the US representation but not in such a way that the US is expected to occur here-and-now. The activation of the US only changes the perception of the CS: the activated US-valence is 'misattributed' to the CS. This activation of the US-valence only becomes apparent in situations with low (differential) response cost, like expressing liking or choosing between two stimuli. Because this response is relatively un-costly for the organism, reliably and non-redundancy are less important. Also, because performing this tasks requires less load than the expectancy system, awareness of the CS-US contingency is not required.

This model can explain many functional characteristics of EC. As the referential system only processes the co-occurrences of CS and US, it can explain resistance to extinction: presenting a CS by itself does not reduce the CS-US referential relation. Counter-conditioning results in new associations of the CS with another US of the opposite valence, which influences the balance of the valenced events activated by the CS. Susceptibility to post-acquisition US-revaluation can also be explained, as the valence of the CS still depends on its (merely referential) relation with the US. If the US-valence changes, the CS can still track this change. Sensory preconditioning can be explained, as pairing two neutral stimuli, CS1-CS2, will create at least a referential relation. The subsequently created CS2-US association, including the US-valence, can be activated by presenting the CS1. The observation that US pre-exposure trials diminish the EC-effect can be explained by claiming that these pre-exposures diminish the salience of the US(-valence).

This ES/RS model is a functional model that does not specify how or by what processes both systems fulfil their functions. Baeyens (1998) and De Houwer (1998a; see also De Houwer, Thomas, & Baeyens, 2001) proposed different ways to implement the

model. One approach is to put the difference between the two systems at the encoding end of the process: ES and RS both have their own encoding system and store the information in different knowledge bases. Associations in the ES could be formed using a Hebbian learning rule: only co-occurrences between stimuli create or influence the CS-US association. CS- or US-only presentations do not influence the knowledge base of the ES. In the expectancy system a more complex learning rule would be needed, that updates the CS-US associative strength based on information about co-occurrences and separate presentations of CS and US. In other words, the contingency between CS and US would determine the associative strength. The second approach puts the difference between ES and RS on the performance end of the process: both systems use the same learning rules and knowledge base but differ in the way they translate this information into behaviour. In this approach, the knowledge base contains information about co-occurrences and separate presentations of the CS and the US. The RS however only uses the information of the USs that co-occurred with a CS, while the ES compares both the information of co-occurrences and the separate presentation of the CS or US to calculate whether the CS is a reliable predictor of the US.

Whatever the implementation of the ES/RS might be, Baeyens, Vansteenwegen, et al. (2001a) stress that both systems are simultaneously active to help the organism deal with stimulus relationships in his environment.

Chapter 2

Beyond Evaluative Conditioning?

In the previous chapter we discussed EC and the theories that tried to explain this phenomenon. One of the more promising theoretical accounts for EC, the referential learning account (e.g., Baeyens, Eelen, & Crombez, 1995), claimed that in EC a merely referential relation is established between the CS and the US representations. As a consequence, post-acquisition presentations of the CS will activate the representation of the US. The valence transfer is caused by ‘misattributing’ the activated US-valence to the CS. In this theory, however, there is no clear reason why only the valence would be transferred to the CS. If EC is caused by misattributing the US-valence to the CS, than maybe other properties of the US can also be misattributed to the CS.

Not only the referential learning account suggests the existence of transfer of non-evaluative stimulus properties. The conceptual categorization account (Davey, 1994) argued that by pairing a CS with a US, the features that both stimuli have in common will be made salient. As explained in the previous chapter, the valence of the CS will therefore subsequently shift in the direction of the US-valence. Field and Davey (1997) pointed out that such an account suggests that not only evaluative but also other, non-evaluative properties can be transferred by pairing two stimuli. For instance, if one pairs the picture of a baby whose gender is difficult to determine with certainty (e.g., because he/she possesses both male and female features) with the picture of a baby who is clearly perceived as male, this could increase the salience of the male features of the first baby and thus increase the likelihood that the baby is categorised as male (but remember that the conceptual categorization account cannot explain all functional characteristics of EC, see previous chapter for more details). Moreover, most other accounts for EC that we discussed in the previous chapter do not, in se, exclude the possibility that non-evaluative properties can be transferred.

The search for this Associative Transfer of *Non-Evaluative* Stimulus Properties (ATNESP) forms the central theme of this dissertation. If ATNESP exists, it would not only give some insights into the mechanism that drives EC, but it could also be an

important process that determines the meaning and perception of various kinds of stimuli. In the next paragraph we will look at the existing evidence for ATNESP. We will end this chapter by specifying our research objectives.

Existing evidence for ATNESP

There is little existing evidence for general ATNESP. The most direct evidence for non-evaluative transfer comes from studies conducted by Stevenson and colleagues. In these experiments the authors used an odour-taste paradigm, in which they associatively transferred the sweetness or sourness of a beverage onto an odour (for an overview, see Stevenson and Boakes, *in press*). A prototypical experiment started with a baseline-rating phase, in which participants had to assess how sweet and how sour a series of odours smelled. During a subsequent acquisition phase, participants sampled sweet sucrose solutions (US) containing a CS-odour (e.g., lychee) and sour citric-acid solutions (US) containing another CS-odour (e.g., chestnut). The prototypical result was that in the post-acquisition ratings in which participants had to smell the odours from a scent bottle (without tasting it), the CS-odour that had been paired with a sweet taste was rated sweeter and that the CS-odour that had been paired with a sour taste was rated sourer (e.g., Stevenson et al., 1995, 1998, 2000a, 2000b; Stevenson & Case, 2003)³.

Note that one could argue that this sweetness/sourness transfer was mediated by EC, and that there was no ‘real’ non-evaluative property transfer. Generally a sweet taste is liked better than a sour taste. As a consequence, pairing odours with sweet/sour beverages, might result in EC: the CS-odour that was paired with the sweet taste might become more positive than the CS-odour that was paired with a sour taste. If participants subsequently have to rate the sweetness/sourness of the odour, they might base these judgements on the perceived valence of the odour (e.g., ‘I like this smell, so it must be sweet’). In the

³ In a similar paradigm, Stevenson and colleagues (Stevenson, 2001a, 2001b; Stevenson, Case, & Boakes, 2003) recently observed similar transfer-like effects in an odour-odour paradigm, in which pairing odours made the odours smell more like one another. For example, the flavours L-carvone (minty) and cis-3-hexanol (freshly cut grass) were found to smell mustier after being repeatedly experienced in a mixture with the musty-smelling odorant p-anisaldehyde. In addition, an odour that had been paired with the musty p-anisaldehyde odour, was subsequently judged as more similar to it (than the odour that had not been paired with it). This odour-odour learning seems to behave similarly as odour-taste learning.

experiments of Stevenson and colleagues, however, it is unlikely that the changes in sweetness/sourness occurred as a consequence of changes in liking. In the majority of their experiments they found no significant difference in liking between the sucrose-paired and citric-paired odours (e.g., Stevenson et al., 1995, 1998). Stevenson et al. (2000b) further argued that it is unlikely that changes in liking can cause perceptual chemosensory changes, referring to Cabanac (as cited in Stevenson et al., 2000b) who demonstrated that after a sucrose preload, sucrose solutions tasted less pleasant but equally sweet. It is more likely that perceptual changes result in changes in liking. Stevenson et al. (2000b) concluded that it is very unlikely that the sweetness/sourness transfer observed in their experiments were (completely) mediated by changes in liking.

The non-evaluative transfer that Stevenson and colleagues observed, appeared to have the same functional properties as EC. First, awareness of the CS-US contingencies appeared to be unnecessary for the transfer to occur (Stevenson et al., 1995, 1998). Second, the sweetness/sourness transfer proved to be resistant to an extinction procedure in which, after acquisition, the CS-odour was repeatedly presented in plain water (Stevenson et al., 2000b). However, data suggest that sweetness/sourness transfer is even more resistant to change than the valence transfer of EC: while EC seems to be sensitive to counter-conditioning (see Baeyens et al., 1989a), Stevenson et al. (2000a) failed to find an effect of counter-conditioning in sweetness/sourness transfer. In this experiment, Stevenson and colleagues found no difference in the sweetness/sourness transfer between a CS-flavour that had been paired with citric acid and another CS-flavour that had first been paired with citric acid, and subsequently with a sucrose solution. However, more research is needed to confirm that there really is a qualitative difference between EC and sweetness/sourness transfer concerning sensitivity to counter-conditioning.

These experiments of Stevenson and colleagues could be seen as evidence for ATNESP: by repeatedly pairing a CS-odour with a US-taste, the sweetness/sourness of the US transfers to the CS. Stevenson et al. (2000a, 2000b), however, seem to suggest that the observed property transfer in their odour-taste experiments might not be a general phenomenon: odours and tastes are sensory very close to one another, and when both are presented together they might be encoded holistically. As subsequent presentations of the odour by itself will also activate the taste that was encoded with it, this would explain why the taste-transfer is resistant to extinction. It remains unclear whether this form of

ATNESP would also happen using non olfactory-gustatory stimuli. It is possible that the holistic encoding only occurs when there is perceptual overlap between the CS and US, similar to the sensory overlap between taste and odour. More research is necessary to investigate whether the holistic encoding will also occur with other non-evaluative stimulus properties, outside of the olfactory-gustatory domain.

Other studies that show ATNESP are sparse. Some evidence comes from the marketing literature. In one experiment, Kim, Allen, and Kardes (1996) paired a picture of a pizza-box of a non-existing brand (CS) with a picture of a (affectively neutral) fast racing car (US). In the experimental condition, this crucial pair was presented 10 times intermixed with 30 fillers. In a control condition all the pictures were presented at random, with the exception that the crucial CS was never paired with the US. When asked how long it would take to deliver pizzas from the CS-brand, participants assessed the pizza-brand to deliver faster in the experimental condition than in a control-condition. This could be seen as transfer of the ‘speed’ of the US (car) onto the CS (pizza). But one third of the participants were able to describe the research as some sort of associative learning study and might thus have been demand aware. No subsequent analysis of the impact of demand awareness on transfer was reported. In a similar second experiment, Kim et al. (1996) also succeeded in transferring the ‘softness’ of a picture of kittens (US) onto a picture of a box of tissues (CS). In this experiment, no information about demand-awareness was given. Both experiments suggest ATNESP, but no hard conclusions can be drawn because the observed effects might have been an artefact caused by demand effects.

To summarize, although there exists clear evidence for ATNESP in the olfactory-gustatory domain, it is unclear whether ATNESP is a general phenomenon that exists in a multitude of modalities.

Research objectives

The existing evidence for general (i.e., not restricted to certain modalities or special cases) ATNESP is sparse. Clear evidence has been found only in the olfactory-gustatory domain: in the odour-taste studies the sweetness/sourness of a taste was transferred to a paired odour (e.g., Stevenson et al., 1995, 1998, 2000a, 2000b; Stevenson & Case, 2003). Although the evidence strongly suggests that it is ATNESP, this type of transfer might be

restricted to olfactory-gustatory stimuli (as taste and odour are sensory very close to one another and might induce configural or holistic encoding). In the picture-picture experiments from Kim et al. (1996), non-gustatory stimuli were used. The authors observed a transfer of the ‘speed’ and ‘softness’ of US-pictures onto CS-pictures. However, because they were unable to exclude demand effects, no clear conclusion can be drawn.

The search for ATNESP outside the olfactory-gustatory domain formed the central objective of this dissertation. If property-transfer was found, it was important to exclude that it was caused by some form of demand awareness. The use of cover stories and indirect measures (that are less sensitive or insensitive to demand effects, like for example affective priming) could help to exclude such demand effects. In addition, we had to exclude the possibility that non-evaluative property transfer was caused and fully mediated by evaluative transfer (see an example of this reasoning in the previous section where we discussed the odour-taste experiments of Stevenson and colleagues). For this purpose it was important to choose properties that were not perfectly correlated with valence. In addition, measuring the valence of the CSs and USs could help in determining if EC occurred.

If we would find clear evidence for ATNESP, we planned to subsequently investigate the functional characteristics of this phenomenon. Most importantly, we would investigate to what extent the non-evaluative transfer behaves similarly as EC (e.g., resistant to extinction, independent of contingency awareness,...). An overview of the performed experiments can be found in the Appendix.

Chapter 3

Picture-Picture Studies

The first set of experiments in which we tried to find ATNESP was based on the picture-picture EC paradigm as implemented by Baeyens and colleagues (e.g., Baeyens, Hermans, & Eelen, 1993). In most of these studies, participants were first asked to express their liking for a series of pictures of human faces. On the basis of these ratings, three types of pictures were selected: the most liked, most disliked and most neutral faces. The selected pictures were used to make three types of CS-US pairs: neutral-liked, neutral-disliked and neutral-neutral pairs. After repeated presentations of these pairs, the valence of the previously neutral CS-picture had shifted in the direction of the valence of the US with which it had been paired. In our search for associative transfer of non-evaluative stimulus properties, we used a similar procedure but focussed on other, non-evaluative, properties. This paradigm was chosen because a procedure that yields EC might also be most suitable for transfer of non-evaluative properties. To select the properties that we would try to transfer, we used three criteria. First, there must exist stimuli that are somewhat ambiguous concerning that property. If not, it might be hard to associatively influence the perception of this property. Second, the property must be more or less biologically significant. It might be that only more significant properties can be associatively transferred. Note that valence, the property transferred in EC, is probably one of the most significant properties of a stimuli. Third, we preferred a property that lies on a dimension that is not strongly correlated with valence. If there is a high correlation between the property and valence, then the associative transfer of a non-evaluative property might be mediated by evaluative transfer. The properties we selected were age and gender. For both properties we can find somewhat ambiguous stimuli: for gender we used pictures of babies of which the gender was hard to discern, and for age we used pictures of middle aged men, for which it is hard to give an exact age estimate. Both are relatively important properties of people that determine how we interact with them. Neither of the two properties is strongly correlated with valence.

Experiment P1

In a first experiment, we tried to influence judgements about the age of middle-aged people (CSs) by pairing pictures of those persons with pictures of young or old people (USs). We predicted that after repeated presentations of these pairs, the assessed age of the middle-aged people would shift in the direction of the age of the US with which it had been paired.

Method

Participants.

Twenty-one first year psychology students (18 men and 3 women) took part in the experiment. Note that, unless indicated otherwise, in all experiments participants were Dutch-speaking psychology students at the University of Leuven who participated in exchange for course credit.

Stimuli and measures.

Fifty-seven colour pictures of men of different ages (face and shoulders, approximately 25-85 years old) were used as stimuli⁴. They were digitalized in such a manner that they were about 12 cm high and 10 cm wide on the computer screen. To measure participants' assessment of the age of the depicted men, we used an age-scale (from 1 to 100 years) that was presented under each picture. The experiment was controlled by a custom made Java-program. Note that, unless indicated otherwise, all instructions and verbal stimuli in all the experiments reported in this dissertation were in Dutch.

Procedure.

As a cover-story, participants were told that we had reasons to believe that depressed people look differently at old-age than non-depressed people, and that the aim of this study was to examine whether this would also become apparent in skin-conductance measures. During acquisition, participants were instructed to just look at the pictures while a skin-conductance device (in reality a heart beat measure device in which one can slide a finger) would do the rest.

⁴ These pictures were taken by Hilde Hendrickx, and used in a pilot experiment (Hendrickx, 1999) that was similar in design to Experiment 1.

At the beginning of the experiment, participants were asked to assess the age of the depicted men using the age-scale. On the basis of these scores, three kinds of stimuli were selected for each participant: the youngest, the eldest and the ones closest to the individually determined mean age (which we will call ‘neutral’). These were used to make nine CS-US pairs: three neutral-old pairs, three neutral-young pairs, and three neutral-neutral control pairs. The assignment of the CS-US pairs was as follows: First, the six most neutral pictures were randomly paired with the eldest and youngest pictures to form the neutral-old pairs and neutral-young pairs respectively. Afterwards, the three next most neutral pictures were randomly paired with the following three next most neutral pictures to form the neutral-neutral control pairs. This assignment-procedure was used to ensure that the crucial neutral-old pairs and neutral-young pairs were formed with the most neutral pictures. As a naming convention, we will name the US pictures after the category to which they belong, that is, US-old, US-young and US-neutral. The CSs will be named after the US-category with which they are paired, so for the previous example this would be CS-old, CS-young and CS-neutral respectively.

During acquisition, each of the nine pairs was presented nine times. The CS was presented before the US, and both stimuli were shown for one second. The inter stimulus interval between CS and US (ISI) was 0 seconds, and the inter trial interval (ITI) 6 seconds. The order of the presentation of the pairs was randomised. After this conditioning phase, participants were asked to reassess the age of the 12 pictures used during acquisition on the age-scale. Participants were told that they should base this assessment on their current impressions.

The experiment ended with a contingency awareness test in which we checked to what extent participants recognized which CS had been paired with which US. If participants know which CS had been paired with which US, they *might* consciously use this knowledge as a basis for their post-acquisition CS-assessment. Hence, the awareness test could provide us with information about possible demand effects. This contingency awareness measure was implemented as follows: On the left third of the screen, a CS picture was presented. On the right side, six US pictures were presented: two US-old, two US-young and two US-neutral pictures.⁵ One of these was the US with which that CS had

⁵ Note that not all the US-pictures were presented because the screen was too small to present 12 pictures on the right side of the screen.

been paired during acquisition. Participants guessed, for each CS, which US had been paired with that CS. At the bottom of the screen, they also indicated how sure they were of their choice by checking one of the four checkboxes that labelled respectively *very uncertain*, *uncertain*, *certain* and *very certain*.

Results

Age-scale.

For all participants, the estimated ages for the three pictures of the same CS or same US categories were averaged for each moment of measurement separately. The average estimated age for these categories are presented in Table 1. To test our main prediction, we conducted a 2 x 2 ANOVA with CS category (i.e., CS-old versus CS-young) and moment (i.e., pre- versus post-acquisition) as within-subjects variables⁶. This analysis revealed no evidence for transfer of age: pairing a picture of an average aged man with a picture of a clearly older or younger man did not differentially influence the estimated age of the average aged men, $F(1, 20) < 1$.

Note that for all rating-measures in this dissertation in which a transfer-effect is expected, the effect size of the property transfer is calculated. To compute the effect size we calculate the extent to which CS-ratings changed in the direction of the US with which they were paired. For example, in this experiment we subtracted, for each participant, the age-shift of the CS-young pictures (from pre- to post acquisition) from the age-shift of the CS-old pictures (from pre- to post acquisition). With the resulting transfer effect score, ‘ Δ Δ rating’, we could calculate the effect size of the transfer as $\mu_{\Delta\Delta\text{-rating}} / \sigma_{\Delta\Delta\text{-rating}}$. The resulting Cohen’s d indicates how many standard deviations the CSs have differentially shifted in the direction of the US with which they had been paired. In other words, positive d values suggest a (assimilative) transfer effect, negative d values suggest a reverse transfer effect, and a zero value indicates no differential shift⁷. On the age-scale of this experiment,

⁶ Note that we do not include the CS-neutral pictures into the analysis. Our main hypotheses only predicts that acquisition has a differential effect on the CS-old and CS-young. If this hypotheses is supported, we can use the data from the CS-neutral as a baseline to explore whether the acquisition influences the CS-old, the CS-young or both.

⁷ Cohen (1988) proposed some heuristic value’s to help interpret the d statistic. A d of 0.20 can be considered a small effect (e.g., the difference in height between 15 and 16 year old girls), a d of .50 a medium sized

the effect size d was -0.17 . The transfer effect score (and corresponding d) was calculated in a similar way in all other experiments reported in this dissertation. Note that in the experiments without pre-acquisition rating phase, we simply dropped the pre-ratings variables from the calculation, and with the resulting ‘ Δ rating’ calculated the transfer effect score (and corresponding d) in the same way as in the other experiments. From now on, the resulting effect size d will be added after the p -value of the crucial F -test for the rating scale measure.

Table 1:

Average Age-Ratings Before and After Acquisition for the CS- and US-Categories from Experiment P1 (standard errors in parentheses)

CS-category	Age-Ratings		d
	Pre-acq.	Post-acq.	
CS-old	44.4 (0.6)	45.2 (0.6)	-0.17
CS-young	46.0 (0.9)	47.5 (1.2)	
CS-neutral	46.3 (1.1)	47.8 (1.5)	
US-old	81.7 (2.1)	81.4 (1.3)	
US-young	21.3 (1.4)	21.5 (0.6)	
US-neutral	42.7 (1.6)	42.9 (2.0)	

Note. d = effect size of the transfer

Contingency awareness.

The primary reason for adding the contingency-awareness test was to control for possible demand effects. If participants are aware of the contingencies between the CS and the US, they *might* consciously influence their post-acquisition CS-assessment (towards that of the US). The extent to which participants know which CS was paired with which US, will be called the strict contingency-awareness. But note that even with less information, participants might be able to differentially influence their ratings: even if participants only know with what kind of US a CS was paired (but not exactly with which young, old or middle-aged person), they might be able to induce a transfer-like-effect. The extent to which participants know which CS was paired with which US-*category*, will be called the category-awareness.

The most straightforward way to investigate strict contingency-awareness is to count the number of pairs for which participants selected the correct US. For analyses effect (e.g., the difference in height between 14 and 18 year old girls) and a d of .80 a large effect (e.g., the difference in height between 13 and 18 year old girls).

purpose, we assume each choice is independent of the other choices and that the number of correctly identified pairs is thus binomially distributed when participants guess. Given that there were 6 USs to choose from, we would expect $1/6^{\text{th}}$ of the 9 pairs to be reconstructed correctly if participants guess. These assumptions allow us to later calculate if participants answered better than by chance. To measure the category-awareness, we counted the number of pairs for which participants select a US from the correct category: e.g., for a CS-old, any (of the two presented) US-old picture would be a correct category answer. For analyses purpose we again assume independency of choices and that the number of correctly selected categories is binomially distributed. Given that there were 6 USs to choose from, we would expect the US-category of $2/6^{\text{th}}$ of the 9 pairs to be guessed correctly. We will report both the contingency- and category-awareness score. If subsequent analysis with contingency awareness are required, they will be primarily done with the category-awareness measure. Category-awareness is after all the minimal prerequisite to be able to consciously cause a transfer effect: one can only correctly influence ratings if one has conscious knowledge about the type of US with which the CS was paired.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 4.2 of the 9 pairs ($SE = 0.5$). 67 % of the participants selected the correct US-picture 4 or more times out of 9, thus performing significantly above chance ($p = .048$). The category-awareness measure showed that, on average, the correct US-category was selected for 5.2 of the 9 pairs ($SE = 0.4$). 48 % of the participants selected the correct US-category 6 or more times out of 9, thus performing significantly above chance ($p = .042$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis revealed no significant correlation, $r = -.05$, $p = .83$, between the category-awareness score and the transfer effect score on the gender-scale (the ‘ $\Delta \Delta$ rating’ described above).

Discussion

We found no evidence for associative transfer of age: pairing pictures of middle-aged people with clearly older/younger people had no differential effect on the perceived age of

middle-aged men. Participants were able to reconstruct 47% of the pairs, showing that they were partially aware of the contingencies. Being less or more aware of the contingencies did, however, not influence the results. Note that in a very similar pilot study ($N = 12$), Hendrickx (1999) was also unsuccessful in transferring age.

Experiment P2

Valence, the property transferred in EC, is a basic and easily activated stimulus-property (Zajonc, 1980) that might be very hard to rival in importance. Maybe age was not biologically ‘important’ enough to be a candidate for associative transfer. For our next experiments, we therefore choose another property to transfer. ‘Gender’ was chosen because, intuitively, we believed that gender is an important property of a person. In our next experiment(s), we tried to influence the gender-assessment of gender-ambiguous infants by repeatedly pairing them with pictures of clearly male or female infants. The procedure used was conceptually similar to the one in Experiment P1. We did not, however, include a pre-acquisition measure but decided to select the stimuli on the basis of a pilot study. This was done to suppress possible demand awareness: the presence of a pre-acquisition gender measure might focus participants on the (gender of) pairs during acquisition. From now on we also measured how much participants liked or disliked the stimuli: this allowed us to investigate the possibility that a gender transfer-effect (if any) might somehow have been mediated by an EC effect. For example, if one likes the male pictures more than the female pictures, an evaluative conditioning-effect is expected. Afterwards the baby that had been paired with a male-picture might thus have become more positive than the baby that had been paired with a female-picture. If one then asks participants to assess the gender of these babies, they might (intentionally or unintentionally) use the valence as basis for their gender-assessment (‘I like that picture, so it is probably a boy’).

Method

Participants.

Thirty-one first year psychology students (4 men and 27 women) participated in exchange for course credits.

Stimuli and measures.

A subset of fifty-eight colour pictures of infants (face and chest, 15 cm high and 10 cm wide on the computer screen) between 2 and 18 months old were used as stimuli for conditioning and testing. The CS and US pictures were selected from this set on the basis of a small rating study ($N = 20$) in which participants assessed the gender of these depicted children on a scale from *very certainly a boy/man* ($= 0$) to *very certainly a girl/women* ($= 100$). Twelve pictures received an average rating between 45 and 55 (and $SDs < 25$) on this scale. These 12 pictures were used as ‘gender-neutral’ stimuli (CSs and US-neutrals). The 3 highest and 3 lowest scoring pictures scored above 90 and below 10 (and $SDs < 15$) respectively and were used as US-females and US-males respectively.

To measure participant’s gender-assessment during this experiment, we used the scale from *very certainly a boy/man* ($= 0$) to *very certainly a girl/women* ($= 100$) that was presented beneath the pictures. In the same way we used a valence-scale from *very disliked* ($= 0$) to *very liked* ($= 100$) to measure linking of the pictures.

Procedure.

The cover story for this experiment was that men look differently at children than women, and that we were testing if this also affected people’s heart-rate when seeing infant pictures. During acquisition, participants were instructed to just look at the pictures while a heart-rate device (see Experiment P1) would do the rest.

The pre-selected 12 neutral, 3 clearly male and 3 clearly female pictures were used to generate nine CS-US pairs for acquisition: three neutral-male, three neutral-female, and three neutral-neutral pairs. For each participant, each CS was randomly assigned to a US. In accordance with our naming convention, we will name the US pictures after the category to which they belong, that is, US-male, US-female or US-neutral. Their corresponding CSs will be named CS-male, CS-female and CS-neutral respectively.

During acquisition, each of the nine pairs was presented nine times, using the same timing parameters as in Experiment P1. The order of the presentation of the pairs was randomised. After this conditioning phase, participants were asked to assess the age of the 12 pictures used during acquisition on the gender-scale. Next a contingency awareness test was administered, in which we checked to what extent participants recognized which CS

had been paired with which US. This procedure was identical to the one described in P1, except that the US-old and US-young pictures were now US-male and US-female pictures. The experiment ended with participants rating the 9 CS and 9 US pictures on the valence-scale.

Results

Gender-scale.

For all participants, the estimated gender for the three pictures of the same CS or same US categories were averaged. To make the results comparable to the results of the next transfer of gender experiments in which the scale was reversed (i.e., a score of 100 corresponded to *very certainly a boy/man*), we took the complement of the gender scores (i.e., $100 - x$, x being the raw gender-scores in this experiment). After this transformation, the gender-scale went from *very certainly a girl/women* (= 0) to *very certainly a boy/man* (= 100). The transformed average estimated gender for these categories are presented in Table 2. To test our main prediction, we conducted an ANOVA with CS category (i.e., CS-male versus CS-female) as within-subjects variable. This analysis revealed no evidence for transfer of gender: pairing a picture of a gender neutral infant with a picture of a clearly male or female infant did not differentially influence the estimated gender of the neutral infant, $F(1, 30) < 1$, effect size d was 0.10.

