Charles University in Prague Faculty of Philosophy and Arts Department of Psychology

Psychology – Clinical Psychology

Jiří Lukavský

### Physiological correlates and semantic distances in Word Association Test

Dissertation

Dissertation advisor Mgr. Radovan Šikl, PhD.

2008

I hereby declare that I have written this dissertation solely by myself and list all relevant literature and other sources in References section.

In Prague, June 23, 2008

**Title:** Physiological correlates and semantic distances in Word Association Test

Abstract: The project is focused on the physiological reactions during Word Association Test (WAT). The author tests the hypothesis originally proposed by Šlechta (2002b) about separable cognitive and emotional workload, where the cognitive workloads manifests in reaction times (RT) and emotional workload in physiological reactions (e.g. electrodermal activity). The hypotheses were tested on the sample of young healthy adults (N=80), with protocol consisting of 40 Czech nouns. The stimuli varied in their level of concreteness and their emotional valence. During the experiment RTs, electrodermal activity (EDA) and pupillary responses (PR) were measured.

Data support the original findings that RTs and EDA reactions reflect in WAT different situation requirements: cognitive workload and emotional workload, respectively. The effect of cognitive workload is larger and manifests also in EDA reactions, while EDA demonstrated the best sensitivity to emotional valence condition from all employed methods (RTs, EDA, pupillary reactions). The effects are unfortunatelly too small to allow a reliable classification of responses.

The effects of other factors (subjective valence, corpus frequency of responses, semantic proximity measured with Semantic Selection Test) were investigated, but even when significant correlations were found, it didn't help to reduce the variance in physiological data. Název: Fyziologické koreláty a sémantické vzdálenosti ve Slovním asociačním experimentu

Abstrakt: Projekt se zabývá studiem fyziologických reakcí v průběhu Slovního asociačního experimentu (AE). Autor ověřuje hypotézu navrženou Šlechtou (2002b) o oddělených projevech kognitivní a emoční zátěže v AE, kde se kognitivní zátěž projevuje v reakčních časech a emoční zátěž ve fyziologických reakcích, např. elektrodermální aktivitě (EDA). Hypotézy byly ověřovány na vzorku N=80 mladých lidí, s protokoly obsahujícími 40 českých podstatných jmen. Podnětová slova se lišila v úrovni konkrétnosti a emoční valenci. V průběhu AE byly zaznamenávány reakční časy, EDA a pupilární aktivita.

Získané údaje podporují původní tvrzení, že reakční časy a EDA odpovídají odlišným situačním nárokům v AE: kognitivní a emoční zátěži. Efekt kognitivní zátěže je větší a projevuje se i v EDA reakcích. EDA prokázala největší citlivost k emoční valenci ze všech testovaných metod, tento efekt je ale poměrně malý a neumožňuje splehlivě klasifikovat individuální odpovědi.

Byl zkoumán vliv dalších faktorů (subjektivní valence, frekvence odpovědí v jazykovém korpusu, sémantická blízkost měřená Testem sémantického výběru). Přestože byly prokázány stastisticky významné korelace, znalost těchto faktorů neumožňuje efektivně redukovat variabilitu fyziologických dat.

# Contents

Introduction 3			
3			
9			
13			
15			
19			
25			
29			
33			
40			
42			
44			
45			
46			
47			

	4.2	Physic	ological results	50
		4.2.1	Correlations of physiological parameters	50
		4.2.2	Effect of emotional workload	51
		4.2.3	Effect of cognitive workload	53
		4.2.4	Comparison of both effects	54
		4.2.5	Discrimination based on physiological parameters	56
	4.3	Seman	tic and linguistic results	56
		4.3.1	Individual differences in SST	56
		4.3.2	Effects of order and attribute picture sets	58
		4.3.3	Proximities measured with SST and MI	60
		4.3.4	Subjective valence rating	60
		4.3.5	WAT responses	64
		4.3.6	Stimulus – response relationship in language	65
	4.4	Combi	ined results	66
		4.4.1	Response frequency and its behavioral correlates	66
		4.4.2	Subjective valence and physiological reactions	68
		4.4.3	Subjective semantic proximity and physiological reac-	
			tions	69
<b>5</b>	Disc	cussion	1	<b>74</b>
	5.1	Procee	lure	74
	5.2	Stimul	li material	75
	5.3	Stabili	ty of results	77
6	Con	clusio	n and implications	78
	6.1	Result	s replication	78
	6.2	Interp	retation of physiological responses	80
	6.3	Using	SST within WAT procedure	83
	6.4	Summ	ary	84
$\mathbf{A}$	App	endix		85
	A.1	Word	stimuli	85
Re	efere	nces		89

### Preface

Five years ago I started to work with Dr. Petr Slechta in the research on physiological reactions during Word Association Test (WAT). It was an exciting experience to administer WAT to various people in various health conditions, to see their behavior and inspect the data. In the end we could conclude (e.g. Šlechta, 2002a) that there are statistically significant differences between healthy people and psychiatric patients and this method could be used to discriminate between them.

When I was thinking about this research, I was curious whether this approach can be used on a finer level. Can we use it for discrimination in healthy people? For example, if we observe higher reaction in word A and we know higher reactions are usually associated with emotional valence – can we reliably conclude that there is some emotional charge connected with word A?

It is probable that there are differences in understanding words between people – the words are the same, people understand their meanings, but the words can evoke different reactions based on the previous personal experiences, which are unique. Can we estimate the perceived meanings by some other measure?

For this reason I started to use Semantic Selection Test, which is described later, and I also tried to use the language data from corpora to better understand the relationships between stimuli and reactions. In short, this work is an attempt to inspect individual differences in WAT observed in healthy people and to measure the differences on the level of word content or meaning.

I would like to thank my dissertation advisor Dr. Radovan Šikl, for his helpful comments about my work, and for the trust to do my work well. Dr. Petr Šlechta was a great source of inspiration for me, and I am grateful for the experience working with him. Roman Mlejnek helped me with data collection and analysis of the audio recordings. The work on this project was supported by Czech Science Foundation (406/06/1772).

## Chapter 1

## Introduction

In this section, the background for the following research will be covered. First, the brief history and development of Word Association Test will be described. Later, we will focus on the combined measurements of WAT and physiological correlates and describe the physiological background behind the methods, which were used in the experiment.

Further the Semantic Selection Test will be described and its basic psychometric properties introduced. Finally, Mutual information (MI) index will be introduced.

The research questions will be described in detail in Chapter 2.

### 1.1 Word Association Test

The concept of associations belongs among the oldest and likely among the most productive constructs in psychology. This concept based a theoretical essence of associanism, but it proved to be independent and was found very useful in neuropsychology and later in cognitive psychology, particularly in connectionism (McClelland et al., 1986), where it was theoretically

assimilated in a novel way.

The early origins of the association concept can be traced back to Aristotle and his manuscript on Memory and Recollection, in which he described the mechanisms of recollection based on similarity, opposition and contiguity. The English empiricists, John Locke (1689) and especially David Hume (1748), further analyzed the ways, how ideas were being associated. Various "laws of association" were formulated by Thomas Brown, James Mill, and John Stuart Mill (see Spitzer, 1994). Brown distinguished between primary and secondary laws of association. The primary laws were similar to Aristotle's ones i.e. similarity, contrast, and temporal and spatial contiguity. The secondary laws affected the association strength and included e.g. the duration and vividness of the sensation, or the frequency of repetition.

With the rise of physiology and neuroanatomy in the second half of the nineteenth century, the concept of association of ideas was combined with the findings on neuron associations. This combination gave birth to the speculative neurophysiological metaphor, which influenced important scientists of the time - e.g. Sigmund Freud or William James. In this metaphor, the associations between ideas correspond to the associations on the brain level.

The Word Association Test (further only WAT) as a research method was introduced to psychology by Francis Galton (1879, 1883). By the same year, Wilhelm Wundt founded the first psychological laboratory in Leipzig, where a great deal of research on word associations was performed. Wundt's assistant James McKeen Cattel (1886) constructed an apparatus (voice key) to measure reaction times in association tasks, which enhanced further scientific development of the method. J.M.Cattell in the summary of his research (Cattell & Bryant, 1889) reported for example the effects of concreteness (responses to concrete nouns were faster compared to responses to abstract nouns). He also pointed out the automaticity of some responses, when one stimulus word led to the same responses in different people. During this time several classification schemas for association responses were developed (Cattell & Bryant, 1889; Spitzer, 1992, for review).

In the early studies, WAT was used as a research method to reveal the operations of mind. Emil Kraepelin, who also worked for two years in the Leipzig laboratory, was the first one, who started to use WAT not only as a research method but also as a diagnostic tool. In his studies (Kraepelin, 1892), partly conducted on himself, he reports the effects of alcohol or coffein on the word association responses. His co-worker Gustav Aschaffenburg performed a large number of experiments in normal subjects and mentally ill patients. They both constructed their own classifications of the association types (Spitzer, 1992, for review).

Carl Gustav Jung in the early years of his career devoted himself to the study of word associations. Under the supervision of Eugen Bleuler he conducted a number of experiments in normal people and in patients. Jung attempted to construct tentative association norms, which could be used in clinical practice. Jung's contributions to the WAT methodology were highly important. First, he drew attention to gender and education level as other explanatory variables. Second, he proposed a novel design using divert attention (similar to dual task method). Subjects were asked to make strokes with a pencil in a rhythm of metronome during the WAT, which led to the increase of clang (i.e. superficial) associations. The observed effect was similar to the previous findings of Kraepelin and Aschaffenburg on the effect of alcohol, which they attributed to a general somatic fatigue. According to Jung's results, both could be explained as attentional phenomena. Third, Jung started to interpret association failures (subject is not able to respond to a stimulus word, or responds with very long delay), which were formerly ignored by Kraepelin's school as experimental errors. Jung assumed the existence of emotional complexes, which are able to interfere with the association task Jung (1910). Forth, Jung was also the first researcher, who studied the changes in skin conductance and breath frequency during WAT (Jung, 1907/1991). His research was based on the findings of Feré (1888) and Tarchanoff, who are considered to be the main discoverers of electrodermal activity (further only EDA).

While Jung's approach relied on interpretative work and Kraepelin's school used the association classifications, Kent and Rosanoff (1910a, 1910b) developed a different approach. They were not satisfied with a priori classification of responses and compiled large empirical norms (N=1000, 100 words). They classified the subject's responses as common, individual, and doubtful based on the frequencies in their norms. The common responses were included in the norms for an appropriate stimulus word, the individual responses were not. The doubtful responses were grammatical variants of the common responses. Kent and Rosanoff (1910a) found out that individual and doubtful responses form only 8.3% of all responses in healthy subjects, but in patients this tendency to report common associations is weakened (Kent & Rosanoff, 1910b).