Table 2:

Average Gender-Ratings before and/or after acquisition for the CS-Categories from Experiments P2, P3, P4, P5, P6 and P8 (standard errors in parentheses)

Experiment	CS-category	Gender-Ratings		<i>d</i>
		Pre-acq.	Post-acq.	
P2	CS-male	–	57.9 (1.9)	0.10
	CS-female	–	56.6 (2.6)	
	CS-neutral	–	60.6 (1.8)	
	US-male	–	96.9 (1.0)	
	US-female	–	2.6 (1.1)	
	US-neutral	–	55.6 (2.1)	
P3	CS-male	57.8 (1.2)	60.4 (1.9)	0.52 **
	CS-female	57.4 (1.0)	54.1 (1.7)	
	CS-neutral	56.0 (1.1)	57.3 (2.4)	
	US-male	95.7 (1.2)	90.6 (2.2)	
	US-female	1.9 (0.7)	4.5 (1.8)	
	US-neutral	56.1 (1.5)	55.6 (2.0)	
P4	CS-male	50.0 (0.6)	51.5 (1.9)	0.14
	CS-female	51.0 (0.7)	50.1 (2.1)	
	CS-neutral	50.0 (0.7)	51.5 (2.4)	
	US-male	97.9 (0.6)	93.9 (1.3)	
	US-female	2.7 (0.9)	8.5 (2.1)	
	US-neutral	50.7 (1.2)	53.8 (1.7)	
P5	CS-male	57.4 (0.6)	58.8 (2.2)	0.12
	CS-female	58.2 (1.1)	57.0 (3.4)	
	CS-neutral	57.3 (1.1)	60.1 (1.9)	
	US-male	99.3 (0.4)	96.3 (1.0)	
	US-female	0.1 (0.1)	4.9 (2.2)	
	US-neutral	57.4 (1.5)	58.0 (2.0)	
P6	CS-male	50.5 (0.6)	51.9 (2.1)	-0.03
	CS-female	51.2 (0.6)	52.9 (2.3)	
	CS-neutral	50.2 (1.1)	57.8 (2.9)	
	US-male	95.8 (1.3)	95.6 (1.5)	
	US-female	2.4 (0.9)	2.5 (1.3)	
	US-neutral	52.0 (1.2)	55.6 (3.0)	
P8	CS-male	57.7 (2.4)	56.9 (2.1)	0.19
	CS-female	58.4 (2.0)	54.0 (1.7)	
	CS-neutral	-	-	
	US-male	92.7 (1.5)	93.7 (1.5)	
	US-female	4.1 (1.3)	4.4 (1.2)	
	US-neutral	-	-	

Note. *d* = effect size of the transfer

** $p < .01$

Contingency awareness.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 2.5 of the 9 pairs ($SE = 0.4$). 26 % of the participants selected the correct US-picture 4 or more times out of 9, thus performing significantly above chance ($p = .048$). The category-awareness measure showed that, on average, the correct US-category was selected for 3.9 of the 9 pairs ($SE = 0.4$). 16 % of the participants selected the correct US-category 6 or more times out of 9, thus performing significantly above chance ($p = .042$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis revealed no significant correlation, $-.11$, $p = .55$, between the category-awareness score and the transfer effect score on the gender-scale.

Valence scale.

The valence measure was included to investigate the possibility that a gender transfer-effect (if any) might have been mediated by an evaluative learning effect. For evaluative conditioning to have occurred two effects must be found. First, there must be a significant evaluative difference between the US-male and the US-female. Second, after acquisition, there must be a significant evaluative difference between the CS-male and the CS-female. If both effects are found we have evidence for EC and further analysis and research must reveal whether the non-evaluative transfer (if any) was mediated by the EC-effect.

There was a significant valence-difference between the US-male and the US-female: participants liked the US-male pictures ($M = 48.7$) less than the US-female pictures ($M = 61.9$), $F(1, 30) = 25.90$, $p < 0.001$. However, there was no difference between the liking of the CS-male ($M = 64.6$) and CS-female pictures ($M = 62.0$), $F(1, 30) < 1$. As such we have no evidence that evaluative conditioning could have influenced the (absent) transfer of gender.

Discussion

We found no evidence for associative transfer of gender: pairing pictures of gender-neutral infants with clearly male/female infants had no differential effect on the neutral infants. Awareness of the contingencies did not influence the effect.

One of the weaknesses of this experiment might be that the 12 gender-neutral pictures were selected on a priori basis (i.e., the rating study). Although these pictures were on average neutral, there was a lot of variance (average *SD* of 17 on the 101-point scale). For quite a few participants some of the CS-neutrals were probably somewhat male- or female-like. This could have been responsible for the absence of transfer if associative transfer of stimulus properties works only with neutral stimuli. It is not unlikely that the perceived gender of infants that are already perceived as somewhat more male or female, is harder to influence. In the next experiment, we tried to resolve this weakness by using an individual pre-rating to select the neutral pictures.

Experiment P3

Three important alterations were made on the design of Experiment P2. First, the pictures (CSs and USs) were now selected on the basis of an individual pre-measure instead of being selected from the rating study. Although this might increase the chance of demand effects (because participants focussed on gender before acquisition), we wanted to ensure that the gender-neutral pictures were really neutral for each participant. An additional advantage of this pre-measure was that we could track changes in the CS-ratings from pre- to post-acquisition. This allowed us to correct for baseline differences and thus increase the power of our analyses.

Second, because the gender-scale measure is very transparent and thus vulnerable to demand-effects, we developed a more indirect gender measure. In this experiment, we added a gender priming procedure (similar to an affective priming procedure⁸, e.g. Fazio et al., 1986). In this priming procedure, target-pictures were presented that showed either a male or female face. The task was to determine the gender of target pictures as quickly as

⁸ In affective priming studies a variant of the standard sequential priming procedure is used, in which the affective relation between primes and targets is manipulated. In the standard procedure introduced by Fazio et al. (1986), a positive or negative noun was presented as prime for 200 msec on a computer monitor, and was followed by a 100-msec blank screen after which the target was presented. Participants were asked to indicate, as fast as possible, whether the target-adjective had a positive or a negative valence by pressing one of the two response-keys. Fazio et al. (1986) observed shorter reaction times when the valence of the prime and the target was identical (i.e., positive-positive or negative-negative) compared to when the valence of the prime and the target differed (i.e. positive-negative and negative-positive).

possible by pressing a left or right key. A prime-picture was presented 300 ms before each target. In a pilot study, we used pictures of adult men and women as primes and targets, and observed that when the gender of the prime and target picture were the same (congruent trials), participants were faster to classify the gender of the target picture than when the gender of prime and target differed (incongruent trials). This suggests that this priming procedure can be used to measure the gender-perception of the primes: If we find that, after a certain prime-picture, participants are faster to classify male targets than female targets, and for another prime-picture the opposite effect is obtained, then we have evidence that the first prime is seen as more male than the second prime (see De Houwer, Hermans, & Eelen, 1998; Hermans, Vansteenwegen, et al., 2002, for studies in which priming tasks were also used as an indirect measure of associative learning). In the current transfer of gender experiment, the CSs were used as primes.

Third, the scale-measure was reversed: while in the previous experiment the gender-scale ran from *very certainly a boy/man* (= 0) to *very certainly a girl/women* (= 100), it now ran from *very certainly a girl/women* (= 0) to *very certainly a boy/man* (= 100). This was done because in the pilot gender-priming experiment described above, there was a marginal significant interaction between the key that participants had to press for a male/female target and the congruency affect: the congruency-effect tended to be larger when the male response key was on the right and the female response key on the left, then when this was the other way around. This could indicate that male is somewhat associated with a right response, and female with a left response. In that case, it is prudent to also put the male/female anchor points on those sides of a rating scale.

Note that an overview of the most important differences between all the transfer of gender experiments of this chapter, can be found in Table 3 (p. 43).

Method

Participants.

Thirty-seven second year psychology students participated (7 men and 30 women) in exchange for course credits.

Stimuli and measures.

The set of fifty-three colour pictures of infants (face and chest, between 2 and 18 months old) were used as stimuli for conditioning and ratings. To measure participant's gender-assessment before acquisition, we used a scale from *very certainly a girl/women* (= 0) to *very certainly a boy/man* (= 100). Due to programming limitations, the post measure was on a 21-point scale instead of a 101-point scale. To measure liking of the selected CSs and USs, a valence-measure from 0 (*very disliked*) to 20 (*very liked*) was used.

A separate set of four relatively gender-neutral (as determined by the pilot study of Experiment P2) infant-pictures were used as practice primes in the priming procedure. Four adult male and four adult female pictures were used as targets for the priming measure. Another eight were used as practice targets for the priming measure. In the gender-priming measure, the 9 CSs were used as primes and combined with four target pictures of adult men and four target pictures of adult women. The task was to determine the gender of target pictures as quickly as possible by pressing a left or right key.

While the conditioning and rating phases were controlled by custom made Java-programs, reaction-time tasks were implemented using the computer software Affect (Hermans, Clarysse, Baeyens, & Spruyt, 2002).

Table 3:
Most Important Procedural Differences Between the Transfer of Gender Experiments

Experiment	Pre-view	Pre-measure	CS/US selection		Dependent variables	Contin. Awa.	Participants
			CS-selection	US-selection			
P2	no	no	predetermined	predetermined	Gender-scale	yes	1 st year economy
P3	no	yes	individually	individually	Gender-scale Gender-priming	no	2 nd year psychology
P4	no	yes	individually	individually	Gender-scale Gender-priming	yes	paid students
P5	yes	yes	individually	individually	Gender-scale Naming measure	yes	1 st year psychology
P6	yes	yes	individually	predetermined	Gender-scale Naming measure	yes	1 st year economy
P8	yes	yes	predetermined ^a	predetermined	Gender-scale Gender-priming ^b	yes	1 st year psychology

Note. Preview = whether participants get to see each picture *before* the pre-acquisition scale measure; Premeasure = whether there was a pre-acquisition gender-scale measure; CS/US selection = how the CSs and USs are selected for acquisition, i.e. individually selected on the pre-measure or predetermined for everyone on the basis of a rating-study; Contin. Awa. = whether there was a contingency-awareness measure at the end of the experiment

^a In this experiment we selected (in our eyes) somewhat more differential stimuli as CSs.

^b Due to a programming error the data from this gender-priming task was unusable.

Procedure.

The cover story was the same as the one used in Experiment P2. The experiment started with the pre-acquisition gender-rating: the participant was asked to rate all the pictures on the 101-point gender-scale and based on their ratings, the 3 most male, 3 most female, and 12 most neutral pictures (i.e., closest to the individually determined mean) were selected for each participant separately. As in the previous experiment, nine pairs were generated for acquisition: three neutral-male, three neutral-female, and three neutral-neutral pairs. The assignment of the CS-US pairs was identical to the one used in Experiment P1: the crucial neutral-male pairs and neutral-female pairs were formed with the most neutral pictures, the neutral-neutral control pairs with the remaining neutral pictures. The nine pairs were each shown nine times during acquisition, using the same timing as in Experiment P1.

After this conditioning phase, the priming-measure was administered. The prime-picture was presented for 200 ms. The target picture appeared 50 ms after the prime had disappeared (Stimulus Onset Asynchrony, or SOA, was 250 ms). Participants were told to ignore the prime, but to respond based on the gender of the target picture by pressing a left or right key. Whether the left/right key had to be pressed for a male/female target was counterbalanced. Each prime-target pair was presented twice, resulting in two equal blocks of 72 trials each. After every 36 trials, there was a short break. Before the test priming trials were presented, participants received 10 practice trials with the practice stimuli. After completion of the priming measure, participants were asked to reassess the age of the 12 pictures used during acquisition on the 21-point gender-scale. Participants were told that their assessment might have changed or might have stayed the same compared to the first rating phase, and that they should base their assessment on their current impressions. These instructions were given in all the remaining experiments in this Chapter. The experiment ended with participants rating the nine CS and nine US pictures on the valence-scale. Due to programming limitations, the contingency-awareness measure that was used in Experiment P2, could not be added.

Results

Gender-scale.

For all participants, the estimated gender for the three pictures of the same CS or same US categories were averaged for each moment of measurement separately. To make the post-acquisition data comparable to the pre-acquisition data, they were transformed from a 21-point scale to a 101-point scale (i.e., multiplied by five). The transformed average estimated gender for these categories are presented in Table 2 (p. 38). To test our main prediction, we conducted a 2 x 2 ANOVA with CS category (i.e., CS-male versus CS-female) and moment (pre- versus post-acquisition) as within-subjects variable. Results showed that acquisition did have a differential effect on the CS-male compared to the CS-female, $F(1, 31) = 8.24$, $p = .007$, effect size $d = 0.52$. The neutral pictures' gender-assessment had shifted in the direction of the US with which it had been paired.

Table 4:

Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Gender-Priming Measure from Experiments P3 and P4 (standard errors in parentheses)

Experiment	Trial type	RT	RT d	%Errors	%Error d
P3	Congruent	489 (9)	0.01	1.5 % (0.4 %)	0.18
	Incongruent	489 (9)		2.0 % (0.5 %)	
	Noncongruent	491 (9)		1.9 % (0.4 %)	
P4	Congruent	526 (12)	0.04	1.8 % (0.4 %)	-0.02
	Incongruent	527 (11)		1.8 % (0.4 %)	
	Noncongruent	524 (12)		2.1 % (0.6 %)	

Note. d = effect size of the transfer

Priming measure.

In the priming data, we will call the trials on which the CS-category matches the target category (e.g., CS-male prime with a male target picture) congruent trials. The trials where the CS-category is the opposite of the target category (e.g., CS-male prime with a female target picture) will be called incongruent trials. Trials which had a CS-neutral as prime will be called non-congruent.

The average reaction times (over blocks) for the congruent, incongruent and non-congruent trials from the priming-phase are presented in Table 4. As was the case for every reaction time task described in this dissertation, reaction times on trials with an incorrect response were discarded and a cut-off procedure was used to remove outliers. For each cell in the design, reaction times that fell above or below 2.5 standard deviations from a

person's mean reaction time in that cell were discarded. To test our main prediction, we conducted a 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variable. Analysis did not reveal a significant difference between congruent and incongruent trials, $F(1, 31) < 1$, and thus no evidence for transfer of gender. There was no interaction between congruency and block, $F(1, 31) < 1$.

We also analysed the percentage of errors or incorrect responses. The average percentage of errors (over blocks) for the congruent, incongruent and non-congruent trials from the priming-phase are also presented in Table 4. The percentage of errors were analysed using a non-parametric Wilcoxon matched pairs test, because the distribution of the percentage of errors diverged strongly from ANOVA criteria (i.e., the errors were distributed geometrically, combined with very few errors per participant and very few differences in errors between participants). The matched pairs test revealed no significant difference between the percentage of errors made on congruent versus incongruent trials, $Z(32) = 1.0, p = .34$.

Note that, just like for the rating scale measures, for all reaction time measures in this dissertation (e.g., priming and Simon measures), the effect size of the congruency-effect was calculated. In order to calculate the effect size of the congruency-effect in the reaction times, we subtracted (per participant) the congruent reaction-times from the incongruent reaction-times. With the resulting ' Δ RT' we could calculate the effect size of the transfer as $\mu_{\Delta\text{-RT}} / \sigma_{\Delta\text{-RT}}$. For the reaction time data of the current experiment, the resulting Cohen's d was 0.01. The computation of the effect size of the congruency-effect in the percentage of errors was identical, except that the reaction times were replaced by the percentages of errors. For the percentages of errors data of the current experiment, the resulting Cohen's d was 0.18. From now on the resulting d 's will be added after the p -value of the crucial F or Z -tests.

Valence scale.

There was only a marginally significant valence-difference between the US-male and the US-female stimuli: participants liked the US-male pictures ($M = 62.7$) slightly less than the US-female pictures ($M = 67.3$), $F(1, 32) = 3.67, p = .07$. However, there was a significant difference between the linking of the CS-male ($M = 54.7$) and CS-female ($M = 60.7$), $F(1, 32) = 6.06, p = .02$. This suggests that some form of evaluative conditioning had occurred.

To further investigate the possibility that the effect on the gender scale was mediated by an evaluative conditioning effect, we first calculated the size of the evaluative condition effect for each participant, by subtracting the valence for the CS-female category from the CS-male-category. This valence-effect variable ($M = 5.83$) was added as a covariate in our crucial analysis of the gender-scale data. This resulted in a 2×2 ANCOVA with CS category (i.e., CS-male versus CS-female) and moment (pre- versus post-acquisition) as within-subjects variable, and the size of the evaluative conditioning as continuous predictor. Results showed no interaction between the valence-effect and the transfer-effect (effect of acquisition on the CS-male versus CS-female), $F(1, 30) < 1$. The transfer-effect also remained significant, even with the valence-effect included as covariate, $F(1, 30) = 7.60, p = .001$. The results from this section indicate that, although there could be some evaluative conditioning at work, the transfer of gender effect is more or less independent of it.

Discussion

We found evidence for associative transfer of gender: after repeatedly pairing pictures of gender-neutral infants with clearly male/female infants, the gender of the previously neutral infant pictures had shifted in the direction of the infants with which they had been paired. There also was some evidence of evaluative conditioning: clearly female pictures were in general slightly more positive than clearly male pictures, and this valence-difference was also found between the CS-male and CS-female pictures. There was, however, no indication that the transfer of gender effect was mediated by this evaluative transfer effect: they seemed independent of one-another. In this experiment we had no contingency-awareness data, so we were unable to investigate if it had an effect on transfer of gender. It could be that participants were more aware of the pairs in this experiment, and maybe consciously manipulated their ratings on the post-acquisition measure with the US-gender. We had added the priming-measure, which is more demand insensitive, to investigate this possibility. If the CSs had indeed become more male/female, we would have expected a congruency effect. Unfortunately, we found no evidence for transfer of gender in the priming-measure. We cannot discriminate whether this is because our priming measure is not sensitive enough for observing associatively induced congruencies, or whether there was no real (i.e., not caused by a demand effect) transfer of gender effect. We thus decided to replicate this experiment with a contingency awareness measure. Note

that when we find that contingency-awareness is a pre-requisite for transfer of gender, we cannot simply conclude that it is a demand effect: it is possible that contingency-awareness is simply necessary for this association to form. But when we find that contingency-awareness does not influence the transfer of gender, we have evidence that it is *not* a demand effect. A few other small elements of the design were also optimised.

Experiment P4

Method

Participants.

Forty students from various faculties at the University of Southampton (6 men and 34 women) were paid £6.00 for taking part.

Stimuli and measures.

The stimuli and measures were identical to the ones used in Experiment P3, with a few exceptions: First, in this experiment the post-acquisition gender-scale was identical to the pre-acquisition gender-scale (both were on an a 101-point scale). Second, this experiment was run in English with English speaking participants. All other experiments reported in this dissertation were in Dutch. Finally, in the priming measure, there was only a pause after 72 trials (i.e., between the two blocks), and not after each 36 trials.

Procedure.

The procedure was identical to the one used in Experiment P3, with three exceptions: First, the gender-neutral pictures were selected around the middle of the gender-scale (i.e., 50 on the scale from 0 to 100), instead of around the individually determined mean. Second, as already mentioned, the post-acquisition gender measure was, as the pre-acquisition gender-measure, on a 101-point scale. Third, the contingency-awareness measure used in Experiments P1 and P2 was also added at the end of this experiment.

Results

Gender-scale.

For one participant all rating-data, including contingency-awareness data, were lost. For all remaining participants, the estimated gender for the three pictures of the same CS or same US categories were averaged for each moment of measurement separately. The average estimated gender for these categories are presented in Table 2 (p. 38). To test our main prediction, we conducted a 2 x 2 ANOVA with CS category (i.e., CS-male versus CS-female) and moment (pre- versus post-acquisition) as within-subjects variable. Contrary to expectations, results showed that acquisition did not have a differential effect on the CS-male compared to the CS-female, $F(1, 38) < 1$, effect size $d = 0.14$.

Priming measure.

The average reaction times (over blocks and after cut-off) for the congruent, incongruent and non-congruent trials from the priming-phase are presented in Table 4 (p. 45). To test our main prediction, we conducted a 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variable. Analysis revealed no significant reaction time difference between congruent and incongruent trials, $F(1, 39) < 1$, effect size $d = 0.04$, and thus no evidence for transfer of gender. There was no interaction between congruency and block, $F(1, 39) < 1$.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent, incongruent and non-congruent trials from the priming-phase are also presented in Table 4 (p. 45). The percentage of errors were analysed using a non-parametric Wilcoxon matched pairs test, because the distribution of the percentage of errors diverged strongly from ANOVA criteria. The matched pairs test revealed no significant difference between the percentage of errors made on congruent versus incongruent trials, $Z(40) = 0.3$, $p = .79$, effect size $d = -0.02$.

Contingency awareness.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 2.1 of the 9 pairs ($SE = 0.3$). 26 % of the participants selected the correct US-picture 4 or more times out of 9, thus performing significantly above chance ($p = .048$). The category-awareness measure showed that, on average, the correct US-category was

selected for 3.4 of the 9 pairs ($SE = 0.3$). 15 % of the participants selected the correct US-category 6 or more times out of 9, thus performing significantly above chance ($p = .042$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis revealed no significant correlation, $r = -.02$, $p = .90$, between the category-awareness score and the transfer effect score on the gender-scale. We also investigated the effect of contingency awareness on the priming data. There was no significant correlation, $r = -.17$, $p = .28$, between category awareness and the congruency effect in the reaction time data. Also no correlation, $r = .00$, $p = .99$ was found between category awareness and the congruency effect in the percentage of errors.

Valence scale.

There was no significant valence-difference between the US-male ($M = 62.2$) and the US-female ($M = 64.9$), $F(1, 38) < 1$. There was also no significant difference between the liking of the CS-male ($M = 61.0$) and CS-female ($M = 61.1$), $F(1, 38) < 1$. As such we have no evidence that evaluative conditioning could have influenced the (absent) transfer of gender.

Discussion

We were unable to replicate the transfer of gender effect found in Experiment P3: pairing pictures of gender-neutral infants with clearly male/female infants had no differential effect on the neutral infants. We found no effect on the gender-scale, nor in the priming measure. Being less or more aware of the contingencies did not influence the results. The next experiment was another attempt to replicate the results of Experiment P3.

Experiment P5

The gender-priming measure, which had failed to give an effect twice, was replaced by another indirect measure. We decided on a naming measure, in which participants saw the CSs together with six first names (three boy and three girl names) and were asked to guess what the first-name of this baby was.

We also started the experiment with a preview phase, in which participants saw each of the infant-pictures for a few seconds with the knowledge that they would have to assess the gender of these infants. This preview phase was added to help participants define anchor-points before the rating phase so that they would subsequently rate all the pictures on the same (internal) scale. We hoped for more stable ratings this way, and thus ‘better’ neutral, US-male and US-female pictures. In the previous experiments, some of the US-males and US-females were in our eyes not the most male/female pictures in the set: we had expected that the older infants, which were in our eyes the most male/female, would be given the most male/female scores. This could have been caused by changes in anchor-points during the pre-acquisition gender-rating: if participants start the first gender-rating with only young infant-pictures, they might give extreme scores to the slightly more male/female young infants. When subsequently seeing an older infant-picture, that is clearly more male/female, they can not give a more extreme score. This argument is somewhat corroborated by the fact that visual inspection of the data showed that for some of the participants (25% and 28% of participants in Experiment P3 and P4 respectively) at least one US-male or US-female picture had shifted 25 points or more on the 101-point scale from pre- to post-acquisition: if those US-male/US-females were truly male/female for that participant, then we would not have expected the ratings to shift that much between moments of measure. The preview phase might solve this problem. Remember that an overview of the most important differences between all the transfer of gender experiments of this chapter, can be found in Table 3 (p. 43).

Method

Participants.

Thirty-seven first year psychology students (7 men and 30 women) participated in exchange for course credits in this experiment.

Stimuli and measures.

The same set of fifty-three colour pictures of infants as in Experiment P3 and P4 were used for conditioning and ratings. To measure participant’s gender-assessment we used the gender-scale from *very certainly a girl/women* (= 0) to *very certainly a boy/man* (= 100) for pre- and post-acquisition. To measure liking of the selected pictures, the valence-scale from *very disliked* (= 0) to *very liked* (= 100) was used. The gender-priming measure that

was used in Experiment P3 and P4 was replaced by a naming measure. In this naming measure participants saw the CS-pictures and were presented for each CS picture with a different set of three male and three female first names. Participants were instructed to guess what the first name of the presented infant was. We predicted that infant-pictures that were paired with a male US would be assigned a male name more often than infant-pictures that were paired with a female US.

Procedure.

The procedure was identical to the one in Experiment P4, with the following exceptions. In this experiment, the pre-acquisition gender-rating was preceded by a preview-phase in which participants were told that they would see a series of infant-pictures which they would have to rate on a gender-scale in a subsequent phase. They were told that they had a chance to preview the pictures to get an idea of what kind of pictures they would have to rate. Another change in comparison to Experiment P4 was that the most gender-neutral pictures were once again, like in Experiments P2 and P3, selected around the individually determined mean (instead of around the middle of the gender-scale). Pair-generation and acquisition were identical to Experiment P4. After acquisition, the (second) gender-rating and new naming-measure were administered. In the naming-measure, all the CS pictures were presented at random, together with three male and three female names each. Note that the order of these two post-acquisition measures, gender-scale and naming, was counterbalanced. The experiment again ended with the valence rating and the contingency-awareness procedure that was identical to the one used in Experiment P4.

Results

Gender-scale.

A manipulation check showed that the preview phase did not stabilise the US-ratings as hoped: for 22 % of the participants at least one of the US-male or US-female changed more than 25 points from pre- to post-acquisition.

For all participants, the estimated gender for the three pictures of the same CS or same US categories were averaged for each moment of measurement separately. The average estimated gender for these categories are presented in Table 2 (p. 38). To test our main prediction, we conducted a 2 x 2 ANOVA with CS category (i.e., CS-male versus

CS-female) and moment (pre- versus post-acquisition) as within-subjects variable. Again, results showed no evidence that acquisition had a differential effect on the CS-male compared to the CS-female, $F(1, 36) < 1$, effect size $d = 0.12$.

Table 5:

Average Percentage of Male Names (%Male) Given in the Naming Tasks of Experiments P5 and P6 (standard errors in parentheses)

Experiment	CS-category	%Male	d %Male
P5	CS-male	60.4 % (5.8 %)	0.26
	CS-female	52.3 % (6.1 %)	
	CS-neutral	65.8 % (4.4 %)	
P6	CS-male	57.7% (6.0 %)	0.06
	CS-female	55.0% (6.0 %)	
	CS-neutral	61.7% (5.0 %)	

Note. d = effect size of the transfer

Naming measure.

For all participants, the number of male names selected for the three pictures of the same CS-category was divided by three and multiplied by 100. These percentages of male-names given to CS-male, CS-female and CS-neutral pictures during the naming phase are presented in Table 5. To test our main prediction, we conducted an ANOVA with CS category (CS-male versus CS-female) as within-subjects variable. Although 8 % more male names were given to CS-male compared to CS-female pictures, this difference was not significant, $F(1, 36) = 2.40$, $p = .13$.

The effect sizes for the naming measures used in Experiment P5 and Experiment P6 were calculated by first subtracted (per participant) the percentage of male names given to the CS-female pictures from the percentage of male names given to the CS-male pictures. With the resulting ‘ Δ %Male’ we calculated the effect size of the transfer as $\mu_{\Delta\text{-}\%Male} / \sigma_{\Delta\text{-}\%Male}$. In this experiment, the resulting Cohen’s d was 0.26. The resulting d will from now on be added after the p -value of the crucial F-test.

Contingency awareness.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 3.3 of the 9 pairs ($SE = 0.4$). 32 % of the participants selected the correct US-picture 4 or more times out of 9, thus performing significantly above chance ($p = .048$).

The category-awareness measure showed that, on average, the correct US-category was selected for 4.8 of the 9 pairs ($SE = 0.4$). 38 % of the participants selected the correct US-category 6 or more times out of 9, thus performing significantly above chance ($p = .042$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis revealed no significant correlation, $r = .27$, $p = .10$, between the category-awareness score and the transfer effect score on the gender-scale. We also investigated the effect of contingency awareness on the naming data. There was also no significant correlation, $r = .01$, $p = .94$, between category awareness and the transfer effect score in the naming data.

Valence scale.

There was no significant valence-difference between the US-male ($M = 57.7$) and the US-female ($M = 62.4$), $F(1, 36) = 2.20$, $p = .15$. There was also no significant difference between the liking of the CS-male ($M = 65.4$) and CS-female ($M = 61.8$), $F(1, 36) = 2.70$, $p = .11$). As such we have no evidence that evaluative conditioning could have influenced the (absent) transfer of gender.

Discussion

We were again unable to replicate the transfer of gender effect found in Experiment P3: pairing pictures of gender-neutral infants with clearly male/female infants had no differential effect on the neutral infants. We found no effect on the gender-scale and although the effect in the new naming-measures was in the right direction, it was not significant.

Experiment P6

In two consecutive experiments, we were twice unable to replicate the transfer of gender effect from Experiment P3. In Experiment P6, we tried a last time to replicate the results of Experiment P3. We started with the design of Experiment P5, with a few adjustments. The most important one was that the US-male and US-female pictures were no longer selected on the basis of the individual pre-acquisition gender-measure, but were a priori selected from the rating study described in Experiment P2. So, while the neutral pictures were still

individually selected by means of the gender-measure, the US-male and US-females were chosen a priori and were the same as in Experiment P2. This was done in the hope of making the US-male and US-females ratings more stable over the moments of measure. We had hoped that the preview phase in the previous experiment would have resulted in only extreme male/female scores for the older infants which were in our eyes the most male/female. But some participants still gave extreme scores to the more younger infants, sometimes resulting in a large shift in US ratings from pre- to post-acquisition. By pre-selecting the USs, we could prevent younger infants that were also given an extreme score, to become a US-male or US-female. Remember that an overview of the most important differences between all the transfer of gender experiments of this chapter, can be found in Table 3 (p. 43).

Method

Participants.

Twenty-six second year economy students (13 men and 13 women) participated in exchange for course credits.

Stimuli and measures.

The stimuli and measures were identical to those of Experiment P5.

Procedure.

The procedure was identical to the one of Experiment P5, with the following exceptions. The US-male and US-female pictures were a priori determined and the same as the ones used in Experiment P2. These pictures were, however, still rated in the pre-acquisition gender measure, like all the other infant-pictures. The neutral stimuli were still, like in Experiment P5, selected individually. As such we had the advantages of a pre-measure (i.e., being able to correct for baseline differences) and individually selected CSs (i.e., selecting the most neutral CSs for that particular participant), and at the same time were able to pre-select clear male and female pictures as USs. Note that that the most gender-neutral pictures were selected around the middle of the gender-scale (instead of around the individually determined mean, like in Experiment P5).

Results

Gender-scale.

A manipulation check showed that pre-selecting the US-male and US-female pictures did somewhat stabilise the US-ratings: for only 15 % of the participants at least one of the US-male or US-female changed more than 25 points from pre- to post-acquisition (compared to 25 %, 28 % and 22 % in Experiments P3, P4 and P5 respectively).

For all participants, the estimated gender for the three pictures of the same CS or same US categories were averaged for each moment of measurement separately. The average estimated gender for these categories are presented in Table 2 (p. 38). To test our main prediction, we conducted a 2 x 2 ANOVA with CS category (i.e., CS-male versus CS-female) and moment (pre- versus post-acquisition) as within-subjects variable. Again, results showed no evidence that acquisition had a differential effect on the CS-male compared to the CS-female, $F(1, 25) < 1$, effect size $d = -0.03$.

Naming measure.

For all participants, the number of male names selected for the three pictures of the same CS-category was divided by three and multiplied by 100. These percentages of male-names given to CS-male, CS-female and CS-neutral pictures during the naming phase are presented in Table 5 (p. 53). To test our main prediction, we conducted an ANOVA with CS category (CS-male versus CS-female) as within-subjects variable. There was no significant difference in the percentage of male names given to CS-male versus CS-female pictures, $F(1, 25) < 1$, effect size $d = 0.06$.

Contingency awareness.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 2.7 of the 9 pairs ($SE = 0.4$). 23 % of the participants selected the correct US-picture 4 or more times out of 9, thus performing significantly above chance ($p = .048$). The category-awareness measure showed that, on average, the correct US-category was selected for 4.0 of the 9 pairs ($SE = 0.4$). 23 % of the participants selected the correct US-category 6 or more times out of 9, thus performing significantly above chance ($p = .042$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis revealed no significant correlation, $r = .26$, $p = .21$, between the category-awareness score and the transfer effect score on the gender-scale. We also investigated the effect of contingency awareness on the naming data. There was a marginally significant correlation, $r = .35$, $p = .08$, between category awareness and the transfer effect score in the naming data, indicating that the more participants were aware of the ‘gender-category’ of the US-picture that was paired with a CS, the larger the transfer-effect tended to be.

Valence scale.

There was a significant valence-difference between the US-male and the US-female: participants liked the US-female pictures ($M = 63.5$) more than the US-male pictures ($M = 47.4$), $F(1, 25) = 17.83$, $p < .0003$. There was however, no significant difference between the liking of the CS-male ($M = 54.3$) and CS-female ($M = 56.4$), $F(1, 25) < 1$. As such we have no evidence that evaluative conditioning could have influenced the (absent) transfer of gender.

Discussion

We were again unable to replicate the transfer of gender effect found in Experiment P3: pairing pictures of gender-neutral infants with clearly male/female infants had no differential effect on the neutral infants. We found no effect on the gender-scale, nor in the naming-measures.

In the five transfer of gender experiments performed, we only found evidence for associative transfer of gender in Experiment P3. In the three subsequent attempts, we were unable to replicate this effect. One possible reason for why we did not find good evidence for associative transfer of non-evaluative properties might be that our procedure is not suited for transferring stimulus properties, evaluative or non-evaluative. Although our procedure is based on the one that Baeyens and colleagues successfully used to investigate EC (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992), this does not guarantee that it will yield EC. In recent EC experiments that used picture-picture paradigms, there were several failures to replicate the basic EC effect. Even exact replications of the experiments

of Baeyens and colleagues sometimes failed to result in EC (e.g. Field, 2002). For these reasons, we decided to try to replicate EC with our paradigm.

Experiment P7

We used the same stimuli and a procedure very similar to the one used in Experiment P4, but focused on valence transfer instead of gender transfer. The gender-scale was replaced by a valence scale and the gender-priming by an affective priming procedure. If we find EC with our procedure, we know that at least some kind of associative transfer is possible. We do *not* claim that a procedure that is (not) able to produce EC will also be (un)able to produce associative transfer of non-evaluative stimulus properties.⁹ We only suggest that if our procedure is able to yield EC, that at least some procedural elements were appropriate to enable associative transfer.

Method

Participants.

Forty first year psychology students (15 men and 25 women) participated in exchange for course credits.

Stimuli and measures.

The same set of fifty-three colour pictures of infants as in Experiment P4 were used for conditioning and ratings. To measure participant's liking of the pictures, a valence-scale from *very disliked* (= 0) to *very liked* (= 100) was used.

As indirect measure, we used an affective priming procedure (e.g., De Houwer et al., 1998; Hermans, Vansteenwegen, et al., 2002) instead of the gender priming procedure. A separate set of four gender-neutral (in the rating study) infant-pictures were used as practice primes in the priming procedure. As targets for the affective-priming procedure,

⁹ Data from Stevenson and colleagues already suggests that there is not necessarily a reciprocal relation between EC and transfer non-evaluative properties. Stevenson, Boakes, and Prescott (1998), for example, did find transfer of the sweetness/sourness of a taste onto a flavour without finding transfer of the valence of the taste to the flavour. Because a sour taste is normally more negative than a sweet taste, one would expect EC to occur.

we used four clearly positive and four clearly negative words. Another eight were used as practice targets for the priming measure. During the priming task, participants received target-words that had a clearly positive or negative valence. The task was to determine the valence of target words as quickly as possible by pressing a left (negative) or right (positive) key. A prime-picture was presented before each target, but participants were instructed to ignore this prime. The nine CSs were used as primes, and were combined with the eight valenced target words. Each prime-target pair was presented twice, resulting in two equal blocks of 72 trials each. After every 36 trials, there was a short break. All other parameters of the priming-procedure were identical to the one in Experiment P4.

Procedure.

The procedure was identical to Experiment P4, with a few exceptions. In this experiment, the pre-acquisition evaluative rating was preceded by the preview-phase, just like in Experiments P5 and P6. The gender-measure's were replaced with their affective variants. In accordance with our naming convention, we will name the US pictures after the category to which they belong, i.e. US-positive, US-negative or US-neutral. Their corresponding CSs will be named CS-positive, CS-negative and CS-neutral respectively.

Results

Valence-scale.

For all participants, the valence scores for the three pictures of the same CS or same US categories were averaged for each moment of measurement separately. The average valence for these categories are presented in Table 6. To test our main prediction, we conducted a 2 x 2 ANOVA with CS category (i.e., CS-positive versus CS-negative) and moment (pre- versus post-acquisition) as within-subjects variable. Contrary to expectations, results showed that acquisition did not have a differential effect on the ratings for CS-positive compared to the CS-negative stimuli, $F(1, 39) < 1$, effect size $d = 0.09$.

Table 6:

Average Valence-Ratings for the CS-Categories from Experiment P7 (standard errors in parentheses)

CS-category	Valence-Ratings		<i>d</i> Rating
	Pre-acq.	Post-acq.	
CS-positive	49.6 (0.5)	52.9 (1.1)	0.09
CS-negative	50.0 (0.4)	52.6 (1.2)	
CS-neutral	49.9 (0.6)	53.2 (1.4)	
US-positive	85.2 (1.7)	78.4 (1.9)	
US-negative	21.7 (2.0)	31.5 (2.4)	
US-neutral	50.3 (0.9)	53.8 (1.6)	

Note. *d* = effect size of the transfer

Priming measure.

The average reaction times (over blocks, after removal of incorrect responses and after cut-off) for the congruent, incongruent and non-congruent trials from the priming-phase are presented in Table 7. To test our main prediction, we conducted a 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variable. Analysis revealed no significant reaction time difference between congruent and incongruent trials, $F(1, 39) < 1$, effect size $d = -0.05$, and thus no evidence for EC.

We also analysed the percentage of errors or incorrect responses. These average percentages of errors (over blocks) for the congruent, incongruent and non-congruent trials from the priming-phase are also presented in Table 7. The matched pairs test revealed no significant difference between the percentage of errors made on congruent versus incongruent trials, $Z(40) = 1.4$, $p = .17$, effect size $d = -0.19$.

Table 7:

Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Affective Priming Measure from Experiment P7 (standard errors in parentheses)

Trial type	RT	<i>d</i> RT	%Errors	<i>d</i> %Errors
Congruent	579 (12)	-0.05	2.8 % (0.5 %)	-0.19
Incongruent	578 (12)		2.0 % (0.6 %)	
Noncongruent	580 (12)		2.9 % (0.6 %)	

Note. *d* = effect size of the transfer

Contingency awareness.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 2.7 of the 9 pairs ($SE = 0.3$). 25 % of the participants selected the correct US-picture 4 or more times out of 9, thus performing significantly above chance ($p = .048$). The category-awareness measure showed that, on average, the correct US-category was selected for 4.3 of the 9 pairs ($SE = 0.3$). 28 % of the participants selected the correct US-category 6 or more times out of 9, thus performing significantly above chance ($p = .042$).

We subsequently investigated the influence of contingency awareness on the valence transfer. Analysis revealed no significant correlation, $r = -.04$, $p = .81$, between the category-awareness score and the transfer effect score on the valence-scale. We also investigated the effect of contingency awareness on the priming data. There was no significant correlation, $r = .15$, $p = .37$, between category awareness and the congruency effect in the reaction time data. Also no correlation, $r = -.05$, $p = .76$, was found between category awareness and the congruency effect in the percentage of errors.

Discussion

We were unable to find EC in this experiment: pairing affectively neutral infant pictures with positive/negative infants had no differential effect. This is not an isolated result in EC research. As previously mentioned, there have been other failures to replicate EC in picture-picture paradigms (e.g., Field, 2002). This does not mean that EC does not exist, but it does suggest that there are some important boundary conditions that are still unknown. In the present experiment, two possible boundary conditions might not have been met. First, it is possible that our pictures were not sufficiently positive/negative: participants seemed to be reluctant to give extreme positive/negative valence-ratings to the pictures in our set. While the experimental USs in our transfer of gender experiment on average scored in the top/bottom tenth of the scale (i.e. above 90 or below 10), the US-positive and US-negative pictures scored on average 78 and 31 respectively. Another possible reason is that our stimuli might not have been distinct enough. Especially the younger infant-pictures (which were often the most gender-neutral) were more alike than the pictures used in most other picture-picture EC experiments. It might be that participants were not able to discriminate between CSs. This is somewhat corroborated by the low contingency-awareness scores in the gender-transfer experiments (using baby pictures),

compared to the age-transfer experiment (using pictures of adults): if participants have a hard time discriminating between the baby-pictures, then we would expect lower contingency-awareness. The inability to clearly differentiate between stimuli could prevent any associative transfer of stimulus properties: evaluative or non-evaluative. This explanation does have implications for the interpretation of our transfer of gender experiments: if the stimuli are too similar for valence to transfer, then probably they are also too similar for gender to transfer. In a next experiment, we therefore tried to use more distinct pictures and again tried to transfer gender.

Experiment P8

In Experiment P8, we again tried to replicate the transfer of gender effect. Because in the previous experiments the neutral infant pictures might have been too similar and hard to distinguish from one another, we decided to select more distinct CSs. We chose for the same set of stimuli, but pre-selected the neutral CSs in such a way that they were (relatively) distinct from one another. We also removed the neutral-neutral control pairs, to keep the number of CSs low and better distinguishable. Note that an overview of the most important differences between all the transfer of gender experiments of this chapter, can be found in Table 3 (p. 43).

Method

Participants.

Forty-four first year psychology students (7 men and 37 women) participated in exchange for course credits.

Stimuli and measures.

The same set of fifty-three colour pictures of infants were used as in the other transfer of gender Experiments. For the conditioning phase, we used a subset of the infant-pictures that were selected using the rating study of Experiment P2. As clearly male and clearly female infants we used the same pictures as in Experiment P2. From the 12 gender-neutral infant pictures of Experiment P2, the author and another researcher selected the 6 pictures that were in their opinion the most distinguishable from one another. The stimuli for the

gender-priming measure, were the same as the ones used in Experiment P3 and P4. Both the gender-scale and the priming measure used were the same as in Experiment P4.

Procedure.

The procedure was identical to the one used in Experiment P4, with two exceptions. First, the experiment started with the preview-phase that was also used in Experiments P5, P6 and P7. Second, due to the decrease in the number of CSs, we removed the neutral-neutral control-pairs from this experiment. As a result of this change in design, only six pairs were generated for acquisition: three neutral-male and three neutral-female pairs. Each CS was randomly assigned to a US for each participant separately. From acquisition onwards, this experiment was identical to Experiment P4, except that there were fewer priming and contingency awareness trials due to the decrease from nine to six CSs.

Results

Gender-scale.

For all participants, the estimated gender for the three pictures of the same CS or same US categories were averaged for each moment of measurement separately. The average estimated gender for these categories are presented in Table 2 (p. 38). To test our main prediction, we conducted a 2 x 2 ANOVA with CS category (i.e., CS-male versus CS-female) and moment (pre- versus post-acquisition) as within-subjects variable. Contrary to expectations, results showed that acquisition did not have a differential effect on the CS-male compared to the CS-female, $F(1, 38) < 1$, effect size $d = 0.19$.

Priming measure.

Due to a programming error in this experiment, the data from the gender-priming procedure was unusable.

Contingency awareness.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 1.7 of the 6 pairs ($SE = 0.2$). 30 % of the participants selected the correct US-picture 3 or more times out of 6, thus performing significantly above chance ($p = .009$). The category-awareness measure showed that, on average, the correct US-category was

selected for 3.6 of the 6 pairs ($SE = 0.2$). 27 % of the participants selected the correct US-category 5 or more times out of 6, thus performing significantly above chance ($p = .018$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis did reveal a significant correlation, $r = .39$, $p = .008$, between the category-awareness score and the transfer effect score on the gender-scale, indicating that the more participants were aware of the ‘gender-category’ of the US-picture that was paired with a CS, the larger the transfer-effect. To examine this relation further, we divided participants into those who did ($N = 12$) and did not ($N = 32$) perform above chance on the category awareness score. Analyses revealed a significant positive transfer effect for the category aware participants, $F(1, 11) = 5.90$, $p = .03$, effect size $d = 0.73$, but not for the other participants, $F(1, 31) = 0.05$, $p = .83$, effect size $d = 0.04$.

Valence scale.

There was no significant valence-difference between the US-male ($M = 62.2$) and the US-female ($M = 64.9$), $F(1, 38) < 1$. There was also no significant difference between the liking of the CS-male ($M = 61.0$) and CS-female ($M = 61.1$), $F(1, 38) < 1$. As such we have no evidence that evaluative conditioning could have influenced the (absent) transfer of gender.

Discussion

Although we believe we used more clearly differentiable stimuli in this experiment, we were not able to (re)produce a clear transfer of gender. We found no overall effect on the gender-scale. In this experiment contingency awareness did influence the results: for participants who were category-aware for 4 or more of the six pairs, did we find a transfer effect on the gender-scale. This is the first experiment in which we find evidence that higher awareness results in a transfer-effect. We are unable to determine whether this gender-transfer is caused by a demand effect, or whether awareness of the contingencies is simply a prerequisite for gender-transfer.

Our inability to find general transfer of gender in this experiment suggests that either our stimuli were still not distinct enough or that it was not the differentiability of our stimuli (alone) that caused the absence of transfer of gender in our previous experiments.

Overall discussion of picture-picture studies

In seven conditioning experiments, we tried to associatively transfer non-evaluative stimulus properties of depicted persons, namely age and gender. We once found clear evidence for transfer of gender in Experiment P3, but were unable to replicate it in Experiments P2, P4, P5, P6, P8. In Experiment P8 we did also find a significant transfer effect, but only for the participants who performed well on the category awareness test.

These results indicate that the transfer effect was at best small and unreliable. A simple meta-analysis was conducted on the gender-scale results to investigate whether, overall, there remained evidence for associative transfer of gender. For this meta-analysis, we used the program and manual of Schwarzer (1989) to combine the effect sizes from the gender-transfer experiments¹⁰. Overall (i.e., over Experiments P2, P3, P4, P5, P6, and P8) there still remained a significant transfer-effect on the gender-scale, $Z = 1.74$, $p = .04$. The analysis revealed that the overall effect size of this transfer effect is rather small (mean effect size $d^+ = 0.17$), and that there is no significant difference between the transfer-effect over the different experiments, homogeneity $Q = 2.37$, $p = 0.80$.

We subsequently investigated whether contingency-awareness had any influence on transfer of gender. Only in Experiment P8 did we find a direct connection between category awareness and transfer of gender. To further explore the influence of category awareness on the transfer-effect, we aggregated the data of all the transfer of gender experiments. We subsequently performed analysis on the transfer effect score (with which the Cohen's d was calculated) of the gender-scale¹¹. A one sample t-test on the transfer effect score of all 209 participants showed that the transfer effect was small but significant,

¹⁰ Schwarzer used the 'weighted integration model' of Hedges and Olkin to calculate the overall effect size and significance (as cited in Schwarzer, 1989). To acquire the input values for this analysis, we transformed the relevant d -values of the transfer-effects into Hedge's g -values (by dividing our reported population value Cohen's d by the $\sqrt{(N/df)}$). As the reliability coefficient of our gender-scale measure was not available, perfect measures were inserted for analysis purpose (i.e., $r_{tt} = 1$, as suggested by Schwarzer, 1989).

¹¹ We are aware of the fact that aggregating data of different experiments is not entirely appropriate when the design and procedure of the experiments is not identical. However, the different gender-transfer experiments were similar to a large degree. Also, when the transfer effect scores of all 209 participants were analysed using an ANOVA with experiment as between-subjects variable, the interaction between experiment and the gender-shift was not significant, $F(5, 203) < 1$.

$t(208) = 2.28, p = .02$, effect size $d = 0.16$. . We also performed an analysis on all data except the data of Experiment P3 in which no awareness measure was present. This allowed us to examine the relation between category awareness and gender transfer. In this set of data, the transfer effect was in the right direction but no longer significant, $t(176) = 1.59, p = .11$, effect size $d = .12$. The correlation between the relative category awareness score (i.e., number of CSs for which a US from the correct category was selected divided by the number of CSs) and the transfer effect score was, however, significant, $r = .15, p = .04$. For the 43 participants who were significantly category aware, we found a significant positive transfer effect, $t(42) = 2.14, p = .04$, effect size $d = 0.33$. For the remaining 134 participants, who were not significantly category aware, we found no trace of a transfer effect, $t < 1, d = 0.04$, even though this test had sufficient power (.82) to detect a small effect ($d = 0.25$)¹².

The meta-analysis and combined analysis of Experiments P2, P3, P4, P5, P6 and P8 strengthens the conclusion that the transfer effect was small at best, especially when taking into account only those participants who performed below chance on the category awareness. The fact that a significant effect was only found for participants who performed above chance on the category awareness measure had several implications. First, it raises the possibility that the transfer-effect observed in Experiment P3 was caused by category aware participants. Second, it suggests that associative transfer of gender depends on conscious knowledge of the CS-US contingencies. If this is the case, it would imply that associative transfer of gender differs from EC which is often found to be independent of awareness of the contingencies (for reviews, see De Houwer, Baeyens, & Hendrickx, 1997; Field, 2000). But it could also indicate that the observed transfer effects were actually demand effects: participants who are aware that a CS is always paired with a male/female US and are subsequently asked to rate the gender of the CS might infer that the experimenters expects them to rate that CS as more male/female. According to this demand interpretation, the observed transfer was not caused by a genuine change in gender-perception of the CSs, but merely do to the fact that the participants intentionally wanted to conform with what they thought was expected of them. This would also explain why in none of the indirect gender measures an effect was found, as we expected those indirect

¹² We used the computer program Gpower (Faul & Erdfelder, 1992) to perform power-calculations. The power of the transfer-test was calculated as for a two-sided ‘matched-pairs t-test’, using the Gpower manual of Buchner, Erdfelder, and Faul (1997).

measures to be less susceptible to demand effects. Also note that in Experiment P3, in which a significant transfer effect was found, was the only experiment in which second year psychology students participated. These students have already participated in many previous psychological experiments, and thus might be more apt to see through the cover story (see also Page, 1969). Although these considerations do not prove that the transfer effects were caused by demand-effects, they do imply that demand effects were at least possible. This casts doubts on whether the observed transfer effect were genuine.

We also investigated whether some of our procedural elements might have prevented or suppressed associative transfer of stimulus properties. One of the elements that might have played a role was the similarity of our stimuli: the used baby-pictures might have been hard to discriminate from one another. This is somewhat corroborated by the low contingency-awareness ratings for the gender-transfer experiments (compared to the gender-transfer experiment). To investigate this we used somewhat (in our eyes) more differential stimuli in Experiment P8, but we were still unable to replicate/find clear transfer of gender. We were however unable to excluded the possibility that are stimuli were still not distinct enough. Moreover, in a control experiment (Experiment P7), we were even unable to replicate EC with our procedure. Taken together, we were unable to exclude the possibility that some procedural elements (possibly the similarity) might have prevented us from finding any form clear of property transfers, evaluative or non-evaluative. So there seem to be clear boundary conditions to EC, which might also play a role in ATNESP.

Chapter 4

Symbol-Word Studies

Due to the difficulty in consistently finding any kind of transfer in our picture-picture paradigm, we decided to search for ATNESP in another paradigm. In this second line of experiments, we used a symbol-word paradigm in which we tried to transfer the gender of first names onto Japanese characters. We chose a symbol-word paradigm because EC had recently been demonstrated with such a paradigm by Diaz et al. (2002) and a procedure that yields EC might enhance the chance of finding non-evaluative transfer. In their EC-experiments, Diaz et al. observed that after repeated paired presentations of Japanese characters (the CSs) with either a clearly positive or a clearly negative word (the US), the Japanese characters had become more positive or negative respectively. They had not only observed EC on a valence scale, but also in an affective priming procedure which is less sensitive to demand-effects. This paradigm was easily adaptable for researching gender-transfer, by replacing the positive and negative words that were used as USs by clearly male or female first names.

In this line of experiments we also assessed more directly if demand awareness could have played a role. In the picture-picture studies, discussed in the previous chapter, we found indications that transfer effects only occurred when participants were aware of the CS-US contingencies. Although this allows for the possibility that the transfer effects were due to demand characteristics, it is also possible that contingency awareness is merely necessary for ATNESP to occur. In the present experiments we therefore interviewed our participants about whether they consciously and intentionally used their knowledge of the CS-US contingencies as a basis for their post-acquisition ratings.

Experiment S1

This experiment is similar to the EC experiment of Diaz et al. (2002), but concentrates on gender transfer instead of valence transfer. During acquisition, Japanese characters were paired with a male or a female first name, instead of the positive or negative word used by Diaz and colleagues. After acquisition a gender-scale measure was used to assess the

perceived gender of the characters, instead of a valence scale. As a more indirect measure, a gender-name priming procedure was used, similar to the one used by Draine and Greenwald (1998), instead of an affective priming procedure.

Method

Participants.

Twenty-six first year psychology students (6 men and 20 women) participated in exchange for course credits.

Stimuli and measures.

Four clearly differentiable Japanese characters were used as CSs for acquisition. Four first names (two unambiguous male and two unambiguous female) were used as USs. The first instrument to measure participant's gender-assessment, was a gender-scale. For this scale measure, participants were told that the Japanese characters (the CSs) referred to a male/female Japanese first name and they were asked to indicate, on a scale from *female* (= 0) to 20 *male* (= 20), what they thought the gender of that Japanese name was.

Our second instrument to measure gender-assessment, was a gender-priming measure. This priming measure was based on an existing gender-name priming procedure. In previous gender-priming experiments, participants responded faster when the gender of a target first name, that had to be categorized as male or female, was the same as the gender of the preceding prime first name (e.g., Draine & Greenwald, 1998). We successfully replicated this gender priming effect in a pilot study. Therefore, if we find that after a certain prime, participants are faster to classify the target as a male than as a female name, and if for another prime-picture the opposite effect is obtained, then we have evidence that the first prime is seen as more male than the second prime. In the current experiment, the four CSs were used as primes and combined with a new set of eight target first names (four male and four female). The participants' task was to determine the gender of the target names as quickly as possible by pressing a left or right key. Note that the priming procedure started with a few practice trials in which two new Japanese characters were combined with another set of six names (three male and three female). We used Affect software to implement the experiment (Hermans, Clarysse, et al., 2002).

Procedure.

The experiment was presented as a study on differences in psycho-physiological reactions when seeing familiar and unfamiliar words and symbols. During acquisition, participants were instructed to just look at words and symbols while a heart-rate device (see Experiment 1) would do the rest.

The experiment started with an acquisition phase in which each of the Japanese characters (CSs) was paired either with a male or with a female name. In accordance with our naming convention, we named the US names after the gender-category to which they belonged, that is, US-male and US-female. Their corresponding CSs will be named CS-male and CS-female, respectively. Which CS was paired with which US was partially counterbalanced: over four conditions each character had been paired once with each US. Each of the four pairs was presented 10 times. During each presentation of a pair the CS was presented for 1 second, immediately followed by the US that also lasted 1 second (ISI of 0). The ITI was 2 seconds.

In the gender-priming procedure, the four CSs plus the two fillers were used as primes together with the four new male and four new female target-names. Participants were instructed to ignore the prime and to quickly categorize the target-name as male or female by pressing a right or left button on the keyboard respectively. Each prime-target combination was presented trice, resulting in three blocks of 32 trials each. After every block there was a short break. The prime-picture was presented for 200 ms. The target name appeared 50 ms after the prime had disappeared (SOA of 250 ms). The target name stayed visible until participants pressed one of the two response keys or 2000 ms had passed. Note that the priming-procedure started with a few practice trials, using the practice stimuli. The practice phase ended when on each of the 12 practice-combinations a valid response was given (trials with no response were re-added at the end of the practice phase until a valid response was given). During the scale measure, participants were asked to indicate the gender of the CSs and also of the two Japanese characters used for the practice trials of the priming measure. The order of these two gender measures was counterbalanced (priming-measure first or scale-measure first).

After the gender-measures, participants also indicated how much they liked the Japanese characters on a valence scale from *very disliked* (= 0) to *very liked* (= 20). The

experiment ended with a contingency awareness measure during which participants had to indicate for each character with what name it had been paired, and were asked a few additional awareness questions. The most important of these questions was whether, during the gender-scale measure, they were consciously influenced by the pairs shown during acquisition. If participants claimed that they consciously had used the CS-US pairings to make their gender-ratings, we will call them ‘demand aware’ in subsequent analysis.

Results

Gender-scale.

For all participants, the estimated gender for the two CS-male and the two CS-female characters were averaged. The average estimated gender for these categories are presented in Table 8. To test our main prediction, we conducted an 2 x 2 ANOVA with CS category (i.e., CS-male versus CS-female) as within-subjects variable, and the order of the measures (i.e., priming-measure first or scale-measure first) as between-subjects variable. Results showed that overall acquisition did have a differential effect on the Japanese characters: the CS-male characters were seen as more male than the CS-female characters, $F(1, 24) = 42.94, p < .0001$, effect size $d = 1.25$. We thus had evidence for a transfer-effect. Although the transfer-effect tended to be slightly larger when the scale-measure came immediately after acquisition (16.3 vs. 3.0, $F(1, 24) = 31.63, p = .0001$, effect size $d = 1.62$), compared to when it came after the priming measure (14.2 vs. 7.0, $F(1, 24) = 12.40, p = .004$, effect size $d = 1.02$), this interaction was only marginally significant, $F(1, 24) = 3.69, p = .07$.