After the wave of big interest in the first decade of the 20th century there comes a decrease in interest in WAT. Spitzer (1992) suggested several possible reasons for this trend. The lower interest is likely to be connected with the rise of gestaltism and the subsequent criticism of associanism. Similarly, behaviorism probably also contributed to this trend. Jung's hermeneutic approach pushed WAT away from the original scientific thinking and on the clinical field WAT had to compete with new methods (Rorschach test, Thematic Apperception Test). Further, in the beginning of 20th century there were no psychometric tools to analyze WAT and it was dependent on detailed, extensive and up-to-date norms.

In 1960's and 70's the interest in WAT rose again (with 170 publications in 1968, according to PsycLIT database, see Fig.1.1). The rise was caused by the neobehavioristic interest in language behavior and increasing interest in cognitive functioning. WAT as a diagnostic method has not been subject of major changes, however in the research field new techniques allowing more experimental control occurred. The standard WAT benefits from being flexible and easy to administer, but the controllability of the experiment is limited. From one association pair we obtain two dependent variables: the reaction time and the response word. The latter is generally unpredictable and therefore the analysis of association types can be done only post hoc. This disadvantage was overcome by the introduction of priming experiments combined e.g. with lexical decision task<sup>1</sup>.

The idea of associations between concepts appeared again in that time. Several network models of semantic memory were proposed in order to explain the code, in which the information is stored. In general, the networks consist of nodes and links, where nodes represent concepts and the links represent their mutual connections. There are two traditions of semantic network models (Griffith, Steyvers, & Tenenbaum, 2007). In the first tradition (Collins & Quillian, 1969), the links between the nodes represent well-defined relation ships (e.g. "is-a" or "has" relations) and the network forms a hierarchical structure. The closeness of two concepts is defined by the number

<sup>&</sup>lt;sup>1</sup>In these experiments the task is to report, whether the target words are meaningful or meaningless (lexical decision). The targets are preceded by prime words and the relationship between primes and targets influences the performance in the task.



Figure 1.1: Publications on word associations in PsycLIT database for individual years (Šlechta, 2002b)

of links, which must be passed to connect both nodes. The second tradition follows the revision of the model (Collins & Loftus, 1975). In this approach the network structure is more flexible (not hierarchical) and the link length represents the closeness of interconnected nodes. The networks can be constructed e.g. using language corpus data or word association data, which offered a new opportunity for the use of WAT.

Although WAT as a diagnostic method has not been subject of major changes, there are attempts to apply modern psychometric methods, such as Rasch model on WAT results (Ivanouw, 2006). Several currently used diagnostic methods resemble WAT in their name: Implicit Association Test (IAT), Benton Controlled Oral Word Association Test (COWAT) or Karp Objective Word Association Test (KOWAT).

IAT (Greenwald, McGhee, & Schwartz, 1998) measures the association between two concepts (e.g. race and danger, or self and valence) to conclude about subject's attitudes (e.g. implicit racism or self-esteem). COWAT is a verbal fluency test used in neuropsychology. KOWAT (described e.g. in Klionsky et al., 1998) is a standard WAT with rating scheme giving the result scores for 12 scales, such as Antonyms, Rejections or Aggression. The rating of the responses is performed by a computer software and meant to be independent on the administrator. The reliability of the scales ranges from 0.60 to 0.85. The test proved to be useful for discrimination between healthy subjects and patients (Klionsky et al., 1998; Cecero & Karp, 1997).

In Czech professional environment WAT is accessible in Kondáš's adaptation (1989), which is methodically identical with classical method. The diagnostic usage is based on the analysis of reaction times, categorizing association disturbances and less quality associations, and comparison with norms. Association norms of children and adults created by Novák (1992, 1996) are also available.

#### **1.2** Electrodermal activity

As it has been already mentioned in the context of Jung's experiments, there is a long tradition of EDA measurement in psychology. The term "electrodermal activity" was introduced to the literature only in 1966 as a general term for all electrical properties of the skin. In 1967 a terminology commission of the Society of Psychophysiological Research (Brown, 1967) published a proposal for standardization, which is now accepted. It replaces the terms psychogalvanic reflex or galvanic skin reactions, which are now discouraged to use.

The term electrodermal activity (EDA) includes two basic methodological approaches: 1) *endosomatic methods* measuring skin potentials without application of external source of electrical current, 2) exosomatic methods for measuring skin resistance changes, which use external sources of either direct or alternating electrical current. Beside this distinction, differentiation between tonic electrodermal measures and phasic electrodermal measures is also made (Boucsein, 1992). In tonic electrodermal measures the values of skin resistance (or other parameters) are studied in longer, relatively stable periods of time (e.g. in sleep or in certain disorders). In phasic electrodermal measures, the short-term deviations from the general electrodermal level are measured and manipulated by experimental conditions. In this study we focus on exosomatic methods especially on phasic electrodermal changes (skin resistance reactions, SRR). The common terms and their abbreviations are in Table 1.1.

Table 1.1: Methods of electrodermal recording, corresponding parameters and their abbreviations (modified from Boucsein, 1992, p.2)

	A 11 1	Abbreviations		ıs		
Measure	Applied current	in general	tonic	phasic		
Endosomatic methods						
Skin potential	none	SP	SPL	SPR		
Exosomatic methods						
Skin resistance	direct current	$\operatorname{SR}$	SRL	SRR		
Skin conductance	direct current	SC	SCL	SCR		
Skin impendance	alternating current	SZ	SZL	SZR		
Skin admitance	alternating current	SY	SYL	SYR		

The history of EDA research goes back to the 19th century. The first experimental manifestation of electrical properties of skin was performed in 1849 by Emil DuBois-Reymond, but in accordance with the shared opinion of his time he attributed the effect to muscle action potentials. Later Romain Vigouroux, a collaborator of Jean Charcot, measured tonic skin resistance levels in various patient groups. He did not recognize the measured levels as properties of skin and attributed them to the changes in vascular conductivity. Finally, his colleague, Charles Féré, discovered that the changes in skin conductance can be induced by various external stimuli. Russian physiologist working in Germany, Jean de Tarchanoff, reported the measurements of skin potentials without using an external current. Féré and Tarchanoff discovered the exosomatic and endosomatic methods of recording and they are traditionally recognized discoverers of the EDA measurement.

The psychophysiological significance of EDA was understood in the early research. Several common relations were described, e.g. it was known that the stronger stimuli evoke larger reactions, and the repeated stimulations lead to habituation (Dawson, Schell, & Filion, 2000). Also the effects of cognitive effort, emotions, and surprise on eliciting EDA were demonstrated.

The technology progress in the 20th century allowed to replace the old measurement methods (recording on photoactive material, manual scoring) with more reliable methods based on A/D converters<sup>2</sup> and computer administration of the stimuli. Currently the method is claimed to be sufficiently refined and well established (cf. Boucsein, 1992), although with the rise of neuroimaging techniques the research interest moved towards the brain functions. But still there are paradigms and situations, when EDA measurement

<sup>&</sup>lt;sup>2</sup>Analog-to-digital converters, electronic devices converting a continuous signal to a discrete signal

is the method of choice.

The EDA measurement has both several advantages and limitations (Dawson et al., 2000). First, for some researchers EDA is a valuable source of information on the activation of sympathetic nervous system. Unlike other methods the eccrine sweat glands are entirely under sympathetic control, while e.g. heart rate, pupil diameter or blood pressure are controlled by both the sympathetic and parasympathetic system. Second, the occurrence of EDA response (i.e. SRR) is generally quite discriminable and it is obvious by simple inspection of the data. The heart rate responses require a context of longer measurement (several seconds) and some other methods like event related potentials are dependent on averaging of multiple trials. Third, the EDA reactions are associated with the situations that elicit anxiety or frustration, while e.g. heart rate is rather associated with general activation, reward seeking or active avoidance. Thus, the research studying the reactions to the stimuli that elicit anxiety should prefer to employ the EDA measurements. Forth, among other ANS indicators the individual differences in EDA are most reliably associated with psychopathology. Fifth, the EDA measurements are cheap and compared to other techniques the costs for each measurement are negligible.

The major disadvantages of EDA measurement are relatively slow responses of EDA parameters and complex causations. The SRR reaction appears approximately 1-3 seconds after stimulation. This delay eliminates the research of fast occurring phenomena and methods of evoked potentials or prepulse inhibition of the startle blink are usually used instead. A rise in EDA can be caused by various processes (activation, attention, affective intensity), which can complicate the interpretation of the data. To reduce this problem, it is necessary to control the experimental conditions. The EDA method is generally used in three types of paradigms (Dawson et al., 2000). First, the effects of discrete stimuli are studied (e.g. in WAT or in habituation research). The second paradigm is based on the study of continuous stimuli or the activity during longer ongoing tasks. Usually the changes in tonic parameters (SRL) or the frequency of non-specific skin responses are measured. In the third paradigm, the EDA is assumed to be relatively stable within subjects and to reflect some psychological traits. This project is based on the first paradigm.

#### 1.2.1 Physiological mechanisms

The skin serves as a selective barrier between the body and the environment, and helps to maintain the constant core body temperature and water balance. These goals are accomplished through vasodilatation, vasoconstriction, and a variable production of sweat.

There are two types of sweat glands in human body: the *eccrine glands* and *apocrine glands*. In EDA research the psychophysiologists focus on the eccrine glands, which cover most of the body with the highest density on the palms and soles of the feet. Their primary function is in thermoregulation. Sweating in response to environmental stimuli, which evoke emotional reactions, is considered a preparatory adaptation to the future increase of body temperature resulting from higher metabolic activity or fast motor action (Boucsein, 1992). Moreover, the production of sweat on palmar sites facilitates grasping. The apocrine glands are usually interconnected with hair follicles, and distributed mainly in the armpits and in the genital areas. Their function is much less studied in humans. In mammals (e.g dogs or monkeys) they are suggested to be involved in scent production for identification or mating. In humans, this issue is still in question (Dawson et al., 2000).

The human skin consists of three horizontal layers: *subdermis* in deep, *dermis*, and *epidermis* on the surface. The eccrine sweat glands consist of the secretory portion located in the subdermis and the sweat duct channeling the sweat through the dermis and epidermis to the sweat pore on the skin surface. The most outer part of the epidermis, *stratum corneum*, is relatively impermeable to the body fluids and therefore to the electric current. The rise of sweat in the ducts connects the surface of the body with the ion-rich conductive interior, and leads to the rise of skin conductance.