Table 8:

Average Gender-Ratings for the CS-Categories from the Symbol-Word Experiments (standard errors in parentheses)

Experiment	CS-category	Gender Ratings	d
S1	CS-male	15.2 (0.9)	1.25 **
	CS-female	5.0 (0.9)	
	practice primes	10.3 (0.4)	
S2	CS-male	12.6 (0.8)	0.64 **
	CS-female	8.3 (0.7)	
	practice primes	9.3 (0.6)	

Note. d = effect size of the transfer

** $p < .01$

The awareness questions at the end of the experiment, however, showed that 62% of participants claimed that in their assessment of the CS-gender, they had taken into account the gender of the US with which it had been paired during acquisition. This could cast doubts on the validity of the effect in the scale measure. When we divided participants based on their claim that they consciously used the CS-US pairings to make their gender-ratings ('demand aware' vs. 'demand unaware'), we found a marginally significant interaction-effect with CS-category, $F(1, 24) = 4.01, p = .06$. The transfer effect was strongest in the 'demand aware' group (16.1 vs. 3.5, $F(1, 15) = 44.46, p < .0001$, effect size $d = 1.72$), but also still present in the 'demand unaware' group (13.7 vs. 7.4, $F(1, 9) = 5.62, p = .04$, effect size $d = 0.79$).

Priming measure.

In the priming data, we will call the trials on which the CS-category matches the target category (e.g., CS-male prime with a male target name), congruent trials. The trials where the CS-category is the opposite of the target category (e.g., CS-male prime with a female target name), will be called incongruent trials.

Reaction times on trials with an incorrect response were discarded and a cut-off procedure was used to remove outliers. The cut-off procedure was identical to the one used for the priming-data in Chapter 3: for each cell in the design, reaction times that fell above or below 2.5 standard deviations from a person's mean reaction time in that cell were discarded. The average reaction times (over blocks and after cut-off) for the congruent and incongruent trials from the priming-phase are presented in Table 9. To test our main prediction, we conducted a $2 \times 3 \times 2$ ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first, second and third) as within-subjects variable, and the order of the measures (i.e., priming-measure first or scale-measure first) as between-subjects variable. Analysis revealed a significant reaction time difference between congruent and incongruent trials: participants were faster on congruent than on incongruent trials, $F(1, 24) = 4.56, p = .04$, effect size $d = 0.43$. There was no interaction between congruency and block, $F(2, 48) < 1$.¹³ There was also no interaction between congruency and order of the measures, $F(1, 24) < 1$.

¹³ The congruency effect was significant for the first block, $F(1, 25) = 4.96, p = 0.04$, marginally significant for the second block, $F(1, 25) = 3.71, p = 0.09$, and not significant in the third block, $F(1, 25) < 1$. The

We also analysed the percentage of errors or incorrect responses. The average percentage of errors (over blocks) for the congruent, incongruent and non-congruent trials from the priming-phase are also presented in Table 9. The percentage of errors were analysed using a non-parametric Wilcoxon matched pairs test, because the distribution of the percentage of errors diverged strongly from ANOVA criteria (i.e., the errors were distributed geometrically, combined with very few errors per participant and very few differences in errors between participants). Although participants made slightly more mistakes on incongruent trials than on congruent trials, this difference was only marginally significant, $Z(26) = 1.7$, $p = .08$, effect size $d = 0.37$. In neither of the two order conditions and neither of the three blocks did we find a significant difference between congruent and incongruent trials, all Z 's < 1.57 (i.e., all p -values > 0.11).

Table 9:

Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Gender Priming Tasks from the Symbol-Word Experiments (standard errors in parentheses)

Experiment	Trial type	RT	d RT	%Errors	d %Errors
S1	Congruent	519 (12)	0.43 *	1.9 % (0.4 %)	0.37
	Incongruent	527 (13)		3.5 % (0.8 %)	
S2	Congruent	514 (9)	0.20	3.4 % (0.6 %)	-0.21
	Incongruent	516 (9)		2.9 % (0.7 %)	

Note. d = effect size of the transfer

* $p < .05$

Contingency awareness.

The contingency-awareness data were analysed similar to the contingency-awareness data from Chapter 3. The most straightforward way to investigate strict contingency-awareness is to count the number of pairs for which participants selected the correct US. For analyses purpose, we assume each choice is independent of the other choices and that the number of correctly identified pairs is thus binomially distributed when participants guess. Given that there were 4 USs to chose from, we would expect $1/4^{\text{th}}$ of the 4 pairs to be reconstructed correctly if participants guess. These assumptions allow us to later calculate if participants answered better than by chance. To measure the category-awareness, we counted the number of pairs for which participants select a US from the correct category. For analyses

overall significant congruency effect was caused by the first two blocks (congruency effect for the first two blocks: $F(1, 25) = 4.96$, $p = 0.04$). Adding the third block in this analysis did not noticeably increase the congruency effect, $F(1, 25) = 4.49$, $p = 0.04$.

purpose we again assume independency of choices and that the number of correctly selected categories is binomially distributed. Given that there were 4 USs to choose from, we would expect the US-category of $2/4^{\text{th}}$ of the 4 pairs to be guessed correctly. We will report both the contingency- and category-awareness score.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 3.2 of the 4 pairs ($SE = 0.2$). 58 % of all participants perfectly reconstructed each of the four pairs, thus performing significantly above chance ($p = .004$). The category-awareness measure showed that, on average, the correct US-category was selected for 3.5 of the 4 pairs ($SE = 0.2$). 77 % of all participants always selected the right US-category, thus performing (almost) significantly above chance ($p = .062$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis revealed a significant correlation, $r = .69$, $p < .001$, between the category-awareness score and the transfer effect score on the gender-scale, indicating that the more participants were aware of the gender of the name that was paired with a CS, the larger the transfer-effect. To examine this relation further, we divided participants into those who did ($N = 20$) and did not ($N = 6$) perform above chance on the category awareness score. Analyses revealed a significant positive transfer effect for the category aware participants, $F(1, 19) = 82.21$, $p < 0.001$, effect size $d = 2.08$, but not for the other participants, $F(1, 5) = 0.03$, $p = .88$, effect size $d = -0.07$.

We also investigated the effect of contingency awareness on the priming data. There was again a significant correlation, $r = .42$, $p = .03$, between category awareness and the congruency effect in the reaction time data: the higher the category awareness, the larger the congruency effect. Further analysis revealed that for the category aware participants a significant congruency effect was found, $F(1, 19) = 6.61$, $p = .02$, effect size $d = 0.59$, but not for the other participants, $F(1, 5) = 0.05$, $p = .82$, effect size $d = -0.11$. No correlation, $r = .22$, $p = .28$, was found between category awareness and the congruency effect in the percentage of errors.

Valence scale.

There was no significant valence-difference between the CS-male ($M = 11.1$) and the CS-female characters ($M = 12.4$), $F(1, 25) = 2.54$, $p = .12$. As such we have no evidence that evaluative conditioning could have influenced the transfer of gender.

Discussion

We found a significant transfer-effect on the scale-measures. But when looking closer, the effects could be caused by some kind of demand effect. Approximately two thirds of the participants claimed that in their assessment of the CS-gender, they took into account the gender of the US with which it had been paired during acquisition. Moreover, these ‘demand aware’ participants showed a stronger transfer-effect on the scale measure. The effect on the scale measure did not disappear in the more ‘demand unaware’ group, but one could argue that our awareness questions were not sensitive enough to detect all demand aware participants. It is thus possible that the effect in the demand unaware group might have been due to a number of demand aware participants that were not detected as being demand aware. On the other hand, the priming-results, which are believed to be more demand-insensitive (e.g., Hermans, Vansteenwegen, et al., 2002), also showed a congruency-effect seemingly independent of demand awareness. In conclusion, this experiment showed good evidence for associative transfer of non-evaluative stimulus properties. Before making more definite conclusions, we ran an additional experiment in order to investigate whether we could replicate the present effects.

Experiment S2

We replicated Experiment S1, with a few changes. First the gender-priming measure now always came before the gender-scale measure (as opposed to being counterbalanced in Experiment S1). This was done because during the scale measure, participants were informed that the Japanese characters were actually first names of men or women. As a result of these instructions, participants might have consciously inferred that during acquisition they had been presented with the Dutch translation of the Japanese characters. Such conscious inferences are sufficient to induce priming effects (e.g., De Houwer et al.,

1998)¹⁴. Because we now always present the scale measure after the priming measure, these kind of inferences will not be (or at least be less) present at the time of the priming-measure.

Another change was that we reduced the original number of priming blocks from three to two, and added two new priming blocks with a new set of target-names. This was done because in Experiment S1 the congruency-effect tended to diminish on each successive block. Although this reduction was far from significant, adding data from the last block did not increase the overall congruency effect (see footnote 13 on page 73). It is possible that either the transfer-effect extinguished, but it is equally possible that too many trials with the same targets somehow reduced the power of the priming-measure to detect the transfer (e.g., a floor-effect, boredom). This is why we replaced the last priming-block by two new priming blocks with different targets.

Method

Participants.

Twenty-eight first year psychology students (7 men and 21 women) participated in exchange for course credits.

Stimuli and measures.

The same stimuli and measures of Experiment S1 were used, except that another eight first names (four male and four female) were added as targets for the gender-priming measure. These eight new first names replaced the original ones in the last two blocks of the priming procedure.

Procedure.

The procedure was identical to that of Experiment S1, with two exceptions. First, the gender-priming task always came before the gender scale measure (this order was counterbalanced in Experiment S1). Second, while the first two priming blocks were the same as in Experiment S1, the next priming-block was replaced by two blocks that used a

¹⁴ De Houwer et al. (1998) asked participants to learn the translation of Turkish words (in reality non-words). The so-called translations were clearly positive or negative words. In a subsequent priming phase, in which the non-words were used as primes, they found a significant affective priming effect.

second set of eight new target-names. This adjustment was made to counter a possible floor-effect.

Results

Gender-scale.

For all participants, the estimated gender for the two CS-male and the two CS-female characters were averaged. The average estimated gender for these categories are presented in Table 8 (p. 72). To test our main prediction, we conducted an ANOVA with CS category (i.e., CS-male versus CS-female) as within-subjects variable. Results showed that overall acquisition did have a differential effect on the Japanese characters: the CS-male characters was seen as more male than the CS-female characters, $F(1, 27) = 10.94$, $p = .003$, effect size $d = 0.64$.

In this experiment only 39% of participants claimed that in their assessment of the CS-gender, they had taken into account the gender of the US with which it had been paired during acquisition. This could cast doubts on the validity of the effect in the scale measure. When we divided participants based on their claim that they consciously used the CS-US pairings to make their gender-ratings ('demand aware' vs. 'demand unaware'), we found a significant interaction-effect with CS-category, $F(1, 26) = 28.38$, $p < .0001$. The transfer effect was only present in the 'demand aware' group (15.6 vs. 5.0, $F(1, 10) = 37.46$, $p = .0001$, effect size $d = 1.94$). In the 'demand unaware' group there was no difference between the CS-male and CS-female (10.7 vs. 10.4, $F(1, 16) < 1$, effect size $d = 0.09$).

Priming measure.

In the priming data, we will call the trials on which the CS-category matches the target category (e.g., CS-male prime with a male target name), congruent trials. The trials where the CS-category is the opposite of the target category (e.g., CS-male prime with a female target name), will be called incongruent trials.

Reaction times on trials with an incorrect response were discarded. The average reaction times (over blocks and after cut-off) for the congruent and incongruent trials from the priming-phase are presented in Table 9 (p. 74). To test our main prediction, we conducted a 2 x 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials)

and block set (first two blocks and last two blocks) and specific block (first or second block in a block set) as within-subjects variable. Analysis revealed no significant reaction time difference between congruent and incongruent trials, $F(1, 27) = 1.12, p = .30$, effect size $d = 0.20$. There was no interaction between congruency and block set, $F(1, 27) < 1$, or between congruency and specific block, $F(1, 27) < 1$.

We also analysed the percentage of errors or incorrect responses. The average percentage of errors (over blocks) for the congruent, incongruent and non-congruent trials from the priming-phase are also presented in Table 9 (p. 74). The percentage of errors were analysed using a non-parametric Wilcoxon matched pairs test, because the distribution of the percentage of errors diverged strongly from ANOVA criteria. There was no significant difference between the percentage of errors made on incongruent trials than on congruent trials, $Z(28) = 1.1, p = .25$, effect size $d = -0.21$. In neither of the two block set conditions (first two blocks and second two blocks) did we find a significant difference between congruent and incongruent trials, both Z 's < 1 .

Contingency awareness.

The strict contingency-awareness measure showed that, on average, the correct US was selected for 2.3 of the 4 pairs ($SE = 0.3$). 32 % of all participants perfectly reconstructed each of the four pairs, thus performing significantly above chance ($p = .004$). The category-awareness measure showed that, on average, the correct US-category was selected for 2.9 of the 4 pairs ($SE = 0.3$). 54 % of all participants always selected the right US-category, thus performing (almost) significantly above chance ($p = .062$).

We subsequently investigated the influence of contingency awareness on the gender transfer. Analysis revealed a significant correlation, $r = .38, p = .05$, between the category-awareness score and the transfer effect score on the gender-scale, indicating that the more participants were aware of the gender of the name that was paired with a CS, the larger the transfer-effect. To examine this relation further, we divided participants into those who did ($N = 15$) and did not ($N = 13$) perform above chance on the category awareness score. Analyses revealed a significant positive transfer effect for the category aware participants, $F(1, 14) = 9.69, p = .008$, effect size $d = 0.83$, but not for the other participants, $F(1, 12) = 2.57, p = .14$, effect size $d = 0.46$.

We also investigated the effect of contingency awareness on the priming data. There was no significant correlation, $r = -0.11$, $p = .58$, between category awareness and the congruency effect in the reaction time data. No correlation, $r = -0.04$, $p = .83$, was found between category awareness and the congruency effect in the percentage of errors.

Valence scale.

There was no significant valence-difference between the CS-male ($M = 11.2$) and the CS-female characters ($M = 11.7$), $F(1, 27) < 1$. As such we have no evidence that evaluative conditioning could have influenced the transfer of gender.

Discussion

Although we replicated the transfer effect in the scale measure, only participants who had claimed that in their assessment of the CS-gender they had taken into account the gender of the US with which it had been paired during acquisition, showed the effect. As such we are unable to exclude the possibility that the effect in the scale measure was caused by some kind of demand-awareness. Moreover, we were unable to replicate the transfer effect in the (demand-insensitive) priming task. Therefore, in contrast to Experiment S1, transfer effects seemed to depend entirely on demand-like effects. This raises doubts about whether demand-free transfer effects can be observed using this paradigm.

Overall discussion of the symbol-word studies

In two experiments we tried to transfer the gender of clearly male and female first names onto Japanese characters. In both experiments we did find an overall transfer-effect on the gender-scale measure. But a large part of participants in both experiments claimed that in their assessment of the CS-gender they had taken into account the gender of the US with which it had been paired during acquisition. While in Experiment S1 the transfer-effect was largest for these ‘demand aware’ participants, the effect remained in the remaining more ‘demand unaware’ participants. In Experiment S2, however, the effect had disappeared completely for the ‘demand unaware’ participants. As we are unable to exclude the possibility that some of the participants that were classified as ‘demand unaware’, nevertheless had also consciously influenced their CS-assessment towards that

of the US with which it had been paired, we cannot claim a pure gender-transfer effect on the basis of this scale-measure.

In both experiments we also administered a, more demand-insensitive, gender-name priming measure. While in Experiment S1 we did find a congruency effect, we were unable to replicate it in Experiment S2. One possible reason why we did not find a transfer effect in the priming task of Experiment S2 was that the priming task always came before the scale measure whereas in Experiment S1, the order of two measures was counterbalanced. This might be important because during the scale measure, participants were informed that the Japanese characters were actually first names of men or women. As a result of these instructions, participants might have consciously inferred that they learned the Dutch translation of the Japanese characters. Previous research has shown that such inferences are sufficient to induce priming effects (e.g., De Houwer et al., 1998). In Experiment S2, participants were not told that the Japanese characters were names of men or women until after the priming task (because the scale measure always came last). This could be the reason why no priming effect was found in Experiment S2.

To investigate this issue further, we re-analysed the priming data of Experiment S1 to investigate to what extent priming effects depended on the order of the tasks. Although an Order x Congruency ANOVA failed to show an interaction between both variables, $F(1, 24) < 1$, closer inspection of the data revealed that the priming effect size was smaller in magnitude when the priming task came first, effect size $d = 0.57$, than when the scale measure came first, effect size $d = 0.28$.

Because it is also possible that participants who were classified as ‘demand aware’, were more likely to see the CS as a translation of the US-name (and/or better remember the CS-US pairs when seeing the CS as a translation of the US-name), we also investigate to what extent the priming effect depended on the awareness of participants. In Experiment S1 an Awareness x Congruency ANOVA failed to show an interaction between both variables, $F(1, 24) < 1$, but closer inspection of the data revealed that the priming effect size was larger in magnitude for aware participants, effect size $d = 0.49$, than for unaware participants, effect size $d = 0.33$. In Experiment S2 the same analysis also failed to show an interaction between both variables, $F(1, 26) = 1.27$, $p = .27$, but closer inspection of the

data revealed that the priming effect size was larger in magnitude for aware participants, effect size $d = 0.48$, than for unaware participants, effect size $d = 0.03$.

To investigate the separate and combined effect of both the order of measures and awareness of participants, we broke down the reaction time data of Experiment S1 and Experiment S2 on both variables. Table 10 shows the resulting reaction times and effect size's for each group of participants. The pattern was in the predicted direction in both experiments: the effect tended to be larger a) if the scale measure came first (only applicable in Experiment S1) and b) if participants were aware. Both factors seemed somewhat cumulative. Although this pattern is statistically not significant (possibly because of low power), these data are consistent with the idea that transfer effects in the priming task might have depended on conscious inferences. More research is of course needed before definite conclusion can be drawn regarding this issue.

Table 10:

Average Reaction Times (RT) for the Gender Priming Task, for each Order Condition and Awareness Group from Experiments S1 and S2 (standard errors in parentheses).

Experiment	Measure first	Awareness	N	Trial type	RT	d RT
S1	Priming	Unaware	5	Congruent	503 (11)	0.15
				Incongruent	506 (7)	
		Aware	8	Congruent	518 (28)	0.37
				Incongruent	525 (32)	
	Scale	Unaware	5	Congruent	509 (30)	0.50
				Incongruent	520 (29)	
		Aware	8	Congruent	534 (20)	0.60
				Incongruent	548 (22)	
S2	Priming	Unaware	17	Congruent	518 (12)	0.03
				Incongruent	519 (11)	
		Aware	11	Congruent	506 (15)	0.48
				Incongruent	512 (15)	

Note. Measure first = whether the *priming* measure came first or the *scale*-measure came first for this group of participants. Awareness = group of participants that were either classified as *aware* or *unaware*. d = effect size of the transfer

The question remains, however, whether we had observed associative transfer of gender. Even if the effect only appears when participants make these kind of conscious inferences, one could claim that the effect is still caused by an association. But even though these inferences seem to be made 'voluntarily' by participants, they will only happen in very specific circumstances (e.g., when presented with a symbol and a word that can be perceived as being meaningfully related). The effect will not be as widespread as

EC, which does not seem to require these inferences and where simple associations seem enough.

In conclusion, we only seemed to observe clear transfer of gender for participants who consciously took into account the gender of the US with which it had been paired during acquisition. We have little evidence for gender-transfer caused by mere association (i.e., without conscious inferences).

Chapter 5

Fear Conditioning Studies

Due to the difficulty in consistently finding clear ATNESP, we also started a third line of experiments. The approach of this third paradigm differs somewhat from the picture-picture and symbol-word designs used in the previous two chapters. Remember that the referential learning account of EC implies that the valence transfer is caused by misattributing the activated US-valence to the CS (e.g., Baeyens, Eelen, & Crombez, 1995). One could discern two sub-processes in this explanation: for EC to occur, a) the CS must activate the US-valence, and b) the activated US-valence must be misattributed to the CS. It is possible that, while activated US-valence is easily misattributed to the CS, activated non-evaluative properties of a US are not that easily misattributed. In this chapter we used a fear-conditioning paradigm to investigate whether pairing a CS with a US can cause non-evaluative properties of the US to be subsequently activated by the CS (i.e., without necessarily ‘transferring’ to the CS). We are thus not really searching for associative *transfer* of non-evaluative stimulus properties, but more for the sub-process of associative *activation* of non-evaluative stimulus properties.

The use of a fear-conditioning paradigm, that is most frequently used to investigate expectancy learning, has several advantages. First, fear-conditioning designs have successfully been used to research EC and have produced robust and easily replicable EC effects (e.g., Hermans, Vansteenwegen, et al., 2002). Second, in our design, we are able to check evaluative transfer/activation and non-evaluative activation at the same time. The advantage of this is that if we do find EC but no associative activation of non-evaluative stimulus properties, then the absence of non-evaluative transfer cannot be due to the fact that the procedure does not support any form of associative transfer.

Experiment F1

In all the experiments reported in this chapter, participants were exposed to a contingency between the colour (CS) of a word and the valence (aversive loud noise present vs. no noise) and direction (loud noise from left vs. right) of a US. We expected that, if

associative activation occurred, the colour that had been paired with a loud noise from the left or right would subsequently activate ‘left’ or ‘right’ respectively (non-evaluative activation)¹⁵. At the same time, we would expect that the colour that had not been paired with an aversive noise would become more positive than the colours that had paired with a loud noise (evaluative transfer). To assess this affective and spatial activation, we used associative variants of the Simon paradigm.

Method

Participants.

Thirty-six first year psychology students (5 men and 31 women) participated in exchange for course credits in this experiment.

Stimuli and measures.

Ten animal (e.g., beaver) and ten profession names (e.g., baker) were used in the Simon test phases. An additional two animal and two profession names were used for the practice trials of the Simon phases. Another 10 words that belonged to neither of those categories were used during acquisition. All words were presented in three colours: pastel-blue, pastel-red, and pastel-yellow. As aversive US, a loud white-noise burst of 95 dB(A) was presented for 500 ms through a headphone. This US could either be presented to the left or right ear.

To measure activation of valence or direction, we used associative variants of affective and spatial Simon-procedures. In a standard affective Simon task (e.g., De Houwer & Eelen, 1998; De Houwer, Crombez, Baeyens, & Hermans, 2001), participants are to select between a positive and negative response on the basis of a non-affective stimulus feature (i.e., relevant stimulus feature) while ignoring the valence of the presented stimuli (i.e., irrelevant stimulus feature). Despite the fact that participants should ignore the affective meaning of the presented stimuli, reaction times are faster when the affective connotation of the presented word and the correct response are the same (congruent) than when they differ (incongruent). Take for example Experiment 1 of De Houwer, Crombez, et al. (2001). Participants were presented with words and instructed to say “Positive” when

¹⁵ Note that a colour cannot be ‘left’ or ‘right’. In this case we thus can not expect property transfer, only property activation.

the word referred to an animal and “Negative” when it referred to a person. These words were either positive or negative. Here the semantic category was the relevant feature while the valence of the word was the irrelevant feature. Results showed that, in this task, participants were faster in saying “Positive” to a positive animal (e.g., butterfly) than to a negative animal (e.g., cockroach). Likewise, they were slower in saying “Negative” to a positive person (e.g., friend) than to a negative person (e.g., enemy). Therefore, if we observe that a stimulus facilitates a positive or negative response, then we have evidence that this stimulus is associated with a positive or negative valence, respectively.

In a standard spatial Simon task (e.g., Simon, 1990), participants are asked to respond to a non-spatial stimulus feature (i.e., relevant stimulus feature) by pressing a left or right key, while ignoring the spatial position of the presented stimuli (i.e., irrelevant stimulus feature). Spatial correspondence between stimulus-position and correct response results in faster responding as compared to non-correspondence. Therefore, if we observe that a stimulus facilitates a left or right response, then we have evidence that this stimulus is associated with a left or right location respectively.

A valence-scale from *very disliked* (= 0) to *very liked* (= 20) was also used to measure participants’ liking of the CS-colour. These experiments were run on an IBM-compatible computer using the computer software Affect (Hermans, Clarysse, et al., 2002).

Procedure.

The experiment was presented as a study that tried to measure valence and emotion through psycho-physiological measures. Participants were told that they would be presented unpleasant, but not painful, short loud tones through a pair of headphones. Participants were also informed that they could stop the experiment at any moment if they found the loud tones painful. If still willing to participate they were asked to sign the ‘informed consent’ form and put on the headphones. During acquisition, participants were instructed to just look at the stimuli while a heart-rate device (see Experiment P1) would do the rest.

During acquisition, words were presented in three colours. When a word was presented in the first colour (CS-right), a loud noise was presented in the right ear. When a word was presented in the second colour (CS-left), a loud noise was presented in the left

ear. When the word was presented in the third colour (CS-neutral), no loud noise followed. Which CS-colour was paired with which US was counterbalanced (resulting in 6 colour conditions). Words were presented for 2500 ms and if a US was presented, it was during the last 500 ms of the word presentation. The inter trial interval (ITI) was 4000 ms. Each of the ten words was presented twice in each of the three colours, resulting in 60 acquisition trials (of which 40 included a US-sound). We predicted that, if EC occurred, the CS-left and CS-right would become more negative than the CS-neutral, because the former were always paired with an aversive loud noise¹⁶. If associative activation of non-evaluative stimulus properties occurred, we would expect that the CS-left would activate the representation of ‘left’ and the CS-right the representation of ‘right’.

After acquisition, the headphones were removed and the associative Simon procedures were run ¹⁷. In this experiment, the spatial Simon task came before the affective Simon task. In the spatial Simon task, participants again saw words on the screen. Each word referred either to a profession or an animal. Participants were instructed to press a button on the left or right of the keyboard as fast as possible, based on the category of the word (relevant feature). These words, however, were also presented in one of the three colours (irrelevant feature) that had been used during acquisition. If the CS-colour activates the direction from which a US came, we would expect a compatibility effect in the spatial Simon task. That is, responses should be faster when the colour of the word was associated with the same position as the correct response, than when the colour and response referred to different locations (see Figure 1 for a schematic example). Note that whether participants had to press the left/right key on an animal/profession trial was counterbalanced. The logic underlying the associative affective Simon task was very similar. Again profession- or animal-words were used. This time participants were instructed to say “Positive” or “Negative” based on the category of the word (relevant feature). Words were presented in the three colours (irrelevant feature). If the CS-colour activates the valence of the US, we expect a compatibility effect in the affective Simon. That is, responses should be faster when the irrelevant colour of the word was associated

¹⁶ Note that we expect either a) the CS-neutral to become a predictor of the absence of a loud noise (i.e., a safety signal), thus becoming positive or b) the CS-neutral just staying more or less neutral, but compared to the CS-left or CS-right becoming more positive.

¹⁷ We decided to remove the headphones after acquisition because during the affective Simon participants had to use a voice key. To correctly use a voice key, participants should be able to hear themselves clearly.

with the same valence as the correct response, than when the colour and response referred to different valences (see Figure 2 for a schematic example). For both Simon tasks, 10 animal and 10 profession words were combined with the three CS-colours. These 60 combinations formed a block and the blocks were presented twice, separated by a short break. Note that each of the two Simon tasks started with a few practice trials that used the four practice words (in each of the three colours). The practice phase ended when on each of the 12 practice-combinations a valid response was given (trials with no response or a voice-key failure were re-added at the end of the practice phase until a valid response was given).

After the two Simon tasks, participants were asked to express their liking of the CS-colours, presented as a row of X's in that colour, on the valence-scale. Next, a contingency-awareness test was performed, in which participants again saw a row of X's in each of the three CS-colours. Participants were told that at the beginning of the experiment each of these colours had always been followed by either a loud noise from the left, a loud noise from the right or no noise at all. For each colour participants guessed if a US was paired with it during acquisition, and if so which US (left or right). The experiment ended with a short questionnaire about the aversive noise (i.e., how much participants liked/disliked the loud noise, how intense they found the loud noise, and how much the noise startled them).

Figure 1:

Schematic Example of Successful Associative Spatial Simon.

Note. In this example blue is CS-left, red is CS-right and yellow is CS-neutral. Participants have to press a left key on animal names and a right key on profession names.

Figure 2:

Schematic Example of Successful Associative Affective Simon.

Note. In this example blue is CS-left, red is CS-right and yellow is CS-neutral. Participants have to say “Negative” on animal names and “Positive” on profession names.

Results

Affective Simon

In the affective Simon task, the trials on which the valence of the CS-category matches the valence of the response (e.g., CS-left or CS-right colour when having to say “Negative”), will be called congruent trials. The trials on which the valence of the CS-category is opposite to the valence of the response (e.g., CS-neutral colour when having to say “Negative”), will be called incongruent trials.

Table 11:

Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Affective Simon Tasks from the Fear Conditioning Experiments (standard errors in parentheses)

Experiment	Trial type	RT	d RT	%Errors	d %Error
F1	Congruent	624 (12)	0.01	0.5 % (0.2 %)	0.13
	Incongruent	625 (12)		0.8 % (0.3 %)	
F2	Congruent	595 (8)	0.24	0.6 % (0.2 %)	0.02
	Incongruent	599 (9)		0.6 % (0.2 %)	

Note. d = effect size of the activation

One participant had misunderstood the instruction of the affective Simon, and was excluded from the analysis. Note that 3.6 % of the affective Simon trials were categorized as voice key failures and excluded from the analysis. In addition, reaction times on trials with an incorrect response were discarded and a cut-off procedure was used to remove outliers. The cut-off procedure was identical to the one used for the priming-data in Chapter 3: for each cell in the design, reaction times that fell above or below 2.5 standard deviations from a person’s mean reaction time in that cell were discarded. The average reaction times (over blocks and after cut-off) for the congruent and incongruent trials in the affective Simon task, are presented in Table 11. To investigate whether affective activation had occurred, we conducted a 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variables. Analysis revealed no significant reaction time difference between congruent and incongruent trials, $F(1, 34) < 1$, effect size $d = 0.01$, and thus no evidence for EC. There was also no significant interaction between contingency and block, $F(1, 34) < 1$.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent and incongruent trials in the affective

Simon task, are also presented in Table 11. The percentage of errors were analysed using a non-parametric Wilcoxon matched pairs test, because the distribution of the percentage of errors diverged strongly from ANOVA criteria (i.e., the errors were distributed geometrically, combined with very few errors per participant and very few differences in errors between participants). The matched pairs test revealed no significant difference between the percentage of errors made on congruent versus incongruent trials, $Z(35) = 0.63$, $p = .59$, effect size $d = 0.13$, and thus no evidence for EC. There was also no interaction between congruency and block, $Z(35) = 0.47$, $p = .64$.

Spatial Simon

In the spatial Simon task, the trials on which the direction of the CS-category matches the direction of the response (e.g., CS-left when having to press the left key), will be called congruent trials. The trials on which the direction of the CS-category is opposite to the direction of the response (e.g., CS-left when having to press the right key), will be called incongruent trials. Trials which were in the CS-neutral colour will be called non-congruent.

Table 12:

Average Reaction Times (RT) and Percentage of Errors (%Errors) for the Spatial Simon Tasks from the Fear Conditioning Experiments (standard errors in parentheses)

Experiment	Trial type	RT	d RT	%Errors	d %Errors
F1	Congruent	593 (18)	-0.07	2.6 % (0.5 %)	-0.46 *
	Incongruent	591 (16)		1.4 % (0.4 %)	
	Noncongruent	602 (18)		1.9 % (0.4 %)	
F2	Congruent	558 (10)	-0.15	3.6 % (0.5 %)	-0.30 *
	Incongruent	553 (8)		2.3 % (0.4 %)	
	Noncongruent	556 (10)		2.8 % (0.4 %)	
F3	Congruent	592 (10)	0.20	1.7 % (0.4 %)	0.21
	Incongruent	598 (10)		2.5 % (0.5 %)	
	Noncongruent	602 (11)		2.2 % (0.4 %)	
F4	Congruent	589 (18)	-0.03	2.5 % (0.6 %)	-0.06
	Incongruent	588 (17)		2.3 % (0.6 %)	
	Noncongruent	586 (16)		2.1 % (0.4 %)	
F5	Congruent	531 (11)	-0.32	5.1 % (1.2 %)	0.17
	Incongruent	524 (10)		6.0 % (1.2 %)	
	Noncongruent	534 (13)		4.6 % (1.0 %)	

Note. d = effect size of the activation

* $p < .05$

Just like in the affective Simon, reaction times of trials with an incorrect response were discarded and the cut-off procedure was used to remove outliers. The average reaction times (over blocks and after cut-off) for the congruent, incongruent and non-congruent trials in the spatial Simon task, are presented in Table 12. To investigate whether spatial activation had occurred, we conducted a 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variables. Analysis revealed no significant reaction time difference between congruent and incongruent trials, $F(1, 35) < 1$, effect size $d = -0.07$ and thus no evidence for spatial activation. There was no interaction between congruency and block, $F(1, 35) < 1$.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent, incongruent and noncongruent trials in the spatial Simon task, are also presented in Table 12. The percentage of errors were analysed using a non-parametric Wilcoxon matched pairs test, because the distribution of the percentage of errors diverged strongly from ANOVA criteria. The matched pairs test revealed a significant difference between the percentage of errors made on congruent and incongruent trials, $Z(36) = 2.27$, $p = .023$, effect size $d = -0.46$. This effect was in the opposite direction as expected: participants made more mistakes on congruent trials than on incongruent trials. So when presented with a CS-colour, participants made more mistakes when having to press a key on the side associated with that CS-colour (i.e., congruent trial) than when having to press a key opposite to the side associated with the CS-colour (i.e., incongruent trial). There was no interaction of this contrast-effect with block, $Z(36) = 1.57$, $p = .12$. If anything, the contrast was slightly clearer in the second block, $Z(36) = 2.30$, $p = .02$, than in the first block, $Z(36) = 1.36$, $p = .17$.

Valence-scale.

We also investigated whether the expressed liking for the two colours that had been paired with an aversive noise (i.e., CS-left and CS-right) had become more negative than the colour that was never followed by an aversive noise (i.e., CS-neutral). The average valence ratings for the CS-left, CS-right and CS-neutral colours are presented in Table 13. A planned comparison revealed no significant liking difference between the colours that had been paired with an aversive noise and the colour that had not, $F(1, 35) = 1.90$, $p = .17$, effect size $d = 0.24$, and thus no evidence for EC.

Table 13:

Average Valance-Ratings for the CS-Categories from the Fear Conditioning Experiments (standard errors in parentheses)

Experiment	CS-category	Valence Ratings	<i>d</i> Rating
F1	CS-left	11.9 (0.7)	0.24
	CS-right	11.4 (0.8)	
	CS-neutral	13.0 (0.7)	
F2	CS-left	11.3 (0.7)	0.06
	CS-right	11.9 (0.8)	
	CS-neutral	12.0 (0.7)	
F3	CS-left	11.1 (0.8)	0.41 *
	CS-right	10.6 (0.8)	
	CS-neutral	13.7 (0.8)	
F4	CS-left	10.3 (1.0)	0.32
	CS-right	10.3 (1.1)	
	CS-neutral	12.5 (1.1)	
F5	CS-left	10.4 (0.8)	0.58 *
	CS-right	10.8 (1.1)	
	CS-neutral	14.3 (0.9)	

Note. *d* = effect size of the transfer

* $p < .05$

Contingency-awareness.

On average 2.83 of the 3 colours were correctly linked to their US. Of all participants, 92 % perfectly reconstructed each of the three colour-sound pairs¹⁸.

Discussion

In this experiment we tried to associatively activate the valence and direction of an aversive US by CS-colours. In the spatial Simon task, we did find some evidence that an association had been formed between the CS-colour and the direction from which the US came. Although we did not find an effect in the reaction time data, we did find a contrast effect in the percentage of errors: contrary to expectations, participants made *more* mistakes on congruent trials than on incongruent trials. It seemed that a colour previously associated with an aversive noise from one side, subsequently slightly facilitated

¹⁸ Due to the high percentage of perfectly contingency aware participants in this and the other experiments in this chapter (always more than 80%), no subsequent analyses with contingency-awareness was performed in these experiments.

(incorrectly) pressing a key in the opposite location. The direction of this contrast-effect is hard to interpret. One could utter that the colour activated the memory representation of the loud noise that came from one side, and that participants wanted to ‘avoid’ this direction and thus tended to press a key in the opposite location. But if we indeed see some kind of ‘avoidance behaviour’, then the effect could be caused by expectancy learning. With expectancy learning we mean that participants simply learn that the CS predicts the US, and the CS subsequently causes a preparatory reaction. If this is the case, we would expect the effect to extinguish during the spatial Simon task (in which the CS-colours are presented without the US). Closer analysis of the data however, showed that the effect did not diminish from the first to the second spatial Simon block. If anything, the contrast effect tended to strengthen in the second block. Before speculating further about this effect, we attempted to replicate it. For the next experiment, we also decided to counterbalance the order of the spatial and affective Simon task. If expectancy learning is responsible for the observed spatial contrast effect, we would expect that it would extinguish if the spatial Simon is run after the affective Simon.

Contrary to expectations we found no evidence for evaluative transfer/activation: neither the valence-scale nor the affective Simon measure showed evidence that colours that had been paired with an aversive noise had become more negative than a colour that had not. It is unclear why we did not find evaluative conditioning. Even if our affective Simon procedure was not fitted to measure associatively induced valence, we would still expect an effect on the valence-rating scale. One possibility is that our loud noise (US) was not aversive enough, but we have no data that can support or refute this hypothesis. In our next experiment we made the aversive US louder, in the hope of making it more negative.

Experiment F2

In the previous experiment we did find some evidence for spatial activation. But contrary to expectations, the spatial Simon revealed a contrast effect: It seemed that a colour previously associated with an aversive sound from one side, subsequently slightly facilitated (incorrectly) pressing a key in the opposite location. It is hard to explain this associative effect. To explore whether the associative effect can be explained by pure expectancy learning, we decided to replicate the experiment while adding a second

condition in which the affective Simon would be administered before the spatial Simon. If the spatial contrast effect is solely caused by expectancy learning, the effect should disappear after the affective Simon trials. This because each Simon trial can be seen as an extinction trial (i.e., the CS is presented without the US). A second small change in this experiment was that the loud noises (USs) were made louder, in the hope of making them more negative and increasing the chance of observing evaluative conditioning.

Method

Participants.

Forty-seven first year psychology students (11 men and 36 women) participated in exchange for course credits.

Stimuli and measures.

The same stimuli as in Experiment F1 were used. The only change was that the aversive USs were now 98 dB(A) instead of 95dB(A). The measures were identical to the ones in Experiment F1.

Procedure.

The first condition in this experiment, which we will call the ‘spatial Simon first’ condition, was identical to the condition used in Experiment F1. In the second condition, which we will call the ‘spatial Simon last’ condition, we switched the order of the spatial and affective Simon so that the spatial Simon came after the affective Simon. All other elements of the procedure were the same as in the first condition.

Results

Affective Simon

Note that 3.4% of the affective Simon trials were categorized as voice key failures and excluded from analysis. The average reaction times (over blocks and after cut-off) for the congruent and incongruent trials in the affective Simon tasks, over the two order conditions (i.e., ‘spatial Simon first’ and ‘spatial Simon last’), are presented in Table 11. To investigate whether affective activation had occurred, we conducted a 2 x 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variables, and order (‘spatial Simon first’ and ‘spatial Simon last’) as

between-subjects variable. Although participants were overall slightly faster on congruent than on incongruent trials, this congruency-effect was not significant, $F(1, 45) = 2.62, p = .11$, effect size $d = 0.24$. There was however a significant interaction of congruency with block, $F(1, 45) = 4.60, p = .04$. Closer inspection of this interaction revealed that there was a significant congruency effect in the first block (598 versus 608 on congruent and incongruent trials respectively, $F(1, 46) = 7.08, p = .01$) that had disappeared in the second block (591 versus 590 on congruent and incongruent trials respectively, $F(1, 46) < 1$). Finally, the order in which the Simon phases were presented, had no significant influence on the congruency-effect, $F(1, 45) < 1$. In short, we only found evidence for evaluative conditioning in the first block of the affective Simon procedure.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent and incongruent trials in the affective Simon tasks, over the two order conditions (i.e., ‘spatial Simon first’ and ‘spatial Simon last’), are also presented in Table 11 (p. 91). A matched pairs test revealed no significant difference between the percentage of errors made on congruent versus incongruent trials, $Z(47) = 0.28, p = .78$, effect size $d = 0.02$, and thus no evidence for EC. The order in which the Simon phases were presented seemed to have no influence: in both order-conditions there was no significant effect (both Z ’s < 1). There was no interaction between congruency and block, $Z(47) = 1.13, p = .26$.

Spatial Simon

Due to a computer problem, data from the spatial Simon of one student was not recorded. The average reaction times (over blocks and after cut-off) for the congruent, incongruent and non-congruent trials in the spatial Simon tasks, over the two order conditions (i.e., ‘spatial Simon first’ and ‘spatial Simon last’), are presented in Table 12 (p. 92). To investigate whether spatial activation had occurred, we conducted a $2 \times 2 \times 2$ ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variables, and order (‘spatial Simon first’ and ‘spatial Simon last’) as between-subjects variable. Analysis revealed no significant reaction time difference between congruent and incongruent trials, $F(1, 44) = 1.07, p = .31$, effect size $d = -0.15$, and thus no evidence for spatial activation. The order in which the Simon phases were presented, had no significant influence on the congruency, $F(1, 44) < 1$. Finally there was no interaction between congruency and block, $F(1, 44) < 1$.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent and incongruent trials in the spatial Simon tasks, over the two order conditions (i.e., ‘spatial Simon first’ and ‘spatial Simon last’), are also presented in Table 12 (p. 92). We were able to replicate our spatial contrast effect: participants made significantly more mistakes on congruent trials than on incongruent trials, $Z(46) = 1.99$, $p = .05$, effect size $d = 0.30$. The order in which the Simon phases were presented seemed to have an influence on this contrast effect: only when the spatial Simon came immediately after acquisition did we observe the contrast effect (3.7 % versus 1.3 % errors), $Z(23) = 3.00$, $p = .003$, effect size $d = -0.82$. When the spatial Simon came after the affective Simon, there was no difference between congruent and incongruent trials (3.5 % versus 3.4 % errors), $Z(23) = 0.07$, $p = .94$ effect size $d = -0.02$. There was no interaction of congruency with block, $Z(46) = 0.79$, $p = .43$, however.

Valence-scale.

We also investigated whether the expressed liking for the two colours that had been paired with an aversive noise (i.e., CS-left and CS-right) had become more negative than the colour that had never been followed by an aversive noise (i.e., CS-neutral). The average valence ratings for the CS-left, CS-right and CS-neutral colours are presented in Table 13 (p. 94). A planned comparison revealed no significant liking difference between the colours that had been paired with an aversive noise and the colour that had not, $F(1, 46) < 1$, effect size $d = 0.06$, and thus no evidence for EC.

Contingency-awareness.

On average 2.57 of the 3 colours were correctly linked to their US. Of all participants, 81 % perfectly reconstructed each of the three colour-sound pairs.

Discussion

In this experiment we succeeded in replicating the spatial contrast effect observed in Experiment F1: a colour previously associated with an aversive sound from one side, subsequently slightly facilitated (incorrectly) pressing a key in the opposite location during a spatial Simon procedure. To further explore this phenomenon, we had manipulated the order in which the affective and spatial Simon were presented. If the contrast-effect disappeared when the Spatial Simon did not immediately follow acquisition (i.e., when the

affective Simon was run between acquisition and the spatial Simon), then we have some evidence that the associative contrast-effect is sensitive to extinction, as a Simon phase can be seen as a kind of extinction procedure. Analysis revealed that the presentation-order of the Simons indeed had this effect: only when the spatial Simon immediately followed acquisition did we observe the spatial contrast effect.¹⁹ Alternative explanations for this extinction effect are possible however: maybe there had never been a spatial effect in the group that first had the affective Simon, or maybe too many Simon trials reduce the effectiveness of the Simon as a measure (e.g., due to boredom, practice with the Simon,...). More research is necessary to disentangle these hypotheses.

We also tried to obtain EC in this experiment. Compared to the previous experiment, we had made the US louder, in the hope of making it more negative. The affective Simon task revealed no significant difference between the colour that had been paired with an aversive noise and the colours that had not. Closer analysis, however, revealed that there was a significant effect, but only in the first block. This suggests that there was indeed some form of EC, but this acquired valence seemed sensitive to extinction. This extinction could also explain why we found no EC on the valence scale at the end of the experiment. On the other hand, there was no interaction between the congruency effect and the presentation order: if the observed activation was sensitive to extinction, we would have expected it to be smaller in the condition in which the spatial Simon came before the affective Simon. More research is necessary to determine whether we observed real EC and if so, whether it is sensitive to extinction. But because the main focus of this dissertation is on associative activation of *non*-evaluative stimulus properties, we will not pursue this any further.

Experiment F3

If we wanted to investigate the role of extinction on the spatial contrast effect obtained in the spatial Simons of Experiments F1 and F2, we would have to add an extinction phase

¹⁹ This result was somewhat contrary to the results of Experiment F1: in that experiment (where the spatial Simon always followed immediately after acquisition) we did not observe a decrease of the contrast effect from the first to the second block of the spatial Simon: there even was a slight (albeit non-significant) increase. Note that in the second block already 20 of the CS-left/right trials had been presented without US, so partial extinction should have already happened.

and another Simon phase. This would seriously prolong the experiment, and might reduce the attention of participants. For this reason, we first attempted to reduce the timing parameters of the conditioning experiment. We also dropped the affective Simon procedure because, as already mentioned, our focus was not on EC. To further simplify the design, we also reduced the number of colour counter-balancing conditions.

Method

Participants.

Forty-two first year psychology students (8 men and 34 women) participated in exchange for course credits in this experiment.

Stimuli and measures.

The same stimuli and measures as in Experiment F2 were used, except that the affective Simon was dropped.

Procedure.

The procedure was identical to the one used in Experiment F2, with the following exceptions. First, we reduced the number of counterbalancing conditions for the colours from 6 to 3 in such a way that each colour still served as frequent for each CS-category as each other colour. Second, we reduced some of the timing-parameters of acquisition: CS-presentation were shortened by 500 ms: words were now presented for 2000 ms and if a US was presented, it was during the last 500 ms. We also reduced the ITI of acquisition from 4000 ms to 3000 ms.

Results

Spatial Simon

Due to a computer problem, data from the spatial Simon of one student was not recorded. The average reaction times (over blocks and after cut-off) for the congruent, incongruent and non-congruent trials in the spatial Simon tasks are presented in Table 12 (p. 92). To investigate whether spatial activation had occurred, we conducted a 2 x 2 ANOVA with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variables. Analysis did not reveal a significant reaction time difference between congruent and incongruent trials, $F(1, 41) = 1.68$, $p = .20$, effect size $d = 0.20$, and

thus no evidence for spatial activation. There was, however, a significant interaction between congruency and block, $F(1, 41) = 4.68, p = .04$: while in the first block there was no effect (623 versus 617 respectively, $F < 1$), in the second block participants were significantly faster on congruent trials ($M = 562$) than on incongruent trials ($M = 578$), $F(1, 41) = 11.07, p = .002$, effect size $d = 0.52$.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent and incongruent trials in the spatial Simon task, are also presented in Table 12 (p. 92). We were unfortunately unable to replicate our spatial contrast effect: participants made no more mistakes on congruent trials than on incongruent trials, $Z(42) = 1.24, p = .22$, effect size $d = 0.21$. There was no interaction of congruency with block, $Z(42) = 0.54, p = .59$.

Valence-scale.

We also investigated whether the expressed liking for the two colours that had been paired with an aversive noise (i.e., CS-left and CS-right) had become more negative than the colour that was never followed by an aversive noise (i.e., CS-neutral). The average valence ratings for the CS-left, CS-right and CS-neutral colours are presented in Table 13 (p. 94). Contrary to previous experiments, a planned comparison revealed a significant difference: the colours that had been paired with an aversive noise had become significantly more negative than the colour that had never been paired with an aversive noise, $F(1, 41) = 7.05, p = .01$, effect size $d = 0.41$.

Contingency-awareness.

On average 2.90 of the 3 colours were correctly linked to their US. Of all participants, 93 % perfectly reconstructed each of the three colour-sound pairs.

Discussion

The purpose of this experiment was to replicate the spatial contrast results of Experiment F1 and F2 with a shortened procedure. Contrary to expectations, the contrast effect in the percentage of errors in the spatial Simon had disappeared. Moreover, we even found a significant spatial congruency-effect in the reaction times of the second block in the spatial Simon task. We cannot explain this sudden congruency-effect in the reaction times, or why

it only appeared in the second block. The only difference with the (spatial Simon first condition of the) previous experiment was the timing parameters. This suggests that the direction of the results (contrast effect versus congruency effect) depend, for some unknown reason, on these timing-parameter. We decided to run another replication of the fear-condition effect before speculating further about these results. For this replication we again used our original, slightly longer, timing parameters. As our primary interest was whether the associative effect was sensitive to extinction, we also added an extinction procedure.

Another discrepancy between the results of the present experiment and those of the previous two experiments was that we found evidence for EC on the valence-scale: colours that were paired with an aversive US had become more negative than the colour that was not paired with a US. This experiment however, was also the first one in which the valence scale was administered after only one Simon measure (because the affective Simon was omitted). If precisely this absence of a Simon (which could be seen as an extinction phase) is responsible for the appearance of significant EC, one could suggest that in the other experiments the EC had already extinguished by the time of the valence scale. But more evidence is necessary to support this hypothesis.

Experiment F4

In this experiment we wanted to further investigate whether the spatial contrast effect obtained in the spatial Simons of Experiments F1 and F2, is indeed sensitive to extinction. We exactly replicated (the spatial Simon first condition of) Experiment F2 up until the spatial Simon measure. After the spatial Simons we added an extinction-phase, which was identical to acquisition but without any US-presentations. Following extinction, a second spatial Simon was run. Note that the affective Simon was omitted, to save time and not flood participants with large amounts of Simon trials. Note that we did not use the shortened acquisition timings and reduced number of colour-counterbalancing conditions of Experiment F3, because we failed to find a spatial contrast effect there.

Method

Participants.

Twenty-four first year economy students (12 men and 12 women) participated in exchange for course credits in this experiment.

Stimuli and measures.

The same stimuli and measures as in Experiment F3 were used. In addition another set with 10 new animal and 10 new profession words were added for the post-extinction Simon procedure. We used a new set of words for this second Simon phase, because extended training might suppress the associative effect.

Procedure.

The procedure was identical to the one used in Experiment F2, with the following exceptions. First, the affective Simon measure was omitted. Second, after the (post-acquisition) spatial Simon an extinction procedure was added. This phase was identical to acquisition, but without the presentation of any US-sound.²⁰ After the extinction phase a second spatial Simon was run, but with a different set of animal and profession words. In all other respects this post-extinction spatial Simon was identical to the post-acquisition Simon. Note that which set of words was used for the first or second spatial Simon was counterbalanced.²¹ After the second Simon came the valence scale and the contingency-awareness measure.

Results

Spatial Simon

The average reaction times (over blocks and after cut-off) for the congruent, incongruent and non-congruent trials in the *first* spatial Simon task are presented in Table 12 (p. 92). To investigate whether spatial activation had occurred, we conducted a 2 x 2 ANOVA on the data from the first spatial Simon with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variables. Analysis revealed no significant

²⁰ Note that for the extinction phase participants were again required to put the headphones on as we wanted to keep the acquisition and extinction context as similar as possible.

²¹ In this experiment there were 24 counterbalancing conditions in total: 6 for colour to CS assignment, 2 for animal/profession to left/right response key, and 2 for which set of words for the first/second Simon.

reaction time difference between congruent and incongruent trials, $F(1, 23) < 1$, and thus no evidence for spatial activation. There was also no interaction between congruency and block, $F(1, 12) < 1$, effect size $d = -0.03$. When looking at the results of the *second* spatial Simon (after extinction), we also found no significant reaction time difference between congruent ($M = 587$) and incongruent trials ($M = 591$), $F(1, 23) < 1$, effect size $d = 0.17$.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent and incongruent trials in the *first* spatial Simon task, are also presented in Table 12 (p. 92). We were again unable to replicate our spatial contrast effect: participants did not make more mistakes on congruent trials than on incongruent trials, $Z(24) = 0.22$, $p = .83$, effect size $d = -0.06$. There was however a significant interaction between congruency and block, $Z(24) = 1.97$, $p = .05$: the first block tended towards a congruency effect ($Z(24) = 1.54$, $p = .12$) while the second block tended towards a contrast effect ($Z(24) = 1.34$, $p = .18$). But no effect was significant. When looking at the results of the *second* spatial Simon (after extinction), we found a significant contrast effect in the percentage of errors made on congruent ($M = 2.6\%$) versus incongruent trials ($M = 1.4\%$), $Z(24) = 2.31$, $p = .02$, effect size $d = -0.55$.

Valence-scale.

We also investigated whether the expressed liking for the two colours that had been paired with an aversive noise (i.e., CS-left and CS-right) had become more negative than the colour that was never followed by an aversive noise (i.e., CS-neutral). The average valence ratings for the CS-left, CS-right and CS-neutral colours are presented in Table 13 (p. 94). Although the colours that had been paired with an aversive US had become slightly more negative than the colour had never been paired with a US, this difference was not significant, $F(1, 23) = 2.35$, $p = .14$, effect size $d = 0.32$.

Contingency-awareness.

On average 2.92 of the 3 colours were correctly linked to their US. Of all participants, 96% perfectly reconstructed each of the three colour-sound pairs.

Discussion

In this experiment we tried to investigate whether the spatial contrast effect found in Experiments F1 and F2 was sensitive to extinction. Unfortunately we were unable to find the overall spatial contrast effect in the post-acquisition (baseline) spatial Simon. This notwithstanding that, up until the spatial Simon, we had used the same procedure as in the ‘spatial Simon first’ condition of Experiment F2, which had given us the clearest contrast-effect in our experiments. In our next attempt we replicated the current experiment, but tried to increase the number of mistakes made in the spatial Simon, by verbally stressing the importance of responding as quickly as possible. Increasing the number of wrong responses might make any effect in that variable stronger.

The EC-effect obtained on the valence-scale in Experiment F3 had disappeared: although colours that were paired with an aversive US tended to be more negative than the colour that was not paired with a US, this difference was not significant. Do note however, that in this experiment there again came two (spatial) Simon phases before the valence scale (as in Experiments F1 and F2), while in the Experiment F3 there only was one. As each Simon-phase can be seen as a kind of extinction phase, this again suggests that too many extinction trials abolish or prevent us from detecting EC on the valence scale.

Experiment F5

We decided to replicate Experiment F4, with one change. To increase the chance of finding the spatial contrast effect that we had observed in Experiments F1 and F2, we added an additional oral speed instruction for the Simon phase. We told participants that they really had to categorize the Simon-word as fast as possible, even if that caused them to make a few mistakes. Increasing the number of wrong responses might make any effect in that variable stronger (e.g., Draine & Greenwald, 1998).

Method

Participants.

Twenty-four first year economy students (10 men and 14 women) participated in exchange for course credits.

Stimuli and measures.

The same stimuli and measures as in Experiment F4 were used.

Procedure.

The procedure was identical to the one used in Experiment F4, except for the instruction given before the post-acquisition spatial Simon. In addition to the written instruction that asked participants to respond as fast as possible (this instruction was used in all the experiments in this Chapter), an extra oral speed-instruction was given: participants were told that it was very important to be as fast as possible, even if this resulted in a few wrong responses.

Results***Spatial Simon***

The average reaction times (over blocks and after cut-off) for the congruent, incongruent and non-congruent trials in the *first* spatial Simon task are presented in Table 12 (p. 92). To investigate whether spatial activation had occurred, we conducted a 2 x 2 ANOVA on the data from the first spatial Simon with congruency (i.e., congruent versus incongruent trials) and block (first and second) as within-subjects variables. Analysis revealed no significant reaction time difference between congruent and incongruent trials, $F(1, 23) = 2.41$, $p = .14$, and thus no evidence for spatial activation. There was also no interaction between congruency and block, $F(1, 23) = 1.11$, $p = .30$, effect size $d = -0.32$. When looking at the results of the *second* spatial Simon (after extinction), we also found no significant reaction time difference between congruent ($M = 500$) and incongruent trials ($M = 502$), $F(1, 23) < 1$, effect size $d = 0.09$.

We also analysed the percentage of errors or incorrect responses. These average percentage of errors (over blocks) for the congruent and incongruent trials in the *first* spatial Simon task, are also presented in Table 12 (p. 92). We were again unable to replicate our spatial contrast effect: participants made no more mistakes on congruent trials than on incongruent trials, $Z(24) = 0.93$, $p = .35$, effect size $d = 0.17$. There was no significant interaction between congruency and block, $Z(24) = 0.34$, $p = .74$. When looking at the results of the *second* spatial Simon (after extinction), we neither found a significant

difference between the percentage of errors made on congruent ($M = 6.4\%$) versus incongruent trials ($M = 8.0\%$), $Z(24) = 1.39$, $p = .19$, effect size $d = 0.29$.

Note that the speed instruction did seem to decrease general reaction times and increase the number of mistakes, compared to the previous experiments (see Table 12, p. 92).

Valence-scale.

We also investigated whether the expressed liking for the two colours that had been paired with an aversive noise (i.e., CS-left and CS-right) had become more negative than the colour that was never followed by an aversive noise (i.e., CS-neutral). The average valence ratings for the CS-left, CS-right and CS-neutral colours are presented in Table 13 (p. 94). A planned comparison revealed that colours that had been paired with an aversive US had become more negative than the colour that had never been paired with a US, $F(1, 23) = 7.74$, $p = .01$, effect size $d = 0.58$.