The accepted single-effector model assumes that the changes in EDA can be explained by solely by the activity of sweat glands. It replaced formerly popular two-effector model, which suggested an additional role to an active selective membrane sited in epidermis (cf. Dawson et al., 2000). In the singleeffector model, the sweat ducts are modeled as a group of variable resistors wired in parallel. The rise of sweat in ducts changes the resistor values and results in observable EDA behavior.

The EDA was formerly believed to be mediated by both parts of autonomic nervous system, but later research proved that it is controlled only by sympathetic cholinergic innervation. The central mechanisms of EDA modulation can be summarized into three pathways (Boucsein, 1992). The lowest level mechanism is in the reticular formation of the brainstem and this area links the EDA with general arousal, gross movement and muscle tone. The second level involves the influences of hypothalamus and limbic system, which connects EDA with thermoregulation mechanisms and affective processes. The third and highest level of central EDA control comprises of the influences from cortex and basal ganglia (see Figure 1.2), which link EDA with fine motor control (premotor cortex) and orienting and attention (frontal cortex).



Figure 1.2: Brain centers involved in EDA regulation in humans (Boucsein, 1992). 1. Basal ganglia and cortex, 2. Limbic system, 3. Reticular formation.

#### 1.2.2 Electrodermal measurements during WAT

The first studies employing EDA measurements in WAT were performed by Jung and Peterson in the early years of 20th century. They reported a correspondence between reaction times and EDA reactions. Their results led Peterson to idealized conclusion about galvanometer being "a measurer of the amount of emotion and becoming a new instrument of precision in psychological research" (Peterson & Jung, 1907).

A more thorough study was conducted by Hathaway (1929). He administered to 100 participants a 25-stimulus protocol with words selected from Kent-Rosanoff list. He recorded reaction times and EDA reactions and reported high rank order correlation ( $\rho = 0.60 \pm 0.086$ ). He also concluded that EDA measurements are a more accurate diagnostic measure than reaction time.

Hathaway's experiments were repeatedly replicated with a mixed success. Crosland (Crosland & Beck, 1931) reported insignificant correlations ranging from -0.195 to +0.116, but his study was criticized and his results were later assumed invalid (Hunt & Landis, 1935). Hunt and Landis (1935) addressed these contradictory findings and administered two protocols to 22 participants (14 healthy, 8 patients) on 2 different days. They found mixed results with individual correlations ranging from -0.71 to +0.79. Because only 18% in the first series and 36% in the second series exhibited a reliable positive correlation, they concluded the low reliability on individual level contributed to the mixed results in previous researches and the relationship between EDA and RTs should be interpreted only as a tendency. Heřmanská and Severová (1958) replicated Hunt's and Landis' study in 26 healthy participants. They confirmed consistent but weak positive correlation between RTs and EDA changes.

A similar problem was addressed by Schlegel and Zeier (1982) by different means. They administered a 50-stimulus protocol to 66 students and registered various behavioral parameters (skin resistance response, heart rate, muscle tension, respiration, smiling, body movements) and verbal parameters (reaction times, initial association, reaction time in reproduction test, correct response reproduction). They found significant correlations between reaction time, skin resistance, body movements and smiling in both WAT and subsequent reproduction test.

This project is based on previous research of Slechta (e.g. 2001, 2002b). Šlechta administered WAT in fixed time intervals and measured the association responses, reaction times, and EDA on both healthy subjects and psychiatric patients from various diagnostic groups of ICD-10. He proposed a hypothesis of cognitive and emotional workloads, which cause a RT delay or increased EDA respectively (see Tab. 1.2). His method of computer administered WAT allowed to collect basic WAT data easily and in a short time.

Table 1.2: Typology of stimuli based on SRR amplitude (SRRamp.) and reaction time (RT) proposed by Šlechta (2002b)

SRRamp. RT		Stimuli description
Low Low		Neutral stimuli with little affective loading
		and/or cognitive complications
Low	High/Failure	Stimuli with little meaning for the subject
High	Low	Affectively loaded stimuli which are under
		cognitive control and coped with successfully
High	High/Failure	Conflictive stimuli interfering with cognitive
		processing

In his study, Slechta focused on the relationship of EDA (SRR amplitude), RT and selected stimuli properties (level of concreteness, emotional valence). He observed a significant but weak relationship between EDA and RT. The level of concreteness influenced only RT and not EDA, and the emotional valence affected the EDA parameters, and not RTs.

### **1.3** Pupillary activity

Pupillary activity measurement is in this project used as another method to assess physiological reactions during WAT. The pupillary data are in general measured for three reasons (Martin, 1973). First, the pupillary motility was found to be related to physiological and psychiatric abnormalities. Second, the pupil is easily accessible and the data can be measured with relatively low discomfort of the subject. Third, the changes in pupilla size are related to the mental activity and the difficulty of performed cognitive tasks.

The pupillary response consists of two components: *miotic reflex* and *mydriatic reflex*. The miotic reflex is mediated by Edinger-Westphal nucleus of the midbrain (Synek & Skorkovská, 2004) and this path enervates musculus sphincter pupillae, which shrinks a pupil as a function of light intensity. The mydriatic reflex causing the dilation of a pupil is a part of sympathetic system and it enervates musculus dilatator pupillae.

Recent studies (Steinhauer, Siegle, Condray, & Pless, 2004) suggest that dilation associated with mental processing is caused by cortical inhibition of the parasympathetic pathway at the oculomotor nucleus.

The changes in pupil dilatation are used as a measure of deception (e.g. Lubow & Fein, 1996; DePaulo et al., 2003), cognitive workload (e.g. Metalis, Rhoades, Hess, & Petrovich, 1980), a pain detector or for assessment emotional reaction (e.g. Oka, Chapman, & Jacobson, 2000). Task-evoked pupillary responses are known to reflect the processing load related to cognitive tasks (Beatty, 1982).

In past a direct observation of various types of photography were used for the measurement. Nowadays electronic infra-red scanning devices are used and the eye-tracking systems are also capable of pupilla recordings.

#### **1.4 Semantic Selection Test**

This method (further abbreviated as SST) was developed by biochemist and psychiatrist Doležal (Smékal, 1990). The original variant had some psychometric deficits, but the modern approaches developed by Urbánek (2003) and Filip (2003) offer new possibilities in evaluation.

The idea of the method is to match suitable pictures with word stimuli. The participant is presented a set of 16 schematic pictures (see Fig. 1.3 for an example) and asked to match a certain number (fixed number of 8, or range from 4 to 12) with each word stimuli. The final data have a form of a binary matrix, where rows represent word stimuli and columns the attribute pictures. The semantic proximity of two stimuli words can be measured through the overlap of their corresponding picture sets. Several methods for the analysis of the observed semantic space are available.

According to the original method, the semantic proximity between each two stimuli words was calculated as the overlap of their response vectors (i.e. scalar product). Then, two stimuli words were selected - usually pairs Love/Hate or Joy/Sadness. This pair was used as an anchor for the final two-dimensional projection. The coordinates of all other stimuli words were



Figure 1.3: Original version of the attribute pictures used for Semantic Selection Test (Urbánek, 2003)

the proximities to these anchored stimuli<sup>3</sup>. In this two-dimensional space a psychologist could interpret the positive, negative, ambivalent (high in both dimensions), or indifferent words (low in both dimensions) stimuli for a subject. This method provides an easy and straightforward access to the results, but it is limited by the use of the anchor stimuli pair. First, the traditionally selected pairs focus on the valence rating of the stimuli words, while other possible relevant semantic properties may be ignored. Second, the quality of the projection is based on the assumption that two selected words are strict opposites (i.e. their response vectors are orthogonal). This assumption is usually fulfilled, but it is not guaranteed.

The second approach to the data analysis of SST, which was proposed by Urbánek (2003, 2001), is based on correspondence analysis. The data matrix is transformed using correspondence analysis to the one- or two-dimensional

<sup>&</sup>lt;sup>3</sup>For example, x coordinate for the word "School" was the proximity between "School" and "Love", and y coordinate was equal to the proximity between "School" and "Hate".

space, where the mutual proximity of word stimuli is observed. Because we expect that the selections made by subject were based on the semantic properties of the stimuli, and this space is constructed using the mutual similarity or dissimilarity between the stimuli, we can consider this projection as a possible reconstruction of a subjective semantic space. Moreover this method allows to display the positions of the attribute pictures in the same (semantic) space. In the presented project this method was used to assess the semantic proximity of word stimuli and the related physiological reactions.

The most recent method proposed by Filip (2003) is based on application of the Rasch model in the data analysis. This approach uses a combination of hierarchical clustering and correspondence analysis, and the results have a form of a small number of one-dimensional semantic scales. In these scales the positions of word stimuli and attribute pictures are displayed (see Filip & Urbánek, 2005, for detailed description). In general, not all stimuli words and only a subset of the attribute pictures is present on each scale.

#### 1.4.1 Selected psychometric properties of SST

Because the materials on SST method are available only in Czech, we will briefly describe some basic psychometric findings regarding the method used in this project. An extensive study, Psychosémantika (Urbánek, 2003), includes the psychometric analyses and the reports on joint results of SST with the semantic differential or sociometry. Here, first the comparison between real and random will be reported. Second, we will mention the relationship between individual and grouped results and report on the test sensitivity on the distance level.

In general, the correspondence analysis should be applied only in cases, where a significant relationship exists. Because the binary data matrix in SST is too sparse to use  $\chi^2$  test, Urbánek compared the results he obtained in several studies with the results generated using Monte Carlo method. There were several categories of random generated results, which differed in the number of pictures assigned to each stimulus. It was found that the first singular values are lower in randomly generated results. On the other hand, specifically systematic responses have the singular values higher, close to 1.0. In the generated results the first and the second singular values decrease significantly with increasing number of selected pictures.

Comparison of individual and averaged results shows that there are high correlations between the distances on individual and group level, which suggests a stability and similarity of the perceived meanings, but still some differences exist. These differences can reflect the individual specifics in the understanding of meanings. Important aspect for the use of the method is the sensitivity on the level of the mutual distances. Urbánek (2003) analyzed SST in healthy adolescents and adolescents with disabilities, and found significant differences in the distances between relevant concepts (e.g. self, attractivity, integration).