Contingency-awareness.

On average 2.92 of the 3 colours were correctly linked to their US. Of all participants, 96 % perfectly reconstructed each of the three colour-sound pairs.

Discussion

Because we failed to find the (baseline) spatial contrast effect in Experiment F4, we tried to increase the number of incorrect responses participants make in the spatial Simon. We expected that this would have increased the chance of finding a contrast-effect in this Simon-procedure. But although our extra speed instruction for this Simon did increase the number of incorrect responses made, it did not result in a significant spatial effect.

Unexpectedly, we observed EC on the valence-scale: colours that were paired with an aversive US were significantly more negative than the colour that was not paired with a US. This is the first experiment in which we observed EC on this scale after two Simon phases. This is thus the first Experiment in this chapter that suggests that EC can be observed after prolonged extinction.

Overall discussion of the fear conditioning studies

In five experiments we tried to associate both the valence (aversive loud noise present vs. no noise) and direction (loud noise from left vs. right) of a loud noise with the colour of a word. We predicted that the colour that had been paired with a loud noise from the left or right would subsequently activate 'left' or 'right' respectively (non-evaluative activation). At the same time, we predicted that the colour that had not been paired with an aversive noise would become more positive than the colours that had paired with a loud noise (evaluative transfer). To assess for this spatial and affective activation, we used associative variants of the spatial and affective Simon paradigm.

In all five experiments, we failed to find consistent evidence for spatial activation in the reaction times. Closer analyses of the percentage of incorrectly classified words in the spatial Simon showed a contrast effect in our first two experiments: when participants saw a colour that was paired with a loud noise from a certain direction (e.g., left), they made more errors when they had to press a key at that location (e.g., left) than when they had to press a key at the opposite location (e.g., right). In other words, it seemed that a colour previously associated with an aversive noise from one side, subsequently slightly facilitated (incorrectly) pressing a key in the opposite location. Although it is an associative effect, its direction is hard to explain. Moreover, we were unable to replicate the effect in three subsequent experiments, which suggests that doubt can be raised about the reliability of the phenomenon. A simple meta-analysis was conducted on the percentage of wrong responses in the spatial Simon to investigate whether, overall, there remained evidence for associative activation of spatial direction. For this meta-analysis, we used the program and manual of Schwarzer (1989) to combine the effect sizes from the fear conditioning experiments (see footnote 10 on page 65). Overall (i.e., over Experiments F1, F2, F3, F4, and F5) there was no significant spatial effect in spatial Simon, $Z = -0.95$, $p = .17$ (mean effect size $d^+ = -0.10$). There was no evidence that the spatial effect was different across the separate experiments, homogeneity $Q(4) = 5.80$, $p = .21$.

In the five experiments we reported in this chapter, we also investigated whether EC had occurred using a valence scale. Only in Experiments F3 and F5 did we find a significant transfer effect: the colours that had been paired with an aversive noise had become more negative than the colours that had not. In the first two experiments of this

chapter we also ran an affective Simon procedure. In this reaction time task we only once found evidence for EC: in the first block of the affective Simon task of Experiment F2, we observed a significant congruency-effect. This effect was, however, not found in the second block of Experiment F2. Overall, we had a difficult time finding stable EC. This seems in contrast with other fear-conditioning experiments (e.g., Hermans, Vansteenwegen, et al., 2002) in which researchers seem to have little trouble in finding it. There are however three significant differences between most of those experiments and the one from this chapter: a) in most of those experiments pictures of people are used as CSs, and not the colour in which a word is presented, b) in most of those experiments an aversive electrical stimulus is administered as US, and not a loud noise, c) when using a reaction time measure, most frequently an affective priming measure is used, and not an affective Simon procedure. It could be that one or more of these elements are responsible for the difficulty in finding EC in our experiments.

Chapter 6

Other Studies

In this chapter we will discuss other studies and phenomena that are or seem related to ATNESP. We will divide these studies in two groups. In the first part we will discuss some unpublished attempts of other researchers to obtain evidence for ATNESP (outside of the olfactory-gustatory domain). In the second part of this chapter, we will describe two phenomena, conditional after effects and hidden covariation detection studies, that at first sight resemble ATNESP. In that section, we will investigate to what extent these studies do or do not provide evidence for the existence of general ATNESP.

Other failures

We are aware of a number of unpublished studies that have also failed to produce evidence for ANTESP.

First, Vansteenwegen (1994) investigated whether neutral faces that had been paired with an angry face would subsequently be perceived as being more angry than neutral faces that had been paired with a fearful face, and vice versa. As dependent variables, she used an anger- and fearfulness rating scale and also a forced choice measure. Pairs were constructed on an a priori basis and were presented in a similar way as in the EC picture-picture studies of Baeyens, Eelen, Crombez, and Van den Bergh (1992). The three experiments gave contradictory results: once a transfer effect, once no effect, and once a contrast effect. In a final experiment, Vansteenwegen tried to replicate EC, using the procedure she had used in the other three studies, by pairing some neutral faces with a happy face and others with a sad face and asking participants to rate their liking of the faces. She found no evidence for valence transfer. She argued that these failures to consistently find associative transfer were probably due to the use of facial expressions as CSs; it might be difficult to change the perception of facial expressions as this perception is determined bottom-up, by very specific, purely perceptual characteristics. There is, however, little data to support this hypothesis directly.

Olson and Fazio (2002) paired one Pokemon cartoon with several items that were meaningfully related to the concept ‘small’ (e.g., the word ‘tiny’ or an image of a mouse) and another Pokemon with several items that referred to the concept ‘large’ (e.g., the word ‘humongous’ or an image of an ocean liner). After the acquisition phase, participants were asked to rate the size of several items, including the two Pokemon cartoons that functioned as CSs. The mean size rating for the two CS Pokemon cartoons did not differ after the learning phase. Therefore, no evidence was found for ATNESP. Note that the authors did find EC in an experiment that was identical to the size experiment but in which the Pokemon cartoons were paired with positive or negative items instead of items referring to small or large objects (also see Olson & Fazio, 2001).

De Houwer, Meersmans, Baeyens, and Eelen (2002) presented a series of studies on the ATNESP that all failed to produce reliable effects. In a first study, four pictures of human faces and four pictures of neutral objects were used as CSs. These pictures had an average brightness. During the learning phase, half of the CSs were repeatedly followed by a bright white screen whereas the other CSs were always followed by a dark screen. Participants afterwards rated how bright the CS pictures seemed to be and also judged how much they liked the pictures. Neither the brightness ratings, nor the evaluative ratings were affected by whether a CS was paired with a bright or with a dark screen.

In two further studies, De Houwer et al. (2002) tried to transfer the semantic category of words to non-words. During the learning phase, five nonwords were paired 10 times with the name of an animal (e.g., sparrow) and five other nonwords were paired 10 times with the name of an object (e.g., toaster). Nonwords were randomly assigned to a category for each participant separately. On each trial, the nonword was presented twice on the screen. The three stimuli (nonword twice and animal or object name once) appeared one above the other in the centre of the screen. The position of the animal or object name was determined randomly on each trial. Participants were instructed to name the category (animal or person) of the existing word as quickly as possible. During the subsequent test phase, participants performed a variant of the semantic Simon task (De Houwer, 1998b, 2003) that can be used to assess the acquired semantic category of the nonwords indirectly. Two consecutive experiments failed to provide clear evidence for a transfer of semantic category. De Houwer et al. (2002) also conducted additional studies to examine whether their paradigm could support the transfer of valence. These experiments were identical to

the previous experiments except that the animal and object names were replaced by positive and negative words and a variant of the affective Simon task (De Houwer & Eelen, 1998; De Houwer, 2003) was used to measure changes in the valence of the nonwords. Although a significant EC was found in one experiment, this effect could not be replicated in subsequent experiments.

De Houwer et al. (2002) also conducted an experiment in which participants were asked to drink cups containing flat or carbonated water. Flat water always had a certain flavour (e.g., apricot) whereas the carbonated water always had another flavour (e.g., kiwi). During the subsequent test phase, participants were given cups of water with various levels of carbonation and with one of the two flavours. When asked to judge the level of carbonation of each test drink, judgements did not depend on the flavour of the water.

Eva Walther (personal communication, April 23, 2003) tried to transfer the ambitiousness from one depicted person to another, using a procedure similar to the one with which she had repeatedly found EC (Walther, 2002). The ambitiousness construct was chosen because it was found to be evaluatively neutral but socially meaningful in several pilot studies. Based on baseline ratings of how ambitious depicted people appeared to be, one neutral person (i.e., which had received a low ambitiousness rating) was paired with a highly ambitious person. A control pair was also generated by combining two other pictures that were both rated as low in ambitiousness. After repeated presentations of these pairs, participants were asked to re-rate the pictures on the ambitiousness-scale. No difference was observed between the neutral person that had been paired with an ambitious one and the neutral person that had been paired with another neutral one.

Related studies

Conditional after effects

One type of phenomena that bears resemblance to ATNESP are the so called conditional after-effects. The most well known variant is the orientation-contingent colour after-effect, also called the McCollough effect, after Celeste McCollough, who first demonstrated this after-effect (McCollough, 1965). In the prototypical procedure participants receive alternate presentations of a red vertical grid and a green horizontal grid. After being

presented for a few minutes with these line patterns, participants will perceive horizontal white grids as more reddish and vertical white grids as more greenish. At the same time, white diagonal grids or a homogenous white field will be correctly perceived as white. Conditional after effects also exist for other stimuli and stimulus-properties (see Durgin, 1996, for a listing). Favreau, Emerson, and Corballis (1972), for example, alternated presentations of a red contracting spiral and a green expanding spiral²². During the test phase they observed not only conditional colour after-effects (i.e., white contracting spirals appeared green, and white expanding rotating spirals appeared red), but also conditional motion after effects (i.e., stationary red spirals appeared to be expanding and stationary green spirals appeared to be contracting). Using achromatic stimuli, Allan and Tirimacco (1987) alternated light grey horizontal grids and dark grey vertical grids. This resulted in the same grey appearing darker when presented on a horizontal than on a vertical grid. Conditional after-effects have also been observed using visual properties like orientation, brightness, spatial frequency, size and texture density (Durgin, 1996). Some conditional after-effects were observed outside the visual domain. For example, alternate presentations of a low-pitched tone of long duration and a high-pitched tone of a short duration, resulted in a high test tone of intermediate length being perceived as sounding longer than a low test tone of equal length (Walker & Irion, 1979). Non-visual contingent after-effects have also been observed with properties like duration and loudness (Durgin, 1996).

The orientation-contingent colour after-effect, described as the first example in this paragraph, is the most investigated variant on which also most theorizing has focused. Looking at the colour after-effect from a conditioning perspective, it seems that repeated presentations of a grid with a specific orientation (CS) in a certain colour (US), will result in a shift of the colour-perception of the CS in the complementary direction of the colour in which the CS was originally presented. In other words, it appears to be a reversed transfer effect: after acquisition, the CS-pattern will be perceived as less similar to the US. However, far from all researchers agree that the McCollough effect can be seen as a form of classical conditioning. McCollough herself (1965) explained these orientation specific colour after-effects as a form of adaptation of orientation-sensitive neuron(system)s: repeated presentations of a grid in a certain colour will cause an adaptation of orientation-specific edge-detectors to that colour. Subsequent stimulation of these edge-detectors will

²² Arithmetic spirals can appear to contract or expand, depending on the direction of it's rotation.

result in reduced reaction, explaining the perception of the complementary colour in the white grid.

During the last decades, several variations of classical conditioning accounts and neuronal adaptations models have been developed, but none of the models can explain all the characteristics and empirical data of the McCollough effect (for an overview, see Silverans, 1998). Investigating and describing all the characteristics of the McCollough effect, and the different conditioning and neuronal accounts developed around it, would require a dissertation on its own.

As it is hard to draw definite conclusions about whether the McCollough effect and other contingent after-effects are in fact a form of classical conditioning and not more basal neural adaptation effects, we cannot with certainty claim that the after-effects are forms of ATNESP²³. If, however, they *are* forms of associative learning, then there are some clear discrepancies with EC. First, the transfer-effect is in the opposite direction of the transfer effect that we observe in EC: in EC we observe an assimilative transfer effect (the CS-valence shifts in the direction of the US-valence), while in the McCollough effect a reverse transfer effect is observed (e.g., the CS-colour shifts in the opposite direction of the US-colour). This could suggest that the conditional after-effects are based on a different mechanism than EC. This hypothesis is supported by the observation that the McCollough effect seems, contrary to EC, sensitive to extinction (e.g., Skowbo, Gentry, Timney, & Morant, 1974, but also see Silverans, 1998).

Implicit covariation detection

Another phenomenon that resembles ATNESP, is a form of implicit learning called hidden covariation detection (HCD). In hidden covariation detection studies, stimuli are presented that contain a usually simple, but non-salient covariation between features *x* and *y*, often two visual stimuli or a visual stimulus and a verbal label (Seger, 1994). In order to

²³ This was also one of the reason why we choose not to pursue this line of research in our search for ATNESP. As long as basic stimulus properties (like for example colour, motion, orientation) are used, it is hard to exclude neuronal adaptation theories, and thus claim with certainty that it are associative conditioning effects. Moreover, the experiments of Silverans (1998) suggest that some of the observed associative characteristics, like extinction, blocking and second order conditioning, of the McCollough effect are not always easy to replicate.

suppress participants' awareness of the covariation, different methods are used such as using very small perceptual differences for the crucial correlation (e.g., Lewicki, Hill, & Sasaki, 1989), using theoretically improbable relations (e.g., Lewicki, 1986b), or using very short stimulus presentations (Hill, Lewicki, Czyzewska, & Boss, 1989) (see also Seger, 1994). Different methods are used to demonstrate that, although participants seem unaware of the presence of the covariations, these covariations do influence their behaviour. We will divide the HCD studies into two categories based on whether the response modality with which the implicit covariation learning is demonstrated (see also Seger, 1994), is related to conceptual fluency or efficiency. For each of these two types of experiments, we will also discuss the relation with ATNESP.

HCD studies with conceptual fluency measures

In the HCD studies that use conceptual fluency measures, participants are presented during acquisition with a non-salient covariation between features *x* and *y*. In a subsequent test phase, participants receive (new) stimuli that contain feature *x*, but that provide no explicit information about the presence or absence of feature *y*. Participants have to make judgements (hence the name 'conceptual fluency') regarding *y*. If participants' judgements about *y* reflect the *x*-*y* covariation, then evidence for (implicit) learning is found.

Lewicki and colleagues performed many studies that fit into this category (e.g., Hill, Lewicki, Czyzewska, & Schuller, 1990; Hill, Lewicki, & Neubauer, 1991; Hill et al., 1989, Experiment 2; Lewicki, 1986b; Lewicki et al., 1989, Experiment 1). For example, Lewicki (1986b) presented participants with pictures of women together with verbal descriptions of their personality (either capability or kindness). There was, however, a covariation between the hair length of the depicted women (*x*, short or long hair) and whether the person was described as capable or kind (*y*). During a test phase, participants received new pictures of short- and long-haired women without any information about their personality, and had to affirm or disconfirm whether they thought these women were either capable or kind. Participants assigned the capable characteristic more often to women with the hair length that had previously been paired with the capable trait than women with the hair-length that had been paired with kind. Participants were, however, unaware of the covariation between hair length and the personality trait. When looking at this implicit learning example from a conditioning perspective, it seems that repeated pairings of the hair length (CS) of women with the personality trait capable (US), resulted

in the fact that previously unseen women with that hair length were judged as being more capable. In other words, it seems like the personality trait of a women (US) was transferred via the hair-length (CS) that covaried with it²⁴. Another good example is the so called ‘brain scan’ study (e.g., Lewicki et al., 1989) in which participants were presented with ‘brain scans’ of people and a description of the intelligence of that person. During the acquisition phase a covariation was implemented between the percentwise presence (13% versus 17%) of a certain character in the brain scan (CS) and the degree of intelligence (US). After acquisition, participants were presented with new brain scans (also containing 13% or 17% of the crucial character), this time without information about intelligence. Participants were instructed to assess the intelligence of the person to which the brain scan belonged. When a new brain scan contained the CS-feature that had covaried with intelligence, it was more likely to be judged as belonging to an intelligent person.

Interestingly, these implicit learning effects seem to be insensitive to extinction. Moreover, post-acquisition CS-only presentations (i.e., in which no explicit information was given about the US) seem to *increase* the transfer effect. This so called self-perpetuating effect has been repeatedly demonstrated by Lewicki and colleagues (e.g., Hill et al., 1989; Lewicki et al., 1989). For example, in a brain scan study, participants were exposed to prolonged testing in which new brain scans without intelligence-information were presented. The transfer effect seemed to increase rather than to diminish as this test phase continued (Lewicki et al., 1989).

At first sight these experiment demonstrate ATNESP (e.g., transfer of a personality trait, intelligence,...). We have, however, three remarks. First, these implicit covariation learning effects might not be as general and robust as initially described. In a number of conceptual and exact replications, Hendrickx, De Houwer, Baeyens, Eelen, and Van Avermaet (1997) failed to find any evidence for covariation learning. This suggests that there are still unknown boundary conditions to the effect. Second, in a number of these

²⁴ Note that in this type of HCD studies, the ‘CS’ is in most cases an invariant feature of the stimulus, and not a stimulus on it’s own. Although in EC studies, the CS is most frequently a separate stimulus, there are also examples in which the CS is a stimulus feature. For example, Baeyens, Heremans, et al. (1993), implemented a correlation between the line thickness (CS) of different geometrical figures and the presentation of a liked or disliked picture. EC was observed in these experiments when new stimuli were presented in thick or thin lines.

experiments, the US seems to be a property that has a clear positive/negative connotation. In those cases, the property transfer might be mediated by an evaluative transfer. For example, as intelligence is generally seen as a positive property of a person, the property that was correlated with intelligence (13% or 17% of the crucial character), might become more positive (i.e., though EC). As a consequence, new brain scans that include the property that correlated with intelligence might be seen as more positive, and hence be rated as more intelligent. There are, however, also experiments in which HCD is very unlikely to be caused/mediated by evaluative transfer. For example, Hill et al. (1990, experiment 2), showed property transfer using two abstract personality types (i.e., ‘M6N’ and ‘L6D’). Third, and more importantly, in the majority of Lewicki’s HCD studies, the covariation between the two relevant stimulus features implies more than mere co-occurrences in space and time (Hendrickx et al., 1997). The CS or the stimulus in which the CS is embedded is usually related to the US in a proposition of the type ‘the CS is an instance of the US’ or ‘the CS is characterised by the US’. For example, in the hair-length and personality type example (Lewicki, 1986b), a woman with short hair *is* characterised as capable or kind. In the brain scan study, each brain scan *belongs* to an (un)intelligent person. In EC such propositional relations are normally not used and the simple co-occurrence (simultaneous or sequential) of two stimuli suffices to induce the valence transfer. It might be that this propositional relation between CS and US is necessary for the transfer effect to appear in these HCD studies. As we are not aware of conceptual fluency HCD studies in which a mere association is used (without implying a propositional relation between CS and US), we cannot claim that the mere pairing is enough to produce the transfer.

In short, for some studies we have evidence for ATNESP. However, we cannot exclude that more than a simple association is needed for the transfer to occur, as in these experiments a propositional relation is clearly implied between the (stimulus that contains the) CS and the US.

HCD studies with efficiency measures

In the second group of HCD studies, efficiency measures are used to investigate covariation learning. Participants are presented with a non-salient covariation between features x and y and participants are instructed to respond to the presence of y. Note that because each (type of) stimulus is assigned to a particular response, the x-y covariation is

confounded with a x-response covariation. The feature x thus predicts the presence of y (and the nature of the required response). Covariation learning is inferred by measuring the speed and accuracy of the response to stimuli in which x is present. In most cases a comparison is made between a condition/trials/phase in which x predicts y, and a condition/trials/phase in which x does not covary with y.

For example, Miller (1987) presented participants with a row of letters. Participants were instructed to ignore the two flanking letters and to respond to the central letter by pressing one of two response keys, depending on which predetermined set the central letter belonged to. However, the flanking letters (which belonged to a different set than the target central letters) covaried with the correct response. This covariation between the flanking letters and the correct response, was however not perfect. On a small percentage of trials, called invalid trials, the covariation was reversed, and the flanking letters predicted the wrong response. Miller observed that participants responded slower on those invalid trials than on the valid trials, indicating that they had learned the covariation. However, participants seemed to be unaware of the covariation between flankers and correct response. Some studies of Lewicki and colleagues also belong to this group of experiments. For example, Lewicki (1986a, Experiment 3.1) presented participants with matrixes of letters and asked them to indicate in which quadrant of a matrix a predetermined target letter was present. The presentation of each matrix was preceded by a subliminal exposure of a string of characters (e.g., 'A235A'). The identity of the string covaried with the quadrant in which the target letter was to be found. Participants responded slower in the condition in which after an acquisition phase the covariation was reversed, compared to a condition in which the covariation stayed the same.

These studies resemble ATNESP less than the studies from the previous category of HCD experiments. For one, no property transfer is measured, so associative *transfer* of stimulus properties cannot be demonstrated by these experiments. The maximum that could be inferred from this kind of experiments is that x activates y, or that x activates the response to y. But precisely the fact that x was also associated with the response to y, forms a crucial difference with our fear-conditioning experiments. In our fear conditioning experiments, the colour of a word (CS) was associated with the direction from which a loud noise came, not with a particular response. In the HCD studies that were discussed in this section, feature x (CS) was associated not only with the stimulus(-feature) y (US), but

also with the response that participants had to emit to y . So if evidence is found for covariation learning in these studies, it is not clear whether it is based on the activation of the US by the CS (S-S learning), or on the activation of the behaviour that participants learned to emit to the US (S-R learning). As the effect in these studies could also be explained by S-R learning, it is hard to see them as evidence for associative *activation* of non-evaluative stimulus properties.

Chapter 7

General Discussion and Conclusion

General discussion

The main aim of this dissertation was to find evidence for Associative Transfer of Non-Evaluative Stimulus Properties (ATNESP) outside of the olfactory-gustatory domain. In a first series of experiments (Chapter 3), we tried to transfer the age or gender of a depicted person onto the picture of another person with which it was paired. We found evidence for gender-transfer once, but were unable to replicate it in four subsequent experiments. In Experiment P8, a transfer effect was found, but only in those participants who had good knowledge of the type of US that was paired with each CS. In a second line of experiments (Chapter 4), we twice tried to transfer the gender of clearly male/female first names onto Japanese characters. One experiment revealed evidence for a transfer effect under conditions where demand effects were unlikely to have an impact, but this effect could not be replicated in a subsequent experiment. In the third and final series of experiments (Chapter 5), we did not investigate non-evaluative *transfer* of a property from the US to a CS, but focused on a sub-process that is probably a prerequisite for ATNESP: *activation* of a non-evaluative US-property by a CS. We tried to associate the direction and the valence of aversive USs with the colour of a word. In the first two experiments, we found an unexpected contrast effect: the colour that had been paired with a US from the right (left) subsequently facilitated a left (right) response. Unfortunately, we were unable to replicate this effect in three subsequent experiments. Finally, we briefly described several other studies from various labs that failed to produce evidence for ATNESP, and discussed some phenomena that resemble ATNESP (Chapter 6).

Is there evidence for general associative transfer of non-evaluative stimulus properties?

Overall, the data on ATNESP that has been collected so far are disappointing. On the positive side, one could argue that we did observe significant transfer effects in a number of experiments (i.e., Experiments P3, P8, S1, and S2). Unfortunately, these effects could

either not be replicated or might have been due to demand effects. We thus found little evidence for genuine transfer effects. That is, the observed transfer effects might not have reflected an actual change in the perception of the CSs but could have been due to the fact that participants consciously and intentionally used their knowledge of the CS-US pairs as a basis for their post-acquisition ratings of the CSs. Participants might have followed this strategy simply because they believed that this was what the experimenter wanted them to do. It is interesting to note that in Experiments P8, S1, and S2, where transfer effects were found, all of the presented CS-US pairs involved USs that were clearly male or female whereas in the other experiments, some USs were neutral with regard to the crucial feature. This might have encouraged participants to believe that knowledge about the US property should be used as a basis for their ratings of the CSs.

In several experiments, we tried to minimize the impact of demand effects by using indirect measures. However, most often we did not find a transfer effect on these measures. One exception were the priming data of Experiment S1. Unfortunately, this effect could also not be replicated. Therefore, the data of the indirect measures also give little reason to believe that our procedures resulted in genuine transfer effects. Some might argue that we lend too much weight to the possibility that the observed transfer effects could have been due to demand characteristics. However, previous studies on EC have shown that demand effects are likely to occur in tasks similar to ours (see Field, in press, and Page, 1969, for a discussion). Some caution is thus warranted.

Is there evidence for associative *activation* of non-evaluative stimulus properties?

Our prediction that non-evaluative stimulus properties can transfer as the result of CS-US pairings was partially based on the model of Baeyens and colleagues (e.g., Baeyens, Crombez, et al., 1995). Baeyens and colleagues (e.g., Baeyens, Crombez, et al., 1995) argued that, due to referential learning, a CS can activate the associated US representation without leading to the expectation that the US will actually occur. They hypothesised that this would influence the evaluation of the CS if the activated US-valence is misattributed to the CS. One possible explanation of our failure to find clear evidence for transfer effects is that misattribution of non-evaluative stimulus properties is less likely to occur. It could indeed be argued that judgments about the valence of a stimulus are more subjective and

thus less determined by the objective properties of the stimulus than judgments about other, non-evaluative stimulus features. This line of reasoning suggests that it should be easier to transfer non-evaluative stimulus properties that are perceived in a more subjective manner. It should be noted that we selected features such as the gender of babies precisely because we thought that there were few objective cues on which judgments about these features could be based. Nonetheless, it seems reasonable to argue that, in comparison to valence, all non-evaluative cues that we used are more closely linked to objective features of the stimuli.

However, even if misattribution of non-evaluative stimulus properties is less likely than misattribution of evaluative properties, this does not necessarily mean that no association is formed between the CS and the non-evaluative US-properties. When looking closely at the referential learning account of EC, one could discern two sub-processes: for transfer to occur, a) the CS must activate the US-valence, and b) the activated US-valence must be misattributed to the CS. It is possible that, while activated US-valence is easily misattributed to the CS, activated non-evaluative properties of a US are not that easily misattributed. In that case, however, we could still expect the CS to activate non-evaluative US-properties. To investigate this associative *activation* of non-evaluative stimulus-properties, we have conducted the fear-conditioning experiments. We were, however, unable to find a stable associative effect. Moreover, the effect we did find twice (but were unable to replicate in three further experiments) was a contrast-effect, that could be caused by regular expectancy learning.

Other forms of associative-*like* transfer of non-evaluative properties?

There are indications that propositional relations of the type ‘the CS is a US’ or ‘the CS is characterised by US’ can cause property transfer. Lewicki and colleagues demonstrated that a US-property seemed to transfer to a CS-feature of a stimulus that was said to contain the US-property. For example, if women with long hair (CS) were described as being capable (US), then participants believed other women with long hair to be capable too (Lewicki, 1986b). In most of these experiments however, it was not investigated whether some kind of evaluative transfer mediates the property transfer. More importantly, as we are unaware of hidden covariation detection experiments that do not use the ‘is a’ or ‘is characterised by’ relation between CS and US, we cannot claim that the transfer was

caused by mere association. It is possible that these propositional relations are a prerequisite for the transfer to occur. In contrast, the mere concurrence of a neutral CS and a valenced US seems enough to obtain valence transfer.

It is also important to mention conditional after-effects in this context. McCollough and colleagues, for example, demonstrated that repeated presentations of a grid with a specific orientation (CS) in a certain colour (US), will result in a white CS-grid being perceived as containing less of the US-colour (e.g., McCollough, 1965). Similar contingent after-effects have also been observed using other basic/simple stimulus properties like colour, motion, duration, orientation, spatial frequency, brightness, texture density, size and loudness (Durgin, 1996). Although these are all contrast effects (the CS acquired the US-property less), it still seem associative ‘transfer’ of non-evaluative stimulus properties. However, alternative more low-level neuronal explanations exist for these phenomena (e.g., adaptation of orientation-sensitive neuron systems), and it is not clear whether the conditional after-effects are a form of classical conditioning or are caused by more basal neural adaptation effects. It is thus hard to treat these after-effects as clear evidence for ATNESP. If we accept these phenomena to be caused by associative learning, then there is evidence for ATNESP, but only with basic stimulus-properties (like colour, motion, duration, ...) and the shift is in the opposite direction then the shift in EC.

Why did we not find clear evidence for ATNESP?

The question remains why we did not find clear evidence for genuine associative transfer effects. Although one might be inclined to conclude that ATNESP does not exist when visual stimuli are used, it is also possible that this form of learning does exist but can be found only when strict boundary conditions are met. One possible boundary condition might be the nature of the stimulus features. For instance, it is possible that we selected stimulus-features that are difficult to transfer in an associative manner. Maybe other features are more easy to transfer in such a way. One should note, however, that we and others have conducted many studies in which a variety of features was used: age and gender of photographed persons, gender of first names, emotional expressions, size, brightness of pictures, semantic category of words, and ambitiousness of persons. It is not clear to us what other features might be more suitable to transfer in an associative manner.

On the other hand, the results of the studies by Stevenson and colleagues (e.g., Stevenson et al., 1998, 2000a, 2000b) clearly demonstrate that non-evaluative features of tastes and odours can be transferred in an associative manner. The nature of the stimuli and stimulus features might thus be an important moderating factor. However, it is also possible that the transfer effects that were observed by Stevenson and colleagues were based on processes that operate only during the processing of tastes and odours. For instance, Stevenson and colleagues (Stevenson et al., 2000a, 2000b) argued that such stimuli are processed in a more holistic manner than visual stimuli. That is, the CS and US are experienced as one event and stored within a single memory representation. When the CS is presented during the test phase, this activates the memory representation of those prior experiences and thus biases the perception of the CS in the direction of the US. This hypothesis implies that it should be more difficult to find ATNESP when visual stimuli are used. The data that we presented in this paper seem to support this hypothesis.

Apart from the nature of the stimuli, it could also be that the number of CS-US pairings is an important determinant of associative transfer of (non-evaluative) stimulus properties. In our studies, each CS-US pair was presented nine or ten times. Although such a number of pairings suffices to induce EC (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Diaz et al., 2002), it is possible that more pairings are needed in order to observe ATNESP. Another element that might have prevented any kind of transfer (evaluative or non-evaluative) in our picture-picture studies, was the differentiability of our stimuli. It could for instance be that the infant-pictures that were used in the picture-picture experiments were too similar to one another for participants to be able to differentiate them. However, in the symbol-word experiments, the CSs were four Japanese characters that were, in our eyes, clearly differentiable. In these symbol-word experiments also no clear, demand-free, evidence for ATNESP was found.

It is possible that there are still other boundary conditions that we are not yet aware of but that did prevent us from finding clear evidence for ATNESP. From this perspective, it is interesting to note that EC also seems to be restricted by certain, as yet unidentified boundary conditions. In their reviews of the literature on EC, both De Houwer and colleagues (De Houwer, Thomas, & Baeyens, 2001) and Field (in press) noted that, although EC effects have been found in many studies, there have also been several genuine failures to find EC (e.g., Field, 2002; Field & Davey, 1999; Rozin et al., 1998). Whether

the boundary conditions that apply to EC are the same as those that apply to ATNESP is unclear. Some data suggest that this might not be the case. First, Stevenson and colleagues (Stevenson et al., 1995, 1998) did find transfer of the sweetness/sourness of a taste onto a flavour without finding transfer of the valence of the taste to the flavour. Because a sour taste is normally more negative than a sweet taste, one would expect EC to occur. Second, Olson and Fazio (2002) and Walther (personal communication, April 23, 2003) failed to find ATNESP despite using a procedure that does produce reliable EC effects.

In order to identify the actual boundary conditions, it would of course be helpful to have a procedure that does result in genuine and reliable transfer of non-evaluative properties of visual stimuli. It is, however, possible that such a result will never be obtained simply because the phenomenon does not exist. The available data do not allow for such a strong conclusion because certain boundary conditions might not have been met in the existing studies.

Conclusion

We did not find clear evidence that proves the existence of ATNESP outside of the olfactory-gustatory domain. The few effects we did find either disappeared in replication studies or seemed to be caused by demand effects. Also a number of unpublished studies of other researchers failed to provide evidence for ATNESP. The best evidence for ATNESP using visual stimuli comes from studies investigating colour-contingent after-effects, but numerous low-level neuronal explanations exist for this phenomena (e.g., adaptation of orientation-sensitive neuron systems) and it is not clear whether the conditional after-effects are a form of classical conditioning or are caused by more basal neural adaptation effects. Hidden covariation detection studies by Lewicki and colleagues suggest that property transfer can occur when a propositional relation of the type ‘the CS is a US-property’ or ‘the CS is characterised by the US-property’ exist between CS and US. But apart from the fact that in many of these studies the non-evaluative transfer effect might be caused/mediated by evaluative transfer, it seems to be an effect that needs more than the simple co-occurrence of two stimuli. Moreover these hidden covariation detection effects seem hard to replicate (Hendrickx et al., 1997). Finally, we also failed to find clear

evidence for associative *activation* of non-evaluative stimulus properties, which is likely to be a sub-process of ATNESP.

However, not finding ATNESP does not mean that it does not exist. It is possible that the phenomenon does not exist, but it is also possible that there are boundary conditions to ATNESP that were not fulfilled in our studies. At the very least, the results that we reported as well as the results of several unpublished studies on ATNESP do suggest that the boundary conditions that apply to this type of learning are manifold.

Samenvatting [Dutch Summary]

Inleiding

Wat mensen aangenaam of onaangenaam vinden, is in veel gevallen niet aangeboren, maar wordt geleerd uit ervaring (Rozin & Millman, 1987). Evaluatieve Conditionering (EC) is één van de mechanismen waarmee voorkeuren en afkeren worden verworven. EC verwijst naar het fenomeen dat het gepaard aanbieden van een affectief neutrale stimulus met een stimulus die men reeds aangenaam of onaangenaam vindt, de valentie van de oorspronkelijk neutrale prikkel kan veranderen. De laatste 25 jaar heeft deze vorm van conditionering veel aandacht gekregen (voor een overzicht, zie De Houwer, Thomas, and Baeyens, 2001), aangezien ze zich anders lijkt te gedragen dan meer traditionele vormen van Pavlovianse Conditionering (PC). Zo lijkt EC, in tegenstelling tot PC, ongevoelig aan extinctie en onafhankelijk van het zich al dan niet bewustzijn van de VP-OP contingenties.

Verskillende theorieën werden voorgesteld om de functionele verschillen tussen EC en PC te verklaren. Eén van de meest veelbelovende theorieën postuleert dat EC een vorm is van louter referentieel leren (Baeyens & De Houwer, 1995; Baeyens, Eelen & Crombez, 1995; Baeyens, Eelen, Crombez, & Van den Bergh, 1992): door het herhaaldelijk paren van een VP met een OP, zal het aanbieden van de VP op zichzelf genoeg zijn om de representatie van de (valentie van de) OP te activeren, zonder dat daardoor noodzakelijk de OP verwacht wordt. Deze geactiveerde OP-valentie wordt vervolgens (ten onrechte) toegewezen aan de VP waardoor de gepercipieerde valentie van de VP verandert.

Tot hiertoe is het idee van referentieel leren alleen nog maar toegepast in de context van overdracht van valentie. Echter, als de beoordeling van de VP-valentie beïnvloed kan worden door gepaarde aanbiedingen met een OP, dan kunnen misschien ook andere eigenschappen van de VP beïnvloed worden door deze associatie. Met andere woorden, misschien kan niet alleen de valentie van prikkels overgedragen worden, maar ook andere prikkeleigenschappen. Een aantal studies, uitgevoerd door Stevenson en collega's (o.a., Stevenson, Boakes, & Prescott, 1998; Stevenson, Boakes, & Wilson, 2000a, 2000b; Stevenson & Case, 2003; Stevenson, Prescott, & Boakes, 1995), suggereren dat dit

inderdaad mogelijk is. Deze auteurs slaagden erin om de zoetheid of zuurheid van een drankje over te dragen op de geur van het drankje. De auteurs verklaarden dit fenomeen echter niet als referentieel leren. Stevenson, Boakes and Wilson (2000a, 2000b) suggereerden dat geur-smaak leren iets unieks is, aangezien geur en smaak sensorieel heel dicht bij elkaar liggen en misschien holistische opgeslagen worden. Niet-evaluatieve transfer zou dus iets uniek kunnen zijn voor gustatorische-olfactorische prikkels.

Het hoofddoel van deze verhandeling was het zoeken naar evidentie voor Associatieve Overdracht van Niet-Evaluatieve PrikkelEigenschappen (AONEPE), vooral met andere prikkelmodaliteiten dan geuren en smaken. Als we deze evidentie gevonden hadden, wilden we nagaan of deze overdracht zich hetzelfde gedraagt als EC. In Hoofdstuk 1 geven we een overzicht van onderzoek naar EC. In Hoofdstuk 2 gaan we een stap verder en kijken we naar bestaande evidentie voor AONEPE en eindigen we met het definiëren van onze onderzoeksdoelen. In de volgende drie hoofdstukken bespreken we de experimenten die we uitvoerden in onze zoektocht naar AONEPE: In Hoofdstuk 3 bekijken we de beeld-beeld experimenten, in Hoofdstuk 4 de symbool-woord experimenten en in Hoofdstuk 5 de vreesconditioneringsexperimenten. In Hoofdstuk 6 bespreken we experimenten over AONEPE van andere onderzoekers. We eindigen met een samenvatting van onze bevindingen en de bevindingen van anderen in Hoofdstuk 7.

Hoofdstuk 1: Evaluatieve conditionering

In dit hoofdstuk wordt een beknopt overzicht gegeven van EC. In een eerste paragraaf wordt het universeel karakter van EC besproken. In de tweede paragraaf gaan we in op de functionele eigenschappen van EC. In de laatste paragraaf bekijken we de verschillende modellen die geformuleerd werden om de functionele verschillen tussen EC en traditionele PC te verklaren. Dit hoofdstuk is gestructureerd volgens en grotendeels gebaseerd op het recente overzicht van De Houwer, Thomas, en Baeyens (2001).

EC is aangetoond binnen een variëteit van procedures en paradigma's. De meest bekende variant met visuele stimuli is het beeld-beeld paradigma met menselijke gezichten van Baeyens en collega's (o.a., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Baeyens, Eelen, Van den Bergh, & Crombez, 1989a, 1989b, 1992; Baeyens, Hermans, & Eelen, 1993) waarin foto's van

gezichten die een deelnemer oorspronkelijk neutraal vond (VP) herhaaldelijk gepaard werden met gezichtsfoto's die de deelnemer zeer aangenaam/onaangenaam vond (OP). De typische bevinding is dat de oorspronkelijke neutrale VP na deze acquisitiefase als aangename/onaangename wordt ervaren. Gelijkaardige effecten werden gevonden met andere visuele stimuli. Ook in andere modaliteiten werd EC gevonden. In het smaak-smaak paradigma kon men door het paren van een bepaalde smaak (vb, abrikoos) met een zeer onaangename smaak (Tween20; een zeepachtige smaak), de oorspronkelijk neutrale smaak onaangename maken (voor een overzicht van evaluatieve smaak-smaak conditionering, zie Baeyens, Vansteenwegen, Hermans, & Eelen, 2001b). EC werd ook aangetoond met tactiele prikkels (Hammerl & Grabitz, 2000) en biologisch belangrijke prikkels zoals bijvoorbeeld een elektrische schok (o.a., Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002; Francken, 2003). Het is ook mogelijk om EC over verschillende modaliteiten heen te observeren (o.a., Todrank, Byrnes, Wrzesniewski, & Rozin, 1995) of in een observationeel paradigma EC aan te tonen (o.a., Baeyens, Eelen, Crombez, & De Houwer, 2001). Alhoewel EC in verschillende paradigma's al herhaaldelijk werd gedemonstreerd, wil dit niet zeggen dat elk experiment EC oplevert. Soms faalden conceptuele en zelfs exacte replicaties van EC experimenten.

Alhoewel EC qua procedure sterk lijkt op PC, zijn er op functioneel vlak toch verschillen met traditionele PC. Zo lijkt in EC de (evaluatieve) respons niet gevoelig voor een uitdovingprocedure. Het werd immers meermaals aangetoond dat het herhaaldelijk aanbieden van de VP op zichzelf (dus zonder de affectief geladen OP), de VP niet terug neutraal maakt. EC lijkt ook voor te kunnen komen wanneer deelnemers zich niet bewust zijn van welke VPs met welke OPs gepaard werden. Verder lijkt EC niet afhankelijk van de statistische contingentie tussen de VP en de OP en is EC ook niet onderhevig aan modulatie. Merk wel dat niet iedereen deze verschillen onderkent (zie bijvoorbeeld Lipp & Purkis, in druk).

Om deze verschillen tussen EC en traditionele PC te verklaren, werden door de jaren heen verschillende modellen voorgesteld. Een eerste verklaring was dat EC een intrinsieke verandering in de VP veroorzaakt. Als gevolg van de VP-OP aanbiedingen zou de VP zelf positiever/negatiever worden. Een tweede verklaring werd gegeven door Martin en Levey (1978, 1987, 1994). Volgens hun holistische representatietheorie, zal het samen aanbieden van de VP met de OP leiden tot een holistische VP/OP representatie. Wanneer

vervolgens één van de twee prikkels wordt aangeboden, zal die automatisch het VP/OP complex (en dus ook de geïntegreerde OP-valentie) activeren. Een derde verklaring werd gegeven door Davey (1994). Davey suggereerde dat EC niet gebaseerd is op een associatieve relatie tussen de prikkels, maar op een soort conceptuele categorisatie: Door het gepaard aanbieden van een VP en een OP zullen de prikkelementen die de VP en OP gemeenschappelijk hebben, meer saillant worden. Aangezien de OP vooral aangename/onaangename prikkelementen bevat, zullen de gemeenschappelijke elementen tussen de VP en de OP meestal ook aangename/onaangename karakteristieken zijn. Bijgevolg zal de appreciatie voor de VP verschuiven in de richting van de valentie van de OP waarmee hij gepaard werd. Een vierde verklaring werd uitgewerkt door Baeyens, Eelen, en Crombez (1995). Er werd gesuggereerd dat een OP bestaat uit verschillende componenten die verschillen in specificiteit en de soort van reactie die ze uitlokken. Zo werd bijvoorbeeld een onderscheid gemaakt tussen de emotioneel-affectieve component en de sensorieel-descriptieve component van een OP. De emotioneel-affectieve component genereert meestal algemene en diffuse emotionele toestanden, terwijl de sensorieel-descriptieve component meer specifieke reacties uitlokt. Baeyens, Eelen, en Crombez suggereerden dat EC vooral gebaseerd is op de associatie tussen de VP en de emotioneel-affectieve component van de OP. Een laatste verklaring voor EC werd eveneens geformuleerd door Baeyens en collega's (Baeyens & De Houwer, 1995; Baeyens, Eelen & Crombez, 1995; Baeyens, Eelen, Crombez, & Van den Bergh, 1992). Deze auteurs suggereerden dat er twee leersystemen zijn: het verwachtingssysteem en het referentieel systeem. Het verwachtingssysteem is verantwoordelijk voor het detecteren van betrouwbare voorspellers van biologisch significante gebeurtenissen, met als doel het organisme zo goed mogelijk voor te bereiden op de komende OP. Het referentieel systeem daarentegen is een minder complex systeem dat onze voorkeuren/afkeren voor VPs vorm geeft. In dit referentieel systeem geeft een VP geen aanleiding tot het hier-en-nu verwachten van de OP, maar zal de VP wel de OP-representatie activeren. Het activeren van de OP-representatie leidt alleen tot het anders percipiëren van de VP: de geactiveerde OP-valentie wordt (foutief) toegeschreven aan de VP.

Hoofdstuk 2: Meer dan evaluatieve conditionering?

Als EC veroorzaakt wordt door het foutief toeschrijven van de geactiveerde OP-valentie aan de VP dan kunnen andere eigenschappen van de OP misschien ook toegeschreven worden aan de VP. Met andere woorden, misschien kan niet alleen de valentie van een OP overgedragen worden naar een VP maar ook andere OP-eigenschappen. De zoektocht naar algemene Associatieve Overdracht van Niet-Evaluatieve PrikkelEigenschappen (AONEPE) vorm het centrale thema van dit doctoraat.

De bestaande evidentie voor algemene AONEPE is schaars. Stevenson en collega's (Stevenson, Boakes, & Prescott, 1998; Stevenson, Boakes, & Wilson, 2000a, 2000b; Stevenson & Case, 2003; Stevenson, Prescott, & Boakes, 1995) demonstreerden dat het mogelijk was om de zoetheid of zuurheid van een drankje over te dragen op de geur van het drankje. In een prototypisch experiment kregen deelnemers een zoete sucrose oplossing (OP) die een VP-geur bevatte (vb, lychee) en een citroenzuur oplossing (OP) die een andere VP-geur bevatte (vb, kastanje). Na deze acquisitiefase werden aan de deelnemers geurflacons aangeboden die de VP-geur bevatten. Participanten beoordeelden de VP geuren die voordien gepaard waren met een zoete of zure smaak als respectievelijk zoeter of zuurder. De auteurs verklaarden dit fenomeen echter niet in termen van referentieel leren. Stevenson, Boakes and Wilson (2000a, 2000b) suggereerden dat het geur-smaak leren misschien iets unieks is, aangezien geur en smaak sensorieel heel dicht bij elkaar liggen en misschien holistisch opgeslagen worden. Niet-evaluatieve transfer zou dus iets uniek kunnen zijn voor geur/smaak-prikkels. Er is inderdaad weinig evidentie voor AONEPE binnen andere modaliteiten. Kim, Allen, en Kardes (1996) droegen de 'snelheid' en 'zachtheid' van OP-foto's over op VP-foto's. Hun resultaten zouden echter te wijten kunnen zijn aan vraageffecten. Hoewel er duidelijke evidentie bestaat voor AONEPE met geur- en smaak-prikkels, is het dus onduidelijk of AONEPE een algemeen fenomeen is dat ook bestaat binnen andere modaliteiten.

De zoektocht naar AONEPE met niet-gustatorische en niet-olfactorische prikkels vormt de kern van dit doctoraat. Bij het vinden van niet-evaluatieve transfer, dienen vraageffecten uitgesloten te worden. Daarenboven moeten we nagaan of de effecten niet gemedieerd/veroorzaakt worden door evaluatieve transfer. Als we duidelijke evidentie

voor AONEPE vinden, willen we vervolgens nagaan of het dezelfde functionele karakteristieken heeft dan EC.

Hoofdstuk 3: Beeld-beeld studies

Een eerste reeks van experimenten was gebaseerd op het beeld-beeld EC paradigma dat Baeyens en collega's gebruikten (vb, Baeyens, Hermans, & Eelen, 1993). De eigenschappen die we probeerden over te dragen moesten voldoen aan drie criteria. Ten eerste moesten er prikkels bestaan die relatief ambigue waren met betrekking tot de eigenschap. Ten tweede moest de eigenschap (net als valentie) min of meer biologisch significant zijn. Ten derde moest de eigenschap op een dimensie liggen die niet sterk gecorreleerd is met valentie. We selecteerden uiteindelijk 'leeftijd' en 'geslacht' als eigenschappen. Voor beide bestaan relatief ambigue stimuli: Voor geslacht gebruikten we foto's van baby's bij wie het geslacht moeilijk te onderscheiden was, en voor leeftijd gebruikten we foto's van volwassenen, bij wie de exacte leeftijd moeilijk in te schatten was. Beide eigenschappen zijn vanuit biologische standpunt belangrijk, aangezien ze voor een deel onze sociale interacties bepalen. Bovendien zijn deze kenmerken niet sterk gecorreleerd met valentie.

In Experiment P1 probeerden we, tevergeefs, de geschatte leeftijd van mensen van middelbare leeftijd (VP) te beïnvloeden door foto's van die mensen te paren met foto's van oude of jonge mensen (OP). In de volgende vijf experimenten (P2 tot en met P6) probeerden we het geslacht van geslachtambigue babyfoto's te beïnvloeden door ze te paren met foto's van duidelijk mannelijke of vrouwelijke peuters. In de Experimenten P3 en P4 voegden we, naast de gebruikelijke verbale geslachtsschaalmeting, ook een geslachtsprimingmaat toe. Deze reactietijdmaat is minder gevoelig voor mogelijke vraageffecten. In Experimenten P5 en P6 werd de primingmaat vervangen door een andere indirecte geslachtsmaat, waarin deelnemers probeerden aan te geven wat de voornaam van de VP-baby was. De resultaten van deze geslachtsoverdrachtexperimenten waren niet bemoedigend. Alleen in Experiment P3 vonden we een significante overdracht van geslacht in de verbale geslachtsschaal. We konden dit effect echter niet repliceren in de volgende experimenten en hebben het ook nooit gevonden in één van de meer indirecte maten (priming of benoeming). In een volgende experiment, Experiment P7, probeerden

we met onze parameters en prikkels EC te repliceren. Aangezien dit niet lukte, dachten we dat onze prikkels misschien niet genoeg differentieerbaar waren. Het was immers mogelijk dat de babyfoto's die als VP's gebruikt werden nogal moeilijk van elkaar te onderscheiden waren en we daarom geen enkele vorm van overdracht konden vinden. In Experiment P8 probeerden we wederom geslacht over te dragen, maar dit keer gebruikten we een selectie van meer differentieerbare baby-foto's. Ook deze keer slaagden we er echter niet in om de geslachtstransfer van Experiment P3 te repliceren.

Een eenvoudige meta-analyse wees uit dat, over de geslachtsoverdrachtexperimenten heen, er globaal genomen nog een klein, maar significant, geslachtsoverdrachteffect was. Dit werd echter veroorzaakt door het deel van de deelnemers die zich grotendeels bewust waren van de contingenties tussen de VPs en OPs. Mogelijk is dit overdrachteffect dus gevolg van vraageffecten. We hebben dus geen sluitende evidentie voor AONEPE.

Hoofdstuk 4: Symbool-woord studies

In een tweede reeks experimenten gebruikten we een symbool-woord paradigma om het geslacht van duidelijk mannelijke/vrouwelijke voornamen (OP) over te dragen op Japanse karakters (VP). Na acquisitie werd aan de deelnemers verteld dat de Japanse karakters in feite Japanse voornamen waren waarvan zij het geslacht moesten aangeven. We kozen voor dit symbool-woord paradigma omdat Diaz, Baeyens en Ruiz (2002) hiermee EC al hadden aangetoond. Aangezien de schaarse overdrachteffecten die we in de beeld-beeld experimenten vonden misschien te wijten waren aan vraageffecten, probeerden we in de huidige reeks van experimenten na te gaan in welke mate deelnemers bewust hun kennis van de VP-OP contingenties gebruikten in hun beoordeling van het geslacht van het Japanse karakter. Daarnaast werd ook een geslachtspriming gebruikt als indirecte maat voor het gepercipieerde geslacht van de Japanse karakters.

In Experiment S1 vonden we evidentie voor geslachtsoverdracht, zowel in de verbale schaalmeting als in de meer indirecte geslachtspriming maat. Ongeveer twee derde van de deelnemers gaf echter toe bij hun beoordeling van het geslacht van de VP rekening te hebben gehouden met het geslacht van de OP-naam met dewelke de VP werd gepaard tijdens de acquisitiefase. De transfereffecten waren ook het sterkst voor deze groep van

‘demand aware’ deelnemers. Langs de andere kant vonden we ook een effect in de primingmaat, die normaal ongevoelig zou moeten zijn voor vraageffecten. Vooraleer conclusies te trekken probeerden we dit resultaat te repliceren in Experiment S2. In Experiment S2 vonden we echter alleen nog een effect in de schaalmeting, en dan nog alleen bij de ‘demand aware’ deelnemers. Verdere analyse van de resultaten van Experiment S1 en S2 suggereren dat misschien meer nodig is dan het simpelweg paren van twee prikkels om AONEPE te bekomen. De data suggereren dat het vormen van bewuste inferenties van de vorm ‘het Japanse karakter *is* een vertaling van de mannelijke/vrouwelijke OP-voornaam’ noodzakelijk is om de geslachtstransfer te bekomen. Globaal genomen leken we alleen evidentie te vinden voor geslachtsoverdracht bij deelnemers die bewust rekening hielden met het geslacht van de OP met dewelke de VP gepaard werd tijdens de acquisitiefase. We vonden weinig evidentie voor geslachtsoverdracht door zuiver associatief leren.

Hoofdstuk 5: Vreesconditionering studies

Wegens de problemen die we hadden om consistent AONEPE te vinden in onze vorige paradigma's, besloten we in een laatste reeks studies een subproces van AONEPE te onderzoeken. Volgens het model van referentieel leren (o.a., Baeyens, Eelen, & Crombez, 1995) wordt de valentieoverdracht veroorzaakt door het verkeerd toewijzen van de geactiveerde OP-valentie aan de VP. Hierin kan men twee subprocessen onderscheiden: om EC te bekomen moet a) de VP de OP-valentie activeren en b) deze geactiveerde OP-valentie toegeschreven worden aan de VP. Het is mogelijk dat de geactiveerde OP-valentie makkelijker toegeschreven wordt aan de VP dan andere geactiveerde OP-eigenschappen. In de experimenten in dit hoofdstuk probeerden we na te gaan of het paren van een VP met een OP ervoor kan zorgen dat niet-evaluatieve OP-eigenschappen *geactiveerd* worden door de VP. We zochten nu dus niet echt naar associatieve *overdracht* van niet-evaluatieve prikkeleigenschappen, maar eerder naar associatieve *activatie* van niet-evaluatieve prikkeleigenschappen. Hiervoor gebruikten we een vreesconditioneringsparadigma dat ons toeliet om tegelijkertijd evaluatieve overdracht/activatie en niet-evaluatieve activatie na te gaan.

In vijf experimenten (F1 tot en met F5) werden deelnemers blootgesteld aan een contingentie tussen de kleur (VP) van een woord en de valentie (aversief geluid of geen aversief geluid) en richting (aversief geluid van links of van rechts) van een OP. In één van de condities, bijvoorbeeld, werden woorden in het blauw altijd gepaard met een aversieve toon die via een hoofdtelefoon in het linkeroor werd aangeboden, woorden in het rood werden altijd gepaard met een luide toon in het rechteroor, en gele woorden werden nooit gepaard met een luide toon. We verwachtten dat na acquisitie de kleur die gepaard werd met het luid geluid van rechts of links nadien respectievelijk ‘rechts’ of ‘links’ zal activeren. Tezelfdertijd verwachtten we dat de kleuren die gepaard werden met een aversief geluid negatiever zouden worden dan de kleur die nooit gepaard werd met een luid geluid. Om deze affectieve en spatiale activatie na te gaan, gebruikten we varianten van het Simon paradigma.

In Experiment F1 en F2 vonden we evidentie voor spatiale activatie. De richting van dit effect was echter het tegenovergesteld van onze verwachtingen: de kleur die voordien geassocieerd was met een aversief geluid van één kant, faciliteerde vervolgens het (foutief) drukken op een toets op de tegenovergestelde locatie. We observeerden dus een contrasteffect. Dit effect zou te wijten kunnen zijn aan verwachtingsleren. We konden dit spaciaal activatie-effect echter niet repliceren in de drie volgende experimenten (F3, F4 & F5). Een eenvoudige meta-analyse wees uit dat, over de vreesconditioneringsexperimenten heen, er globaal genomen geen evidentie voor spatiale activatie was.

Hoofdstuk 6: Andere studies

In dit hoofdstuk worden andere studies besproken die ook AONEPE, of op AONEPE lijkende fenomenen, onderzoeken. We beginnen met een aantal niet-gepubliceerde studies die ook geen evidentie toonden voor AONEPE. Vansteenwegen (1994) vond geen stabiele overdracht van emotionele expressie bij het paren van neutrale gezichtsuitdrukkingen met kwade of angstige gezichten. Olson en Fazio (2002) slaagden er niet in om op associatieve wijze de perceptie van grootte te beïnvloeden. De Houwer, Meersmans, Baeyens, en Eelen (2002), voerden een reeks experimenten rond AONEPE uit die ook geen stabiele effecten opleverden. Er werd geen evidentie gevonden voor o.a. de overdracht van helderheid van

foto's of de semantische categorie van woorden. Tenslotte vond Eva Walther (persoonlijke communicatie, 23 April, 2003) geen evidentie voor het overdragen van het kenmerk 'ambitie' tussen foto's van mensen.