#### **1.5** Mutual information

The mutual information (MI) is a concept taken from information theory, which was introduced to linguistic by Church and Hanks (1990). Their work is also directly related to WAT, because they mention word association norms. In their article they introduced MI and association ratio parameters to linguistic and demonstrated several examples of deriving semantic associations from large amount of language corpus data.

The original MI index from the information theory describes the mutual

dependence or independence of two observed phenomena. In the linguistic application the appearance of a word on certain place is considered to be such phenomenon. The probability of the appearance is dependent on the frequency of the word in corpus. When the cooccurence is to be investigated, it is necessary to set context size i.e. how far from each other the words can be to be considered being together, for example they can be separated by up to two other words. In general, the context doesn't have to by symmetrical e.g. we can search for the second word in the three words *following* the first one.

The formula for MI is shown in Figure 1.4. The result MI represents how much the cooccurrence of the investigated words differs from the chance level. The words can be more often together (MI > 0), or their occurrence in the context is about chance level  $(MI \approx 0)$ ; or they are not likely to appear in the same context (MI < 0). The expression in formula corresponds to the association ratio (in the sense of Church and Hanks, they distinguished between MI and association ratio), but nowadays it is often used as mutual information (Český národní korpus, 2005).

$$MI(x, y) = \log_2 \frac{N \cdot f(x, y)}{f(x) \cdot f(y)}$$

Figure 1.4: Mutual information (MI) formula for words x and y used in corpus linguistic; N represents the size of corpus, f(x) and f(y) numbers of occurrences of words x and y, f(x, y) the number of their cooccurrence in context.

The disadvantage of MI index is its vulnerability to the frequency of the input words, because rare words can more easily achieve high MI values. This problem is usually eliminated by the frequency threshold, when MI is not calculated for the words with very low frequencies.

## Chapter 2

### **Research** question

In this chapter the research question is presented and list of hypotheses formed.

The main idea of the project is to evaluate the possible conclusions about stimuli derived from the analysis of physiological parameters in WAT. In the previous studies with patients and healthy subjects (Šlechta, 2002b), we administered WAT while measuring RTs and EDA. In the end of the experiment we usually showed the results to the subjects and together we tried to interpret the results.

If a psychologist faces a person with a question like "Is there anything important/problematic in your life related to A?" (where A can stand for e.g. "death", "mother", "school", "road"), he/she can usually get some relevant answers. Especially with the methods like EDA measurement, which seem objective, the subjects can try hard to come with some interpretation.

In this project we try to investigate the reliability of conclusions about stimuli and their personal importance. If we find low reliability, it will be helpful to know more about other sources, which can influence the outcome. In case of high reliability, we can device some guidelines how to design a stimulus protocol and how to interpret the results.

As noted above, direct questions on selected stimuli words can be tricky and provide misleading results. It would be possible to inquire about each word in the stimuli protocol and then to compare the answers with physiological measurements. This method could be done in double-blind design, which would increase the reliability of the results, but it has several drawbacks. First, it is very time consuming for both the psychologist and subject, which can affect the quality of answers (the later answers could be less elaborate). Second, the presentation order in inquiry could influence the responses in different way compared to original WAT. For example, some stimulus could be so important and related to current personal problems to the subject, that it would evoke much larger answer, and later stimuli could be perceived differently or simply as connected to previous answers. Third, such approach requires some level of confidence between the psychologist and subject. Without this relationship it is not fair to open some issues (without the possibility of later counselling or therapy). On the other hand if we investigated this problem with people within a therapy, we would get a special sample. Therefore, while WAT could be a good method to raise new issues in individual therapy, this approach would be a problematic way how to evaluate the reliability of measurements.

An alternative approach is to add other techniques and observe whether the results are in accord, and we employed this approach in the presented project. As the baseline for the research we used the same experimental/diagnostic situation as in the previous research (Šlechta, 2002b): computer administered WAT, where RT, EDA, and verbal responses are analyzed. For the practical applications the EDA measurement is the most relevant because it is the main point differing from traditional studies, allowing to collect data on common personal computer with minimum additional costs<sup>1</sup>. On Figure 2.1 the schema of original observed parameters is presented, with newly introduced parameters below the line.



Figure 2.1: The parameters observed in the original study (above line), and new parameters investigated in this research (below the line). The parameters are divided on stimulus-related or response-related.

Given that the EDA activity is related to potential personal relevance of the stimuli, the potential errors of EDA can come from two sides. First, the EDA activity measurement can be confounded (by body movements, measurement errors). For this reason we add pupillary activity measurement as a second source of data, and we will compare them. Pupillary activity measurement will be less prone to errors, or at least it will be confounded by other effects (e.g. blinks). Also the workload could be generally lower if the subject responds with a simple, common association. To reveal these

<sup>&</sup>lt;sup>1</sup>The instrument we used (METEX M-4650CR) cost about \$200.

responses we use Mutual information index derived from Czech language corpus.

Second, the potential errors can be related to the stimuli. Given that the EDA activity is related to the personal relevance, words with similar relevance will evoke similar reactions. The original study used words of different emotional valence, which was obtained by rating before the experiment. But words of similar valence (both negative) still could evoke different reactions. Here we used SST to estimate the perceived proximities between stimuli. We also tried to estimate subjective stimuli valence based on SST ratings. To enhance the generalizability of the results we used several alternative protocol and SST versions.

### 2.1 Project subgoals

Given the major goal and the added parameters, the research question can be subdivided into three major subgoals.

First, the experiment of Slechta (2002b) is replicated under more conditions. In addition to original conditions the pupillar reactions are measured as another parameter of sympathetic reaction. The original research used a single protocol administered in the same order. To facilitate generalization of the results, two protocol versions were developed and administered in two different orders.

The second subgoal of the project is to investigate the variability of semantic measurements using SST. The variability of individual results was compared with the results aggregated across several people. The aggregated results were analyzed for the effect of presentation order and attribute picture sets. The aggregated semantic spaces were compared with other association measure (MI proximity in language corpus). SST was also used to detect subjective emotional valence of the stimuli words. This subjective valence is compared to the original experiment structure, and the effect of subjective valence on WAT behaviour is analyzed further.

The third goal is to investigate the relationship between the behavior in WAT, semantic proximity measured by SST and the linguistic properties of the responses. It is tested whether semantically close words ellicit similar reactions, and how linguistically trivial associations affect RTs, EDA and pupillary reactions.

### 2.2 Hypotheses overview

In this section we go through the project outline (see Section 2.1), and state the variables of interest in each part. The hypotheses are not stated in statistical manner, their proper testing is done in Chapterch:results.

The manipulated conditions in first part of the project and the observed parameters (dependent variables) are shown in Table 2.1.

In relation to these factors, the following hypotheses are stated:

- $H_1$ : The stimuli version will not affect RT, EDA and pupillar reactions.
- $H_2$ : The order condition will not affect RT, EDA and pupillar reactions.
- $H_3$ : Concreteness and valence will affect all observed parameters.
- $H_4$ : The relationship of RT and EDA will be weak.
- $H_5$ : The stimilus valence will have greater effect on EDA than on RT.
- $H_6$ : The neutral valence condition will elicit the smallest EDA reaction.

Table 2.1: Manipulated conditions and observed parameters

	Conditions		Dependent variables
1.	concreteness level of stimulus	1.	reaction time
2.	emotional valence of stimulus	2.	electrodermal activity
3.	stimuli order	3.	pupillary activity
4.	protocol version		
5.	SST attribute picture version	4.	semantic distance in individual
			SST
		5.	MI stimulus proximity
		6.	subjective emotional valence de-
			rived from SST rating
		7.	MI stimulus-response proximity

- $H_7$ : The stimulus concreteness will have greater effect on RT than on EDA (with concrete stimuli processed faster).
- $H_8$ : There will be a strong relationship between EDA and pupillary reaction.
- $H_9$ : The relationship between RT and pupillary reaction will be weak.
- $H_{10}$ : The stimilus valence will have greater effect on pupillary reaction than on RT.
- $H_{11}$ : The neutral valence condition will elicit the smallest pupillary reaction.

In hypotheses  $H_1$  and  $H_2$  the equivalence of the protocol versions is tested. If no difference will be observed, further analyses will be done with all data merged together.
The hypotheses  $H_3 - H_7$  revisit the questions stated in the previous study Šlechta (2002b). We expect that concreteness and valence will affect RTs and EDA, with similar patterns as shown in Table 1.2.

The pupillary responses are expected to show similar pattern as EDA reactions  $(H_8, H_9)$ , also responding more to the emotional valence changes  $(H_{10}, H_{11})$ .

In the second part, SST is used and the proximities between stimuli are estimated, both individual with SST and corpus based (mutual information scores). First the effects of presentation order and alternative attribute picture set on SST distance will be measured  $(H_{12}, H_{13})$ . The corpus based proximities are expected to be similar to individually perceived ones  $(H_{14})$ . SST will be used to estimate subjective emotional valence, and the difference between this measure and the experimental valence condition will be tested  $(H_{15})$ .

- $H_{12}$ : The presentation order will have no effect on semantic distances in SST.
- $H_{13}$ : The attribute picture set will affect the semantic distances in SST.
- $H_{14}$ : There will be a significant relationship between semantic distance and proximity derived from corpus data.
- $H_{15}$ : The subjective emotional valence will not differ from the expected valence.

In the third part, the physiological results will be compared with the findings on semantic similarity. It will be tested, whether stimuli with close distances in individual SST will ellicit similar reactions in RT, EDA and pupillary responses  $(H_{16}, H_{17}, H_{18})$ . Also the effect of simple/trivial responses

with high MI association in language will be tested. The simply response could be either a word with high frequency in language  $(H_{19}, H_{20}, H_{21})$  or a stimulus-response pair with common cooccurrence  $(H_{22}, H_{23}, H_{24})$ .

 $H_{16}$ : The stimuli similar in individual SST will ellicit similar RTs.

- $H_{17}$ : The stimuli similar in individual SST will ellicit similar EDA reactions.
- $H_{18}$ : The stimuli similar in individual SST will ellicit similar pupillary responses.
- $H_{19}$ : The responses with high corpus frequency will ellicit high RTs.
- $H_{20}$ : The responses with high corpus frequency will ellicit low EDA reactions.
- $H_{21}$ : The responses with high corpus frequency will ellicit high pupillary responses.
- $H_{22}$ : The stimulus-response pairs with high MI will ellicit high RTs.
- $H_{23}$ : The stimulus-response pairs with high MI will ellicit low EDA reactions.
- $H_{24}$ : The stimulus-response pairs with high MI will ellicit high pupillary responses.