Daarnaast zijn er ook een aantal fenomenen die sterk op AONEPE lijken. Een eerste fenomeen, dat binnen deze categorie valt, zijn de zogenaamde conditionele naeffecten. Het conditionele kleurnaffect, als eerste gedemonstreerd door Celeste McCollough (McCollough, 1965), is het meest onderzochte. Na het herhaaldelijk aanbieden van een rood verticaal raster en een groen horizontaal raster, werden witte horizontale rasters als roodachtig en witte verticale rasters als groenachtig gepercipieerd. Het lijkt alsof het herhaaldelijk aanbieden van een raster met een bepaalde oriëntatie (VP) in een bepaalde kleur (OP) resulteert in een verschuiving van de kleurperceptie van de VP in de tegenovergestelde richting van de kleur waarin de VP oorspronkelijk was aangeboden. Niet iedereen ziet deze conditionele naeffecten echter als een vorm van klassieke conditionering. McCollough zelf verklaarde deze oriëntatiespecifieke kleurnaffecten als een vorm van adaptatie van oriëntatiegevoelige neuronen: Het herhaaldelijk aanbieden van een raster in een zeker kleur resulteert in een adaptatie van de oriëntatiegevoelige randdetectoren voor die kleur. De laatste decennia zijn verschillende variaties van zowel klassieke conditioneringtheorieën als neuronale adaptatie theorieën ontwikkeld, maar geen van deze modellen kan alle karakteristieken verklaren van de conditionele naeffecten. Voor een uitgebreid onderzoek naar conditionele kleurnaffecten en de theorieën hier rond, verwijzen we de lezer naar Silverans (1998). Aangezien het moeilijk is om te besluiten of deze conditionele naeffecten wel daadwerkelijk een vorm van klassieke conditionering zijn, en niet een meer basaal neuronale adaptatie-effect, kunnen we niet besluiten dat het evidentie biedt voor AONEPE.

Een tweede fenomeen dat sterk lijkt op AONEPE, is een vorm van impliciet leren die bekend staat als 'verborgen covariatiedetectie'. In studies rond verborgen covariatiedetectie worden prikkels aangeboden die meestal een simpele, maar niet-saillante covariatie bevatten tussen de elementen x en y. Een deel van deze studies toont impliciet leren (van de verborgen covariatie) aan door nieuwe stimuli aan te bieden die element x bevatten, maar geen informatie bevatten omtrent de aan- of afwezigheid van element y. Als het oordeel van de deelnemers met betrekking tot y de x-y covariatie reflecteert, besluit men dat de x-y covariatie werd geleerd. Lewicki en collega's voerden veel experimenten

uit die in deze categorie passen. Bijvoorbeeld, Lewicki (1986b) liet deelnemers foto's van vrouwen zien samen met een verbale beschrijving van de personaliteit van die vrouw (ofwel als capabel ofwel als vriendelijk). Er was over de stimuli heen een covariatie tussen de haarlengte (kort of lang haar) en de toegeschreven capabiliteit of vriendelijkheid. Na de acquisitiefase kregen de proefpersonen foto's te zien van andere vrouwen. Deze keer moesten de deelnemers aangeven of deze vrouwen capabel of vriendelijk waren. De deelnemers vonden de vrouwen met de haarlengte die voordien geassocieerd werd met capabel, meer capabel dan de vrouwen met de haarlengte die voordien geassocieerd werd met vriendelijk. Het lijkt dus alsof een persoonlijkheidskenmerk van vrouwen (OP) werd overgedragen via de haarlengte (VP) die ermee covarieerde. In de meeste van deze experimenten bleek echter de covariatie tussen de VP en de OP meer te zijn dan het simpelweg samen aanbieden van deze prikkels: De VP was meestal gerelateerd aan de OP via een propositionele relatie van het type 'de VP is een instantie van de OP' of 'de VP is gekenmerkt door de OP' (Hendrickx, De Houwer, Baeyens, Eelen, & Van Avermaet, 1997). We kunnen dus niet bewijzen dat we hier met zuiver associatief leren te maken hebben waarin het simpelweg gepaard aanbieden van de VP en OP tot een transfer leidt. Bovendien zijn deze effecten niet altijd makkelijk te repliceren (zie Hendrickx et al., 1997) en zouden ze deels gemedieerd kunnen zijn door EC. Bijgevolg hebben we hier ook geen evidentie voor zuivere AONEPE.

Hoofdstuk 7: Algemene discussie en besluit

Het hoofddoel van dit doctoraat was het vinden van evidentie voor AONEPE buiten het olfactorische-gustatorische domein. In een eerste reeks experimenten (Hoofdstuk 3) probeerden we de leeftijd of het geslacht van de ene (afgebeelde) persoon op de andere over te dragen, door simpelweg de foto's van deze personen gepaard aan te bieden. We vonden in één experiment evidentie voor geslachtsoverdracht, maar we konden dit resultaat niet repliceren in de vier volgende experimenten. In een tweede reeks experimenten (Hoofdstuk 4), probeerden we twee keer het geslacht van duidelijk mannelijke/vrouwelijke voornamen over te dragen op Japanse karakters. In één experiment vonden we evidentie voor een overdrachtseffect in een maat die normaal ongevoelig is voor vraageffecten, maar dit effect konden we niet repliceren in een volgend experiment. In een laatste reeks experimenten (Hoofdstuk 5) onderzochten we niet de *overdracht* van niet-evaluatieve

prikkeleigenschappen van een OP op een VP, maar richtten we ons op een subproces dat waarschijnlijk een voorwaarde is voor AONEPE: *activatie* van niet-evaluatieve prikkeleigenschappen. We probeerden de richting en valentie van een aversieve OP te associëren met de kleur van een woord. In de eerste twee experimenten vonden we een onverwacht contrasteffect: de kleur die gepaard was met een OP van rechts (links) bleek vervolgens een linkse (rechtse) respons te faciliteren. Jammer genoeg konden we dit effect niet repliceren in de drie daaropvolgende experimenten. Tenslotte beschreven we kort een aantal studies, uitgevoerd in andere laboratoria, die ook faalden om AONEPE te vinden, en bespraken we ook een aantal fenomenen die op AONEPE lijken (Hoofdstuk 6).

De data die totnogtoe werden verzameld rond AONEPE zijn teleurstellend. We observeerden een aantal keer significante overdrachtseffecten (in Experimenten P3, P8, S1, en S2). Jammer genoeg konden we deze effecten niet repliceren of konden ze mogelijk veroorzaakt zijn door vraageffecten. We vonden dus weinig overtuigende evidentie voor reële overdracht. Dit wil zeggen, de geobserveerde overdrachtseffecten reflecteerden misschien geen echte verandering in de perceptie van de VP, maar konden veroorzaakt zijn door het feit dat de deelnemers omwille van vraageffecten bewust en intentioneel hun kennis van de VP-OP paren gebruikten in hun beoordeling van de VP na de acquisitiefase.

De beste evidentie voor AONEPE met visuele prikkels komt van studies die conditionele kleurnaeffecten onderzochten. Echter, verschillende basale neuronale verklaringen bestaan voor dit fenomeen. Het is dus niet duidelijk of deze conditionele naeffecten een vorm zijn van klassieke conditionering. Verborgene covariatiedetectie studies, uitgevoerd door Lewicki en collega's, suggereren dat eigenschapsoverdracht voor kan komen wanneer prepositionele relaties van het type 'de VP is een OP' of 'de VP wordt gekenmerkt door de OP-eigenschap' bestaan tussen de VP en de OP. Maar naast het feit dat in veel van deze studies de niet-evaluatieve transfer veroorzaakt kan zijn door evaluatieve transfer, lijkt het alsof meer dan het louter samen aanbieden van de twee prikkels nodig is om het overdrachtseffect te bekomen. Daarenboven zijn deze verborgene covariatie-effecten moeilijk te repliceren (Hendrickx et al., 1997).

Het is onduidelijk waarom we geen evidentie vonden voor AONEPE. Men zou kunnen besluiten dat associatieve overdracht van niet-evaluatieve prikkeleigenschappen niet bestaat bij visuele stimuli. Het is echter evengoed mogelijk dat deze vorm van leren

wel bestaat maar alleen gevonden kan worden als aan strikte grenscondities wordt voldaan. Wat deze grenscondities dan zijn, is nog onduidelijk. Een mogelijke grensconditie is het soort van prikkeleigenschap dat men probeert over te dragen. Het kan zijn dat we prikkeleigenschappen hebben geselecteerd die moeilijk over te dragen zijn. Maar het is dan onduidelijk welke eigenschappen wel overgedragen kunnen worden. We moeten ook opmerken dat wij én andere onderzoekers al een verscheidenheid van eigenschappen hebben proberen over te dragen: leeftijd en geslacht van afgebeelde personen, geslacht van voornamen, emotionele expressie, grootte, helderheid van foto's, semantische woordcategorie en ambitie van personen.

Het is mogelijk dat er andere (nog niet gekende) grenscondities zijn, die ons weerhielden om evidentie te vinden voor algemene AONEPE. Het is interessant om hier op te merken dat er in EC ook, tot nog toe niet geïdentificeerde, grenscondities zijn. In hun review van de literatuur over EC, merkten zowel De Houwer en collega's (De Houwer, Thomas, & Baeyens, 2001) en Field (in druk) op dat er verschillende studies zijn waarin EC niet gerepliceerd kon worden (o.a., Field, 2002; Field & Davey, 1999; Rozin et al., 1998). Of de grenscondities die voor EC gelden dezelfde zijn als diegene die gelden voor AONEPE is niet duidelijk.

Samengevat kunnen we zeggen dat we geen eenduidige evidentie hebben gevonden voor het bestaan van puur AONEPE buiten het olfactorische-gustatorische domein. Toch kunnen we niet beweren dat AONEPE niet bestaat, aangezien het mogelijk is dat een aantal grenscondities voor het bekomen van AONEPE niet vervuld waren. Maar op z'n minst kunnen we, op basis van onze studies en de verschillende niet gepubliceerde studies van anderen, zeggen dat de grenscondities die gelden voor deze vorm van leren veelvuldig zijn.

References

- Allan, L. G., & Tirimacco, N. (1987). An orientation-contingent achromatic aftereffect. *Bulletin of the Psychonomic Society*, 25, 54-55.
- Baeyens, F. (1998). Koele witte wijn, als plots de deurbel rinkelt: Referentie en evaluatie versus verwachting en voorbereiding bij Pavlovians leren. [Chilled white wine, when all of a sudden the doorbell rings: Reference and evaluation versus expectancy and preparation in Pavlovian learning]. *Gedragstherapie*, 31, 7-47.
- Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). *Behaviour Research and Therapy*, 33, 825-831.
- Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996). No evidence for modulation of evaluative flavor-flavor associations in humans. *Learning and Motivation*, 27, 200-241.
- Baeyens, F., Crombez, G., Hendrickx, H., & Eelen, P. (1995). Parameters of human evaluative flavour-flavor conditioning. *Learning and Motivation*, 26, 141-160.
- Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy*, 10, 179-199.
- Baeyens, F., Eelen, P., & Crombez, G. (1995). Pavlovian associations are forever: On classical conditioning and extinction. *Journal of Psychophysiology*, 9, 127-141.
- Baeyens, F., Eelen, P., Crombez, G., & De Houwer, J. (2001). On the role of beliefs in observational flavor conditioning. *Current Psychology: Developmental, Learning, Personality, Social*, 20, 183-203.
- Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy*, 30, 133-142.
- Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition and Emotion*, 4, 3-18.
- Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989a). Acquired affective-evaluative value: Conservative but not unchangeable. *Behaviour Research and Therapy*, 27, 279-287.
- Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989b). The influence of CS-UCS perceptual similarity/dissimilarity on human evaluative learning and signal learning. *Learning and Motivation*, 20, 322-333.

References

- Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation*, 21, 434-455.
- Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning: Acquired valence is sensitive to US-revaluation. *Learning and Motivation*, 23, 200-224.
- Baeyens, F., Hendrickx, H., Crombez, G., & Hermans, D. (1998). Neither extended sequential nor simultaneous feature positive training result in modulation of evaluative flavor-flavor conditioning in humans. *Appetite*, 31, 185-204.
- Baeyens, F., Heremans, R., Eelen, P., & Crombez, G. (1993). Hidden-covariation detection and imagery ability. *European Journal of Cognitive Psychology*, 5, 435-456.
- Baeyens, F., Hermans, D., & Eelen, P. (1993). The role of CS-US contingency in human evaluative conditioning. *Behaviour Research and Therapy*, 31, 731-737.
- Baeyens, F., Vanhouche, W., Crombez, G., Eelen, P. (1998). Human evaluative flavor-flavor conditioning is not sensitive to post-acquisition US-inflation. *Psychologica Belgica*, 38, 83-108.
- Baeyens, F., Vansteenwegen, D., De Houwer, J., & Crombez, G. (1996). Observational conditioning of food valence in humans. *Appetite*, 27, 235-250.
- Baeyens, F., Vansteenwegen, D., Hermans, D., & Eelen, P. (2001a). Chilled white wine, when all of a sudden the doorbell rings: Mere reference and evaluation versus expectancy and preparation in human Pavlovian learning. In F. Columbus (Ed.), *Advances in psychology research* (Vol. 4, pp. 241-277). Nova Science Publishers: Huntington, NY.
- Baeyens, F., Vansteenwegen, D., Hermans, D., & Eelen, P. (2001b). Human evaluative flavor-taste conditioning: Conditions of learning and underlying processes. *Psychologica Belgica*, 41, 169-186.
- Baeyens, F., Wrzesniewski, A., De Houwer, J., & Eelen, P. (1996). Toilet rooms, body massages, and smells: Two field studies on human evaluative odor conditioning. *Current Psychology: Developmental, Learning, Personality, Social*, 15, 77-96.
- Buchner, A., Erdfelder, E., & Faul, F. (1997). How to Use G*Power [Software Manual]. Retrieved from http://www.psych.uni-duesseldorf.de/aap/projects/gpower/how_to_use_gpower.html
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum: Hillsdale, NJ.
- Davey, G. C. L. (1994). Defining the important theoretical questions to ask about evaluative conditioning: A reply to Martin and Levey (1994). *Behaviour Research and Therapy*, 32, 307-310.

- De Houwer, J. (1998a). Leren is eenvoudig, doen is moeilijk: Een nieuwe visie op het onderscheid tussen Pavlovianse en evaluatieve conditionering. [Learning is simple, performance is difficult: A new look at the distinction between Pavlovian and evaluative conditioning.]. *Gedagstherapie*, 31, 49-66
- De Houwer, J. (1998b). The semantic Simon effect. *Quarterly Journal of Experimental Psychology*, 51A, 683-688.
- De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology*, 50, 77-85.
- De Houwer, J., Baeyens, F., & Eelen, P. (1994). Verbal evaluative conditioning with undetected US presentations. *Behaviour Research and Therapy*, 32, 629-633.
- De Houwer, J., Baeyens, F., & Hendrickx, H. (1997). Implicit learning of evaluative associations. *Psychologica Belgica*, 37, 115-130.
- De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behaviour Processes*, 26, 237-242.
- De Houwer, J., Crombez, G., Baeyens, F., & Hermans, D. (2001). On the generality of the affective Simon effect. *Cognition and Emotion*, 15, 189-206.
- De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion*, 12, 45-61.
- De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with "subliminally" presented stimuli. *Consciousness and Cognition: An International Journal*, 6, 87-107.
- De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition and Emotion*, 12, 145-169.
- De Houwer, J., Meersmans, T., Baeyens, F., & Eelen, P. (2002, April). Associative transfer of non-evaluative stimulus properties. In A. Field & J. De Houwer (Chair), *Evaluative Conditioning*. Symposium conducted at the Experimental Psychology Conference, Leuven, Belgium.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853-869.
- Diaz, E., Baeyens, F., & Ruiz, G. (2002, May). Resistance to extinction on human evaluative conditioning using between-subjects design. *Special Interest Meeting on Human Evaluative Conditioning*, Le Lignely, Belgium.
- Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, 127, 286-303.

References

- Durgin, F. H. (1996). Visual aftereffect of texture density contingent on color of frame. *Perception and Psychophysics*, 58, 207-223.
- Eifert, G.H., Craill, L., Carey, E., & O'Connor, C. (1988). Affect modification through evaluative conditioning with music. *Behaviour Research and Therapy*, 26, 321-330.
- Faul, F. & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS (Version 2.0) [Computer software]. Retrieved from <http://www.pscho.uni-duesseldorf.de/aap/projects/gpower> . Bonn, FRG: Bonn University, Department of Psychology.
- Favreau, O. E., Emerson, V. F., & Corballis, M. C. (1972). Motion perception: A color-contingent aftereffect. *Science*, 176, 78-79.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229-238.
- Field, A. P. (2000). I like it, but I'm not sure why: Can evaluative conditioning occur without conscious awareness? *Consciousness and Cognition: An International Journal*, 9, 13-36.
- Field, A. P. (2002, May). Evaluative conditioning: missing presumed dead. *Special Interest Meeting on Human Evaluative Conditioning*, Le Lignely, Belgium.
- Field, A. P. (in press). Learning to like (or dislike): Associative learning of preferences. In A. J. Wills (Ed.), *New directions in human associative learning*. Lawrence Erlbaum: London, UK.
- Field, A. P., & Davey, G. C. L. (1997). Conceptual conditioning: evidence for an artifactual account of evaluative learning. *Learning and Motivation*, 28, 446-464.
- Field, A. P., & Davey, G. C. L. (1998). Evaluative conditioning: arti-fact or -fiction? A reply to Baeyens, De Houwer, Vansteenwegen, and Eelen (1998). *Learning and Motivation*, 29, 475-491.
- Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 211-224.
- Francken, G. (2003). *Resistance to extinction in evaluative conditioning*. Unpublished doctoral dissertation, University of Leuven, Leuven, Belgium.
- Hammerl, M., & Grabitz, H. J. (1993). Human evaluative conditioning: Order of stimulus presentation. *Integrative Physiological and Behavioral Science*, 28, 191-194.
- Hammerl, M., & Grabitz, H. J. (1996). Human evaluative conditioning without experiencing a valued event. *Learning and Motivation*, 27, 278-293.
- Hammerl, M., & Grabitz, H. J. (2000). Affective-evaluative learning in humans: A form of associative learning or only an artifact? *Learning and Motivation*, 31, 345-363.

- Hammerl, M., Bloch, M., & Silverthorne, C. P. (1997). Effects of US-alone presentations on human evaluative conditioning. *Learning and Motivation*, 28, 491-509.
- Hardwick, S. A., & Lipp, O. V. (2000). Modulation of affective learning: An occasion for evaluative conditioning? *Learning and Motivation*, 31, 251-271.
- Hendrickx, H. (1999). [Associatively transferring age: age ratings]. Unpublished raw data.
- Hendrickx, H., De Houwer, J., Baeyens, F., Eelen, P., & Van Avermaet, E. (1997). Hidden covariation detection might be very hidden indeed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 201-220.
- Hermans, D., Clarysse, J., Baeyens, F., & Spruyt, A. (2002). Affect (Version 1.0 and 3.0) [Computer software]. Retrieved from <http://www.psy.kuleuven.ac.be/leerpsy/affect>. University of Leuven, Belgium.
- Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In S.P. Shohov (Ed.), *Advances in psychology research* (Vol. 12, pp. 17-40). Nova Science Publishers: Huntington, NY.
- Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion*, 8, 515-533.
- Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition and Emotion*, 17, 83-99.
- Hermans, D., Van den Broeck, A., & Eelen, P. (1998). Affective priming using a color-naming task: A test of an affective-motivational account of affective priming effects. *Zeitschrift fuer Experimentelle Psychologie*, 45, 136-148.
- Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, 40, 217-234.
- Hill, T., Lewicki, P., Czyzewska, M., & Boss, A. (1989). Self-perpetuating development of encoding biases in person perception. *Journal of Personality and Social Psychology*, 57, 373-387.
- Hill, T., Lewicki, P., Czyzewska, M., & Schuller, G. (1990). The role of learned inferential encoding rules in the perception of faces: Effects of nonconscious self-perpetuation of a bias. *Journal of Experimental Social Psychology*, 26, 350-371.
- Hill, T., Lewicki, P., & Neubauer, R. M. (1991). The development of depressive encoding dispositions: A case of self-perpetuation of biases. *Journal of Experimental Social Psychology*, 27, 392-409.

References

- Holland, P. C. (1983). "Occasion-setting" in Pavlovian feature-positive discriminations. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behaviour: Discrimination processes* (Vol. 4, pp. 183-287). Ballinger: Cambridge, MA.
- Johnsrude, I. S., Owen, A. M., Zhao, W. V., & White, N. M. (1999). Conditioned preference in humans: A novel experimental approach. *Learning and Motivation*, 30, 250-264.
- Kim, J., Allen, C. T., & Kardes, F. R. (1996). An investigation of the mediational mechanisms underlying attitudinal conditioning. *Journal of Marketing Research*, 33, 318-328
- Lewicki, P. (1986a). *Nonconscious social information processing*. Academic Press: San Diego, CA.
- Lewicki, P. (1986b). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 135-146.
- Lewicki, P., Hill, T., & Sasaki, I. (1989). Self-perpetuating development of encoding biases. *Journal of Experimental Psychology: General*, 118, 323-337.
- Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation*, 34, 219-239.
- Lipp, O. V., & Purkis, H. M. (in press). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition and Emotion*.
- Lovibond, P. F., Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 3-26.
- Mackintosh, N. J. (1983). *Conditioning and associative learning*. Oxford University Press: New York, NY.
- Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy*, 1, 57-102.
- Martin, I., & Levey, A. B. (1987). Learning what will happen next: Conditioning, evaluation, and cognitive processes. In G. Davey (Ed.), *Cognitive processes and Pavlovian conditioning in humans* (pp. 57-81). John Wiley & Sons: Oxford, England.
- Martin, I., & Levey, A. B. (1994). The evaluative response: Primitive but necessary. *Behaviour Research and Therapy*, 32, 301-305.
- McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science*, 149, 1115-1116
- Miller, J. (1987). Priming is not necessary for selective-attention failures: Semantic effects of unattended, unprimed letters. *Perception and Psychophysics*, 41, 419-434.

- Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, 12, 413-417.
- Olson, M. A., & Fazio, R. H. (2002, February). *Implicit covariation learning: evaluative versus non-evaluative associations*. Poster session presented at the annual meeting of the Society for Personality and Social Psychology, Savannah.
- Page, M. M. (1969). Social psychology of a classical conditioning of attitudes experiment. *Journal of Personality and Social Psychology*, 11, 177-186.
- Pavlov, I. (1927). *Conditioned Reflexes*. Oxford University Press: New York, NY.
- Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation*, 32, 84-99.
- Razran, G. (1954). The conditioned evocation of attitudes (cognitive conditioning?). *Journal of Experimental Psychology*, 48, 278-282.
- Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology*, 66, 1-5.
- Rozin, P., & Millman, L. (1987). Family environment, not heredity, accounts for family resemblances in food preferences and attitudes: A twin study. *Appetite*, 8, 125-134.
- Rozin, P., & Zellner, D. (1985). The role of Pavlovian conditioning in the acquisition of food likes and dislikes. *Annals of the New York Academy of Sciences*, 443, 189-202.
- Rozin, P., Wrzesniewski, A., & Byrnes, D. (1998). The elusiveness of evaluative conditioning. *Learning and Motivation*, 29, 397-415.
- Schwarzer, R. (1989). Statistics software for meta-analysis (Version 5.3) [Computer software and manual]. Retrieved from http://www.fu-berlin.de/gesund/gesu_engl/meta_e.htm.
- Seeger, C. A. (1994). Implicit learning. *Psychological Bulletin*, 115, 163-196.
- Silverans, P. (1998). Een experimentele klassieke conditioneringsanalyse van conditionele kleurnaeffecten. [An experimental classical conditioning-analysis of conditional colour aftereffects]. Unpublished master's thesis, University of Leuven, Leuven, Belgium.
- Simon, J.R. (1990). The effects of an irrelevant directional cue on human information processing. In R.W. Proctor & T.G. Reeve (Eds.) *Stimulus-response compatibility: An integrated perspective* (pp.31-86). Amsterdam, North Holland.
- Skowbo, D., Gentry, T., Timney, B., & Morant, R. B. (1974). The McCollough effect: Influence of several kinds of visual stimulation on decay rate. *Perception and Psychophysics*, 16, 47-49.
- Staats, C. K., & Staats, A. W. (1957). Meaning established by classical conditioning. *Journal of Experimental Psychology*, 54, 74-80.

References

- Stevenson, R. J. (2001a). Associative learning and odor quality perception: How sniffing an odor mixture can alter the smell of its parts. *Learning and Motivation*, 32, 154-177.
- Stevenson, R. J. (2001b). The acquisition of odour qualities. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 54, 561-577.
- Stevenson, R. J., & Boakes, R. A. (in press). Sweet and sour smells: Learned synesthesia between the senses of taste and smell. In C. Spence, G. Calvert, & B. Stein (Eds.), *Handbook of multisensory processing*. MIT Press: Boston, MA.
- Stevenson, R. J., Boakes, R. A., & Prescott, J. (1998). Changes in odor sweetness resulting from implicit learning of a simultaneous odor sweetness association: An example of learned synesthesia. *Learning and Motivation*, 29, 113-132.
- Stevenson, R. J., Boakes, R. A., & Wilson, J. P. (2000). Counter-conditioning following human odor-taste and color-taste learning. *Learning and Motivation*, 31, 114-127.
- Stevenson, R. J., Boakes, R. A., & Wilson, J. P. (2000). Resistance to extinction of conditioned odor perceptions: Evaluative conditioning is not unique. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 423-440.
- Stevenson, R. J., & Case, T. I. (2003). Preexposure to the stimulus elements, but not training to detect them, retards human odour-taste learning. *Behavioural Processes*, 61, 13-25.
- Stevenson, R. J., Case, T. I., & Boakes, R. A. (2003). Smelling what was there: Acquired olfactory percepts are resistant to further modification. *Learning and Motivation*, 34, 185-202.
- Stevenson, R. J., Prescott, J., & Boakes, R. A. (1995). The acquisition of taste properties by odors. *Learning and Motivation*, 26, 433-455.
- Todrank, J., Byrnes, D., Wrzesniewski, A., & Rozin, P. (1995). Odors can change preferences for people in photographs: A cross-modal evaluative conditioning study with olfactory USs and visual CSs. *Learning and Motivation*, 26, 116-140.
- Vansteenwegen, D. (1994). *In welke mate wordt de perceptie van een prikkel beïnvloed door zijn referentiele relaties met andere prikkels?* [To what extent can the perception of a stimulus be influenced by its referential relation with other stimuli]. Unpublished master's thesis, University of Leuven, Leuven, Belgium.
- Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (1998). Extinction in fear conditioning: Effects on startle modulation and evaluative self-reports. *Psychophysiology*, 35, 729-736.
- Wagner, A. R., & Brandon, S. E. (1989). Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In S.B. Klein & R.R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory* (pp. 149-189). Lawrence Erlbaum Associates: Hillsdale, NJ.

- Walker, J. T., & Irion, A. L. (1979). Two new contingent aftereffects: Perceived auditory duration contingent on pitch and on temporal order. *Perception and Psychophysics*, 26, 241-244.
- Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, 82, 919-934.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.
- Zellner, D. A., Rozin, P., Aron, M., & Kulish, C. (1983). Conditioned enhancement of human's liking for flavour by pairing with sweetness. *Learning and Motivation*, 14, 338-350.

Appendix

Overview of performed experiments

Experiment	CSs	USs	Property to transfer or activate	Dependent variable
Picture-picture studies				
P1	Middle-aged men	Older/younger men	Age	Age-scale
P2	Gender-neutral infants	Male/female infants	Gender	Gender-scale
P3	Gender-neutral infants	Male/female infants	Gender	Gender-scale Gender-priming
P4	Gender-neutral infants	Male/female infants	Gender	Gender-scale Gender-priming
P5	Gender-neutral infants	Male/female infants	Gender	Gender-scale Naming measure
P6	Gender-neutral infants	Male/female infants	Gender	Gender-scale Naming measure
P7	Gender-neutral infants	Positive/negative infants	Valence	Valence-scale Affective-priming
P8	Gender-neutral infants	Male/female infants	Gender	Gender-scale Gender-priming ^a
Symbol-word studies				
S1	Japanese character	Male/female first name	Gender	Gender-scale Gender-priming
S2	Japanese character	Male/female first name	Gender	Gender-scale Gender-priming
Fear conditioning studies				
F1	Word-colour	Loud noise from left/right	Direction and valence	Spatial Simon Affective Simon Valence-scale
F2	Word-colour	Loud noise from left/right	Direction and valence	Spatial Simon Affective Simon Valence-scale
F3	Word-colour	Loud noise from left/right	Direction and valence	Spatial Simon Valence-scale
F4	Word-colour	Loud noise from left/right	Direction and valence	Spatial Simon Valence-scale
F5	Word-colour	Loud noise from left/right	Direction and valence	Spatial Simon Valence-scale

Note. CSs = conditional stimuli or the stimulus that the property is transfer to; USs = unconditional stimuli or the stimulus that the property is transferred from.

^a Due to a programming error the data from this gender-priming task was unusable.