The hypotheses with the highest relevance to the main goal (i.e. significance of single EDA peak in WAT) are hypotheses  $H_5$ ,  $H_7$ , and  $H_{16-18}$ , inspecting the main relationships described by Šlechta and testing the effect of semantic similarity on the results. The next chapter introduces the methods used to collect experimental data, the results are shown later in Chapter 4.

# Chapter 3

# Methods

This chapter overviews the technical aspects of the methods used to collect the data. First, the structure of the experimental session is described, and the details on each method are provided. Second, the experimental sample described.

# **3.1** Session structure

The experimental session lasted about 60 minutes (see Table 3.1 for overview). The participants were tested individually to ensure the privacy of their responses and the controllability of conditions for instrumental measurements.

At the start of the session the participants were provided with the details about goals of the project and the time track of the session, and signed an informed consent. Then the electrodes were attached to the participant's palm and left for at least 5 minutes before the experiment started to ensure a sufficient absorption of the electrodermal gel. Meanwhile the eye tracking device was prepared and calibrated. Then the participant was instructed about WAT and the test proceeded. After the test the electrodes and the eye tracker headband were removed and reproduction test for WAT was performed. The participant was not informed about the reproduction test to ensure he/she won't be trying to remember the responses and will follow only the instructions for WAT. Then he/she was instructed about SST and proceeded with the test. After the session the administrator talked with the participant, answered questions and showed example results.

Each participant was randomly assigned to 1 of 8 experimental conditions (2 word stimuli sets  $\times$  2 stimuli orders  $\times$  2 SST picture sets). Thus, 9 to 11 participants were tested in each condition.

Time duration	Method
$5 \min$	Introduction, informed consent
$5 \min$	Preparation for EDA and eye tracking
$12 \min$	Word association test (WAT)
$5 \min$	WAT reproduction test
20 - 35 min	Semantic selection test
$5 \min$	Ending, debriefing, last questions

Table 3.1: The structure of the experimental session, with estimated time durations in minutes.

# **3.2** Word Association Test

### 3.2.1 Protocol

The test protocol contained 40 common Czech nouns as stimuli. There were also 4 extra training words, which were not included in the analysis. The protocol was compiled to provide levels of 2 factors: emotional valence and concreteness (see Tab. 3.2). The words differed in their emotional valence (10 positive, 10 negative, 20 neutral) and in their level of concreteness (10 concrete neutral, 10 abstract neutral words). The protocol structure followed the design from previous studies (Šlechta, 2002b).

Table 3.2: Structure of the stimulus protocol with respect to 2 manipulated factors: emotional valence and concreteness

	Positive Neutral		Negative
	stimuli	stimuli	stimuli
Medium concreteness	10 nouns	10 nouns	10 nouns
High concreteness		10 nouns	

There were 4 versions of the stimuli protocol (2 stimuli sets  $\times$  2 stimuli order). One of this versions closely matched the content of Šlechta's protocol (2002b) and contained 36 of 40 original words (90%). This version was used for comparison with previous research using identical design and similar methods. Still, some changes were made. First, the word "friend (přítel)" was replaced, because it is ambiguous. In Czech "přítel" means both friend and boyfriend (but not girlfriend), so we can expect it is perceived differently by men and women. Second, all four versions of the protocols contained 6 emotion names<sup>1</sup> to be used as reference points in SST. Two of these emotion names were already in the original protocol, one word was removed for ambiguity and three additional words were replaced to accomodate all six emotion names.

The altervative version followed the same structure and contained words with comparable frequency in Czech language corpus. Both versions overlapped in 7 words: 6 emotion names and word "love (láska)".

Two order conditions were used in presentation. In the alternative order condition first the stimuli 21–40 were presented and then the stimuli 1–20.

#### 3.2.2 Stimuli frequencies

The stimuli sets were designed with respect to previous version tested in patients. The previous studies used slightly different approach in the measurement of the frequencies in corpus.

To justify the balance of stimuli frequencies in the protocol across the stimuli groups, a new analysis was performed, which was based on the logarithm of the frequency in the most recent and largest corpus of Czech written language SYN2005 (Český národní korpus, 2006).

The frequency data were transformed to the number of occurences in million words and then logaritmized. The frequency distributions for all stimuli groups and protocol versions are shown in Figure 3.1. The stimuli groups introduced earlier (see Table 3.2) were compared as follows: according to cognitive workload condition (neutral concrete/abstract), and according to emotional workload condition (abstract positive/neutral/negative). The

<sup>&</sup>lt;sup>1</sup>Basic emotions in Eckman sense were used: joy, sadness, surprise, disgust, fear, anger



#### Differences in the corpus frequencies of stimuli

Figure 3.1: Differences in the corpus frequencies of stimuli, tested with Wilcoxon rank sum test. The significant differences are marked with \* (p < 0.05) or \*\* (p < 0.01).

differences were tested with Wilcoxon rank sum test.

The abstract and concrete neutral stimuli words did not differ significantly in either stimuli set. There were found significant differences in the emotional workload condition. The positive stimuli had lower frequencies (sets 1 and 2), and negative stimuli had also lower frequencies (set 2) than neutral abstract stimuli. It is important to emphasize that the abstract neutral stimuli group have either higher or equal frequency to the other groups, and therefore possible longer RTs are not caused by rarity of the words compared to the other groups, but rather by their level of abstractness.

The stimuli sets were used in the experiment because they both passed the previously used criteria. This updated comparison will be discussed later in the context of physiological differences (in Section 4.2.4).

The SYN2005 corpus was used as the best and largest source of linguistic data. Still there are some limitations of this method. First, SYN2005 is a corpus of written language, the actual frequency in spoken language can differ, but the size of current oral corpus is much smaller<sup>2</sup>. Second, the actual familiarity of participants with stimuli words can vary with their education, specialization, hobbies or other factors, but the corpus frequency is the best available estimate of common familiarity of words.

## 3.2.3 Administration

The stimuli were presented in fixed interval manner on computer monitor and the participant was instructed to "respond with the first word that comes to his/her mind". The participants were informed that there were no correct or wrong answers and it was important to be authentic, and that only the

<sup>&</sup>lt;sup>2</sup>According to the Bonito software (Český národní korpus, 2006) oral corpus ORAL2006 contains 1.3 million of positions compared to 122.4 million of positions in SYN2005.

first response was being recorded. There were two versions of test protocol containing different stimulus words, each in two minor versions differing in the stimuli order.

The stimuli were presented in fixed 16 second intervals. Each stimulus was presented for 300 ms, with preceding fixation cross for 700 ms. After the stimulus presentation the computer screen was blank for 15 seconds allowing the participants to answer (see Fig. 3.2). The limited presentation time was used to engage subjects' attention and to ensure there is a limited period when the stimuli are presented and from which we measure the reaction times. If we presented stimuli for e.g. 2 seconds, the measurement of effective stimulus onset would be more problematic. The 300 ms presentation time with preceding fixation cross allowed comfortable reading.

The responses were recorded with audio recorder and later the reaction time was measured from these recordings.



Figure 3.2: The stimulus timing during the computer administration of association experiment.

# **3.3** Electrodermal activity

EDA was measured on the palm site of non-dominant hand with METEX M-4650CR at 2 Hz. For the experiment the Ag/AgCl electrodes ( $0.28 \text{ cm}^2$ ) were used. They were filled with isotonic NaCl gel and attached to the skin for at least 5 minutes before the experiment started to allow the hydration of epidermis.

The time of the first reaction peak and the peak height (SRR amp.) were measured and several methods for data extraction were used. These methods differed in the size of time window for peak detection and in standardization of the data. Either the peaks were detected in raw data and later z-transformation was employed (e.g. Lukavský, 2007), or Fast Fourier Transformation was used for filtering out low frequency waves before the peak detection.

It was formely observed that the continuous EDA signal does not have stable EDL level. The decrease during experiment session is attributed to habituation and to the fact that participants become familiar with the experiment task<sup>3</sup>. The Fast Fourier Transformation (FFT) was used in the previous study (Šlechta, 2002b) to reduce this effect.

With FFT it is possible to decompose the EDA signal and obtain two signal components — high frequency signal corresponding to fast changes and low frequency signal corresponding to slow changes. In this analysis the EDA signal was first scanned for occasional missing data, which were replaced by the mean of surrounding values. Second, the data of each participant was transformed with FFT. All frequencies with periods longer than 32 seconds (i.e. 2 stimulus cycles) were suppressed and the remaining high frequency

<sup>&</sup>lt;sup>3</sup>It was also observed that this habituation is more frequent in people with low Lability scores in Eysenck Personality Inventory (Lukavský, 2003)

signal reconstructed with inverse  $FFT^4$ . This process is illustrated with an example on Fig. 3.3.



Figure 3.3: Example of FFT decomposition of EDA signal on low and high frequency components (32 s threshold).

Some authors (e.g. Bernat, Patrick, Benning, & Tellegen, 2006) detect the SRR peaks in a limited time window, for example 0 - 4 s after stimulus onset. Šlechta (2002b) used whole interstimulus interval (14 seconds) for peak detection. The original study (p.34) showed that in some cases average peak times were close to 7 s, and therefore a longer time window should be preffered. Both methods were considered (short 7 s window vs 14 s window) and similar results were obtained. The correlations of parameters obtained with

<sup>&</sup>lt;sup>4</sup>All calculations were performed with R statistical package (R Development Core Team, 2007) using fft function for Fast Fourier Transformation.

these modified methods are shown in Table 3.3. For further analysis the FFT transformed signal with 14 s time window was used. The z-transformation on case level was later used in the analysis to eliminate the variance between cases.

EDA Peak Method EDA Peak Time 1. 2. 3. 4. 1. 2.3. 4. 1. raw data, 7 s 0.908 0.928 0.8860.5410.609 0.068 2. raw data, 14 s 0.7920.8980.2540.4773. FFT, 7 s 0.9240.3094. FFT, 14 s

Table 3.3: Comparison of four methods for peak detection in EDA signal.

Note. All Pearson's correlation coefficients are statistically significant (p < 0.001)

# 3.4 Pupillary activity

PA was measured using Eyelink II system at 250 Hz. For the analysis only data samples from a limited interval (0 - 5000 ms) were used. The first part of this time interval (0 - 1000 ms), when the fixation cross and the stimulus are being displayed, was used to estimate the base level. The maximum dilation from the other part of the time interval (1000 – 5000 ms) was compared with the corresponding base level and relative dilation was calculated (see 3.4).

The trials with blinks were included in the analysis, but the corresponding data were adjusted. The blinks were treated as missing data and the data



Figure 3.4: Method of measuring pupillary activity. Dotted lines show the stimulus presentation interval, solid lines show the base level estimated by mean, and maximum dilation in 1000 – 5000 ms interval.

segments near blinks (closer than 100 ms) were also removed, because they included steep changes in pupil size. Only the first 5000 ms of pupillary signal were analyzed, because the pupillary reactions tend to be faster than EDA changes and the shorter interval decreases the chance of possible measurement artifacts after the response had been reported. The pupillary responses were standaridized using z-transformation on the case level. Several methods how to assess the base level where used. For base level estimate both mean and median were used, but they led to similar results (r = 0.999, p < 0.001).

In further analyses only the base level calculated from mean was used. The base levels were very stable during the experiment. There was no observable time trend (Pearson correlation with trial order r = -0.024, p = 0.15), and the majority of variance in base levels could be explained solely by interindividual variance (*adjusted*  $R^2 = 0.9676$ ). Due to this stability in time no further data adaptation (e.g. FFT) was used. It is reported that the pupillary measures are confounded by the gaze position (Pomplun & Sunkara, 2003). The pupillary data were not adjusted for pupil position, because due to the absence of drift corrections during the data collection the pupil position could not be reliably measured. The drift correction procedures would interrupt the WAT and therefore they were omitted.

# **3.5** Semantic Selection Test

After WAT the participants were asked to complete SST test to assess the stimulus words. They were instructed "to select for each word 4-12 pictures based on its meaning for you and the feelings it evokes". After the test they indicated 4 personally pleasant and 4 unpleasant pictures. The personal valence rating was calculated for each word as the number of personally pleasant pictures selected minus the number of unpleasant pictures selected (i.e. ranged from -4 to +4).

There were two sets of attribute pictures used in the experiment (see Fig. 3.5). One was the original set (found for example in Urbánek, 2003), the second set was created from photographs available on internet.

The individual results were processed using correspondence analysis (for details see Section 1.4). An example of individual results is shown on picture 3.6. When two SST results were being compared, the Peason correlation of



Figure 3.5: Attribute pictures used for Semantic Selection Test, original version on left, photograph version on right.

the distances was performed.

For several analyses the SST results were aggregated (if obtained for identical stimuli and with identical attribute pictures). In the aggregation the individual SST data were summarized and the correspondence analysis was performed with this aggregated matrix.

# 3.6 Calculation of MI

The MI index was based on lemma frequencies in Czech corpus of written language (Český národní korpus, 2005). For the psychological research the oral corpus would be more convenient, but currently it lacks the size of the former one, and the corpus size is for MI calculations very important.

The size of context was set to  $\pm 3$ . The word "hectar" with significantly lower frequency was removed from the analysis. One technical problem had to be solved regarding the occasional no occurences, when according to the



Figure 3.6: Example of individual SST results transformed with correspondence analysis.

formula  $MI = -\infty$ . Two alternatives were tested with similar results: these cases where either considered missing values, or the values were replaced with a low value  $(MI_{low} = mean(MI) - 2 \times SD(MI))$ .

# 3.7 Sample selection

The sample consisted of 80 healthy volunteers, both men and women, age ranging from 18-35 years. They were recruited via advertisement and obtained 200 Kč (\$8 equivalent) for 1 hour experiment session. More detailed statistics on sample are presented in Section 4.1.1.

# Chapter 4

# Results

# 4.1 General results

### 4.1.1 Sample

The sample consisted of 80 healthy young volunteers, 30 men and 50 women. The age ranged from 18 to 34 years (mean = 22.9, SD = 3.6). The education level in the sample was higher than in general population (19 participants (24%) were high school students or graduates, 61 (76%) were university students or graduates).

Because of the physiological recordings, the laterality and sight conditions were registered. The majority of the participants were right-handed (75; 94%) and only 5 participants (6%) left-handed. The majority of participants didn't use any eye correction (50; 63%), 12 participants (15%) wore contact lenses, 11 (14%) wore glasses, and 7 (9%) usually wore glasses, but didn't wear them during the experiment session.

#### 4.1.2 Difference between stimuli protocols

Before we focus on the main hypotheses, the differences between two protocol versions must be examined. Because of the individual differences in scale, the data are z-transformed. The goal of this comparison is not to compare individual performances, but the reactions evoked by stimuli subgroups. The RTs, SRRamp values and pupillary activity peak values were compared in both versions for whole protocols and separatelly for specific stimuli groups within protocols. The results are shown in Table 4.1.

~	6	RT		EDA		PR	
Stimulus category	Count	t	p	t	p	t	p
all words	40	0.000	1.000	0.000	1.000	0.000	1.000
concrete neutral	10	-1.187	0.236	-1.346	0.179	0.897	0.370
abstract neutral	10	-0.918	0.359	-1.162	0.246	-0.239	0.811
abstract positive	10	-0.318	0.750	0.792	0.429	-0.878	0.380
abstract negative	10	2.108	$0.035^{*}$	1.653	0.099	0.192	0.847

Table 4.1: Difference between protocol versions

\*p < 0.05

There is only one significant difference in the observed parameters across the protocol versions. This difference can be attributed to the multiple testing – after Bonferroni correction it is no longer significant. The results demonstrate that there are no significant differences between stimuli subgroups in both protocol versions. Potential differences would be caused e.g. if one version contained neutral words, which evoke either high emotional or cognitive workload.

## 4.1.3 Effect of stimuli order

Both versions were administered under two order conditions. The dependent variables are not expected to differ. The comparison of both order conditions for both protocol versions is shown in Table 4.2.

		RT		EDA		PR	
Stimulus category	Count	t	p	t	p	t	p
Version A							
all words	40	0.000	1.000	0.000	1.000	0.000	1.000
concrete neutral	10	0.679	0.498	-1.000	0.318	-0.851	0.395
abstract neutral	10	-0.299	0.765	-0.091	0.928	1.465	0.144
abstract positive	10	-0.745	0.457	0.275	0.783	-1.188	0.236
abstract negative	10	0.505	0.614	0.827	0.409	0.548	0.584
Version B							
all words	40	0.000	1.000	0.000	1.000	0.000	1.000
concrete neutral	10	-0.194	0.846	-0.024	0.981	0.939	0.348
abstract neutral	10	0.183	0.855	1.342	0.180	-1.465	0.144
abstract positive	10	-0.442	0.659	-1.576	0.116	0.038	0.970
abstract negative	10	0.386	0.699	0.350	0.726	0.467	0.641

Table 4.2: Effect of stimuli order for both protocol versions

The results demonstrate that the order condition does not affect the observed parameters, which was expected. In further text the data form both protocol versions and conditions will be analyzed together, and only the stimuli properties (emotional valence, level of abstraction) will be taken into account.

# 4.2 Physiological results

The main goal of this part of the analysis is to investigate possible causes related to the changes in physiological parameters. This knowledge would facilitate the interpretation of the responses in WAT.

To test this issue stimuli words differing in level of concreteness and emotional valence were included in WAT protocols (see Section 3.2.1). It is hypothetized that abstract words will evoke higher cognitive workload, while positive or negative words will evoke higher emotional workload compared to neutral stimuli.

The observed parameters were reaction time (RT), electrodermal activity (EDA) and pupillary responses (PR). According to the original hypotheses, cognitive workload will manifest in longer RTs, and emotional workload in higher EDA. First the effect of emotional valence is tested, then the effect of cognitive workload is investigated. Finally, the effects are compared, and the efficiency of decisions based on physiological data is evaluated.

#### 4.2.1 Correlations of physiological parameters

First we compared the outcomes from each method. We found low, but significant correlations between all three observed parameters (see Table 4.3). The relationships are significant due to the large total number of samples (N = 3048; 4.8% of samples were removed due to technical problems), but the ratio of explained variance by these correlations is low. It is probably caused by different sensitivity of each method or different vulnerability to possible measurement artifacts (e.g. body movements, blinks).

Table 4.3: Intercorrelations for observed physiological parameters.

Parameter	RT	EDA	PR
Reaction time (RT)	_	0.23	0.17
Electrodermal activity (EDA)		—	0.11
Pupillary response (PR)			_

All Pearson correlation coefficients are statistically significant (p < 0.001).

#### 4.2.2 Effect of emotional workload

In order to assess the effect of emotional workload, the responses to 30 stimuli were analyzed within each participant. The analyzed stimuli included 10 positive, 10 negative and 10 neutral words, all on similarly high level of abstractness (i.e. not representing real-world objects). The differences between positive and neutral, and negative and neutral stimuli were tested with t-tests.

The results are shown in Figure 4.1. In RT only the significant difference between positive and neutral stimuli was found (p < 0.001). The RTs in positive stimuli were very short (near mean RT), and both the negative and neutral (abstract) stimuli required longer reaction times.

The changes in EDA show a different pattern. Both positive and negative stimuli evoke significantly higher EDA reactions than neutral abstract stimuli



Figure 4.1: Effects of emotional workload on reaction time (RT), electrodermal activity peak values (EDA) and pupillary responses (PR). The significant differences (tested with t-test) are marked with \*\* (p < 0.01) or \*\*\* (p < 0.001).

(p < 0.001). The reactions to negative stimuli are slightly higher compared to the positive ones.

The changes in PRs are similar to the trends in RTs, but the effects are smaller. The neutral abstract stimuli evoke the highest responses, and there is a significant difference between responses to neutral and positive stimuli (p < 0.01). The responses to negative stimuli are close to mean values.

The most significant results were found in EDA changes, which was expected. In reaction times the longest times are connected with abstract neutral stimuli, which was also expected, but it is important to note that negative stimuli evoked also comparably long RTs (the difference was not significant). The pupillary activity showed a similar pattern to RTs, with smaller differences.

## 4.2.3 Effect of cognitive workload

The effect of cognitive workload was tested by comparing responses to 10 abstract and 10 concrete neutral stimuli within each subject. The differences were tested with t-tests.



Figure 4.2: Effects of cognitive workload on reaction time (RT), electrodermal activity peak values (EDA) and pupillary responses (PR). The significant differences (tested with t-test) are marked with \*\* (p < 0.01) or \*\*\* (p < 0.001).

The observed results are shown in Figure 4.2. The abstract neutral stimuli evoke higher/longer reactions in all observed parameters (p < 0.001). The

differences are larger than in the previous condition (see Section 4.2.2). Only in EDA the reactions to abstract stimuli are around the mean values (the positive and negative stimuli evoked the above average EDA responses), in RTs and PRs the reactions exceed the mean.

The difference in RTs was expected and it matches the previous experience. On the other hand, the difference in EDA shows that these reactions are also influenced by cognitive workload in WAT. In the following section both effects will be compared.

### 4.2.4 Comparison of both effects

It was shown that in WAT both emotional and cognitive workload conditions affect the results. The results for both conditions are shown together in Figure 4.3. To compare the importance of these effects, the effect size measure (Cohen, 1992) will be used. For effect size there were proposed several bounds, which help to interpret the size of the observed difference. The effect size 0.2 indicates a small effect, 0.5 a medium effect and 0.8 a large effect size.

When we apply this measure, the highest effect size (d = 0.51) is found in RTs for abstract/concrete neutral stimuli. Two other effects can be considered as small effect sizes: PRs for abstract/concrete neutral stimuli (d = 0.21)and EDA changes for negative stimuli (d = 0.24).

This shows that changes observed in WAT are mostly caused by cognitive workload condition. However, the interpretation of behavioral and physiological parameters in WAT usually stressed the emotional workload process.

The longer RTs in abstract neutral stimuli can not be explained by the frequency in language, because their frequency was either equal or higher than in other groups (see Section 3.2.2). It is possible that people have



Figure 4.3: Effects of cognitive and emotional workload on reaction time (RT), electrodermal activity peak values (EDA) and pupillary responses (PR). The significant differences (tested with t-test) are marked with \*\* (p < 0.01) or \*\*\* (p < 0.001).

better representation of the emotional words, since these words are or can be related to themselves, and thus they respond faster, despite the similar frequency in language. This interpretation would suggest positive words to be connected with faster RTs than negative words, because they are more likely to be related to participants' self-concept and therefore more easily accessible. To further test the cognitive/emotional workload differences it would be good to include also concrete positive and negative stimuli (e.g. rose or gun) and investigate the full  $2 \times 3$  design.

Still the emotional valence effect in EDA was confirmed. For the application in clinical assessment it is important to investigate also the reversed perspective: not only there are significant differences in EDA based on emotional valence, but how precise is the stimuli sorting based on these responses?

# 4.2.5 Discrimination based on physiological parameters

From the clinical point of view the possibility to discriminate especially between the negative and neutral stimuli is important. The increased activity associated with a generally neutral word could indicate a potentially problematic topic (i.e. the word is individually perceived as negative), and the further assessment or therapy could be focused in this way. The largest differences between negative and neutral stimuli were observed in EDA reactions, and therefore here we will focus on this parameter.

The Figure 4.4 shows a receiver operating characteristic (ROC) for the discrimination of abstract neutral and negative stimuli based on EDA data. We can see that EDA data help to classify the stimuli, but the previously reported small effect size (d = 0.24) is too small for a successful automatic discrimination.

When physiological parameters are collected in WAT, the classification of the reactions cannot be automatic, because the data variability within stimuli groups is too large compared to the differences between both stimuli groups.

# 4.3 Semantic and linguistic results

## 4.3.1 Individual differences in SST

First the differences between individuals semantic maps obtained from SST and the maps obtained from the aggregated data were compared. The distances in two-dimensional correspondence analysis space were used as a measure of deviation from an aggregated map, and each individual map was

ROC curve of negative/neutral classification based on EDA



Figure 4.4: ROC curve of negative/neutral classification based on EDA. The dotted line shows the level of random guess.

compared with corresponding aggregated map for the same experimental conditions (9–11 participants for each condition). When the distances for the whole protocol were compared, the mean distance correlation coefficient was 0.42 (median = 0.39, SD = 0.15). Because the correlation coefficient is probably affected by the large number of interrelated values, the maps were also compared using a limited part of the protocol containing 6 basic emotion words. The coefficients observed on this protocol subset were higher (mean = 0.70, median = 0.79, SD = 0.25), see Figure 4.5 for coefficient distribution details.

The observed results are similar to the results obtained by Urbánek (2003). In his study the correlation coefficients computed with same methods ranged from 0.37 to 0.84 (mean = 0.63, SD = 0.13; 15 participants,



Figure 4.5: Individual vs aggregated map distance correlations, shown separatelly for complete 40 word protocol and for 6 word subset.

20 words). The values are lower than the results obtained for 6 words protocol and higher that the values for the whole protocol, which suggests that the correlation coefficients as the measures of map similarity are dependent on the number of words in compared maps/protocols.

## 4.3.2 Effects of order and attribute picture sets

The effects of order and attribute picture sets were analyzed using aggregated SST results for each condition. The two alternative stimuli versions were used as independent observations of these effects. The aggregated SST maps obtained with correspondence analysis were compared using Pearson correlation coefficient for corresponding distances. The results are shown in Figure 4.6. The two correlation coefficients in the middle row show the distances under the effect of attribute picture sets, the four correlation coefficients in the bottom row show the distances of versions differing in presentation order.

In all cases the correlations are statistically significant (p < 0.001).



Figure 4.6: Pearson correlation coefficients for distances in aggregated SST maps.

It was expected that the presentation order will have only small effect on the SST map, but the attribute picture sets will have more powerful effect resulting in lower distance correlations. The results show that both effects are comparable. Because the correlation coefficients can be biased by the large number of distances (780), the analysis was repeated with a smaller number of stimuli words (6 basic emotions, i.e. 15 distances). The second analysis showed the same results, correlations under both effects were comparable, but higher (r = 0.82-0.98).

#### 4.3.3 Proximities measured with SST and MI

The distances in aggregated SST maps were compared with MI for corresponding stimuli words. The comparison was made separatelly for each of 4 experimental conditions (2 stimuli sets  $\times$  2 attribute picture sets; 2 order conditions were merged together).

The distance correlation coefficients<sup>1</sup> are low (from -0.22 to -0.11; p < 0.001). The results are statistically significant, but the explained variance ratio is very low. The alternative approach (substituting MI with mean±2SD for word pairs with no common occurrence in corpus) led to similar results (r = -0.28 to -0.18; p < 0.001).

The results show that even both metrics (SST and MI) measure the similarity of meaning to some extend, they are not directly interrelated.

### 4.3.4 Subjective valence rating

The subjective valence rating was measured indirectly. It was derived from the SST responses and the post-test assignment of 4 positive and 4 negative pictures (see Section 3.5). For each word the subjective valence rating was calculated as the number of personally pleasant pictures selected minus the number of unpleasant pictures selected (i.e. ranged from -4 to +4).

The subjective rating was slightly shifted toward the positive values (mean = 0.40, median = 1.0, SD = 2.31). This rating was compared with the expected valence (see Figure 4.7).

Because the subjective valence rating uses a different scale (-4 to +4) than the expected valence (-1 to +1), the subjective valence rating was also collapsed into similar scale (positive valence with ratings +4 to +2, neutral

<sup>&</sup>lt;sup>1</sup>The correlations are negative, because the higher values represent greater distance in SST, but greater proximity in MI.



Figure 4.7: Subjective valence and expected valence fit (left), and aggregated to three groups (right). Kendall correlation coefficient  $\tau = 0.62$ ,  $\tau = 0.64$ respectively.

with +1 to -1 and negative valence with -2 to -4), and compared again.

The fits were similar for both scales, Kendall correlation coefficient  $\tau = 0.62$  or  $\tau = 0.64$  respectively. It means that two words with the same subjective valence rating (across individuals) are likely (with probability  $p = \tau = 0.62$ ) to be in the same expected valence group.

The discordancies in fit were caused by shifts within one valence category (i.e. positive - neutral, or neutral - negative), which constituted 32.3% of the cases. The more substantial shifts (positive - negative) were very rare (0.9%).

The mentioned discordancies can be caused by several factors. First, some can be attributed to the error of the employed method, there can be some difference between assigning pleasant or unpleasant pictures, and subjective valence. Some error also can be introduced by the aggregation down to 3group scale. Second, the differences in subjective valence were reported in previous studies (Lukavský, 2004). It was found that people are sensitive to negative words and their subjective valence reflects the original valence condition better, but they mix the neutral and positive words more easily, perceiving them more like a continuum. This point is supported by the observed shift in subjective valence toward positive values. Third, there can be of course real individual differences in valence perception, and therefore the total fit can never be observed.

To assess the importance of the last point, the results from Section 4.2.2 were replicated using subjective valence instead of original expected valence. We focused only on EDA part, because the most significant differences related to the valence were observed there.

The replication brought mixed results. In the strict replication, using three valence groups, the effects sizes decreased, especially concerning the positive valence group. The difference between negative and neutral abstract words decreased from d = 0.24 to d = 0.19, and the difference between neutral and positive words fell from d = 0.16 to d = 0.01. This suggests that the possible advantage of individualized valence ratings doesn't counterbalance the changes introduced by the method employed and the nature of valence perception. In case of negative/neutral difference the values are similar, both showing a small effect size. Therefore the fall in positive/neutral differences can not be simply attributed to the method, but probably to the mentioned tendency to perceive the neutral and positive words as a gradient also attributes to this difference.

In the subjective valence rating it was possible to replicate previous observations using more fine scale. The relationship between subjective valence and EDA peak values can be found in Figure 4.8. When the finer 9-point scale (-4 to +4) is used, the gradients are more evident. In the negative part of the scale the relationship is steeper, while in positive values the trend is slower and the differences are hardly significant.

Subjective valence and EDA peak values

Figure 4.8: Subjective valence rating and EDA peak values. Whiskers indicate standard errors.

If the subjective valence is measured using SST (or if it is obtained directly by question (Lukavský, 2004)), the relationship between the valence and physiological parameters are more reliably observed in the negative part of the scale. The positive part is biased with the tendency to mix neutral and positive words, which affects the results.

This tendency is probably not a measurement artifact, but it reflects a general assessment bias. Many neutral words ("house", "tree", "road", etc.) can be easily perceived as positive, which makes the neutral/positive continuum fuzzy, while we detect the negative words precisely and we avoid rating neutral words as negative.

#### 4.3.5 WAT responses

The lexical categories of WAT responses were investigated (see Figure 4.9). For the linguistic analyses, 2 participants were excluded (native Slovaks). The vast majority of responses were nouns (84.3%) and also a small number of adjectives (6.8%) was used. In some cases (3.2%) the participants responded with several words at once, which couldn't be easily classified.



Lexical categories of WAT responses

Figure 4.9: Lexical categories of WAT responses

The association failures (no response in WAT within 14 seconds) were observed in 2.5% of reactions. There were on average 0.97 association failures per participant (SD = 1.70, with maximum 10 failures).

The frequency distribution of responses was compared with the frequencies of stimuli in Czech national corpus to find out how rare or common words were used in responses compared to the WAT stimuli (see Figure 4.10). After removing responses of 2 native Slovak speakers, association failures and combined multiword responses, 91.2% of the responses were included in the analysis. The frequencies were transformed to the number of occurences in million words and then logaritmized<sup>2</sup>. The frequency of responses is slightly lower, which probably corresponds with a large number of possible words, which could be said in response. However, this effect is not strong and in a non-parametric test is not found significant (t-test, p < 0.05; Wilcoxon rank sum test, p = 0.34).

#### 4.3.6 Stimulus – response relationship in language

We investigated the bond between stimuli and their responses in language using MI. The stimuli-response pairs with no cooccurence in corpus (263 cases) were excluded from the analysis (N = 2687). The stimuli-response MI (SRMI) was apparently above zero (mean = 5.855, median = 5.750, SD = 2.666, see Figure 4.11), which demonstrates a predictable assumption that people pick the responses, which are also often found together in written language (one sample t-test, p < 0.001).

This measure will be further used to assess the cognitive requirements

<sup>&</sup>lt;sup>2</sup>This transformation helps to adapt the statistical distribution and it also facilitates the reading of the charts – e.g. number 2 in the chart means  $100 = 10^2$  occurences in million words.

Corpus frequency of stimuli



Figure 4.10: Frequency distribution of stimuli and responses in Czech national corpus.

to generate a response and the related behavioral parameters (see Section 4.4.1).

# 4.4 Combined results

## 4.4.1 Response frequency and its behavioral correlates

As shown in Section 4.3.5, the corpus frequency of responses was similar to the frequency of stimuli words. Here, it is to be tested whether very common words are connected with low cognitive workload, because these words are
#### Histogram of WAT stimulus-response mutual information index



Figure 4.11: Histogram of WAT stimulus-response mutual information index.

easily accessible to cognition. The common words could be also connected with low emotional workload, because during the WAT task they come to subject's mind more easily and he/she doesn't need to search his mind for more personally related (and maybe more emotional) response.

This relationship was tested with Pearson correlation between the corpus log frequency and each observed physiological parameter (see Table 4.4). It was found that responding with a frequent word is associated with lower RT, lower EDA reaction, and pupillary response (p < 0.01), but the common variance ratio is very low. It means that knowing the response frequency in corpus, doesn't allow us to efficiently substract this effect from the reaction data to obtain more clean results.

Table 4.4: Correlations of response frequency and behavioral measures

	Response frequency	Stimulus-Response MI		
Reaction time	-0.084 ***	-0.113 ***		
Electrodermal activity	-0.103 ***	-0.032 n.s.		
Pupillary response	-0.053 **	0.006 n.s.		

Note. Pearson correlation coefficients shown. \*\*p < 0.01, \*\*\*p < 0.001

Similarly, it was tested whether the stimulus-response pairs often found together in corpus are associated with lower cognitive or emotional workload (see Table 4.4). This effect was found only in RTs (p < 0.001) suggesting that responding with frequent stimulus-response pair is faster (lower cognitive workload), but doesn't affect physiological reactions). Again, this effect is significant, but very low, so it isn't efficient to substract it from reaction data.

#### 4.4.2 Subjective valence and physiological reactions

One of the main questions of this research is whether the observed reactions can be explained by individual differences in understanding stimuli words. The main condition we used to evoke emotionally related reactions was including stimuli of different valence levels. In Section 4.3.4 we already replicated the emotional workload comparison using semantic valence obtained from SST. This approach didn't prove to be much useful: the observed effects were similar (negative stimuli) or lower (positive stimuli). The error introduced by this method and rounding down to three valence groups could confound the results. The differences in positive–neutral part of the scale can be also explained by different approach in valence assessment with these stimuli.

# 4.4.3 Subjective semantic proximity and physiological reactions

The previous approach was based on the subjective valence derived from SST responses. With SST results it is not necessary to rely on the valence data, but it is possible to inspect the effect of perceived similarity. There is a problem how to test statistically whether subjectively similar stimuli evoke similar reactions. Even in 1-dimensional scale like valence there was no simple linear relationship, but rather U-shaped curve (see Figure 4.8). In 2-dimensional SST projection the testing will be even more complicated: the peaks can be surrounded by lower peaks and separated from other peaks by areas with no points or by points with lower values.

Example of the SST results for one respondent are shown in Figure 4.12. The coordinates are obtained with correspondence analysis (see Section 1.4 for details), the peak heights represent EDA values. The problem is how to measure the impact of similarities on reaction in such space.

The idea is to study overlap between different stimuli subgroups. To test for example whether the stimuli similar to the ones with high EDA reactions also evoke high reactions, we can first pick a subgroup (10 stimuli with highest EDA reactions). Then we select the stimuli which are perceived similarly in SST (10 other with smallest distances to the first group). Finally we can compare the EDA reaction between second group and the unselected stimuli (20 stimuli). If the hypothesis is true, there should be a difference between



Figure 4.12: Example SST projection for one respondent (ID03). Peaks represent EDA reactions. Positive stimuli in red, negative in gray, neutral in blue.

these two subgroups. Analogously, we can test the differences in RTs.

Some subgroups are already present in the protocol by purpose (see Figure 4.13). There are positive and negative stimuli, there are also 6 basic emotions, which were included as benchmarks for the SST projection. We tested the overlap between 20 stimuli with emotional valence and the 20 stimuli most similar to 6 basic emotions (included), and found a significant association  $(\chi^2 = 527.8, df = 1, p < 0.001, Kendall \tau = 0.41).$ 

We picked 5 subgroups: 10 positive stimuli, 10 negative stimuli, 10 similar to basic emotions, 10 with highest RTs and 10 with highest EDA reactions for each respondent. For each of these "core" subgroups, a "similar" subgroup of 10 stimuli with smallest SST distances was selected. The unselected 20 stimuli formed a "rest" subgroup and the parameters (RT, EDA) of "core" and "similar" groups were compared with "rest".

Subgroup	Core group		Similar group			Rest group		
	mean	SD	р	mean	SD	р	mean	SD
10 Positive	0.003	0.959	n.s.	-0.029	0.951	n.s.	0.013	1.017
10 Negative	0.121	1.084	***	-0.039	0.955	n.s.	-0.040	0.952
10 Basic emotions	0.100	1.044	**	-0.012	1.005	n.s.	-0.045	0.947
10  Top RT	1.375	0.991	***	-0.446	0.459	n.s.	-0.434	0.467
10  Top EDA	0.307	1.143	***	-0.075	0.977	n.s.	-0.117	0.864

Table 4.5: RT differences for stimuli subgroups and similar stimuli (by SST).

\*\*p < 0.01, \*\*\*p < 0.001

Table 4.6: EDA reaction differences for stimuli subgroups and similar stimuli (by SST).

Subgroup	Core group		Similar group			Rest group		
	mean	SD	р	mean	SD	р	mean	SD
10 Positive	0.099	1.018	n.s.	-0.127	1.004	**	0.014	0.958
10 Negative	0.177	0.970	***	-0.055	0.968	n.s.	-0.059	0.996
10 Basic emotions	0.079	0.994	*	-0.028	0.999	n.s.	-0.024	0.982
10  Top RT	0.317	1.009	***	-0.157	1.024	n.s.	-0.107	0.903
10 Top EDA	1.058	0.814	***	-0.336	0.752	n.s.	-0.362	0.770

p < 0.05, p < 0.01, p < 0.01, p < 0.001

The results are shown in Tables 4.5 and 4.6. For RTs, all differences are observed only on the level of core groups. The tests for positive and negative stimuli revisits the analysis done in Section 4.2.2, where the difference was observed also only in negative stimuli. Because of the overlap between negative stimuli and the stimuli close to 6 basic emotions, the difference was also found in this category. The significant difference in Top RT group is ensured by the definition of the group, and the longer RTs in Top EDA group represent the correlation between both behavioral parameters. Interestingly, there is no significant difference among "similar" groups. It means that the stimuli that are perceived similarly don't evoke the reactions of similar levels, at least if measured by SST.

For EDA the results show similar patterns. There is only one significant difference among "similar" groups, which could be attributed to multiple testing done in this analysis. Again, we can see that the stimuli similar to Top EDA group don't evoke higher EDA reactions than the rest of the stimuli.



Figure 4.13: Example of response groups in SST space for one respondent (ID03): 10 positive and 10 negative stimuli (top left), 6 basic emotions with 4 most similar stimuli (top right), 10 responses with highest RTs (bottom left), 10 responses with highest EDA reactions (bottom right).

### Chapter 5

### Discussion

After the results were presented, several limitations will be discussed, before we move to conclusion and implications (Chapter 6). The goal of this chapter is to offer explanations and identify limitations of the methods used.

The discussion will be structured into three main topics: procedure, stimuli material, and stability of results. In Procedure section the issues on condition control and the similarity with traditional WAT will be discussed. The section Stimuli material is focused on the nature of reactions evoked by the stimuli words. In the last section the problem of first and repeated WAT administration is discussed.

### 5.1 Procedure

The experimental procedure in this study was designed to mimic the traditional WAT procedure and to allow physiological measurements. If the experiment situation resembles the clinical WAT procedure, it is possible to transfer the results back to clinical assessment. If we can control the parameters in experiment, we can attribute the observed changes in behavior to the effect of these parameters.

These two goals were fulfilled up to some degree, but it is important to note that they are partly incompatible. If the patient is relaxed, the stimuli are presented orally by the administrator, and he/she can move freely, the experimental control is affected. On the other hand, with increasing experimental control, the test situation differs more from the original method.

There are EDA studies employing less experimental control and studying characteristics in longer ongoing tasks, but their authors focus on state or trait questions, e.g. (e.g. Cruz & Larsen, 1995). Such studies correspond to the second or third paradigm (as proposed by Dawson et al. (2000), see Section 1.2).

#### 5.2 Stimuli material

In current study the subject saw stimuli words on screen and their task was to "respond with the first word that comes to his/her mind" (see Section 3.2.3). Several types of questions can appear. Is the verbal response necessary? Do the stimuli words evoke emotions? Can we use pictures instead of words?

If no task was used, the subjects would be just sitting relaxed, watching a screen and there would be no control whether they read the stimuli words. If some other tasks, such as memorizing or word detection, were employed, the cognitive component of the process would be emphasized. The WAT instruction used in this study activates both components: cognitive and emotional. The unusual words, or the words that are hard to imagine (abstract words in this study), evoke the cognitive workload, because the subject has to identify the words and produce some related responses. It is not sure, whether the simple exposition of emotional words without specific instruction would cause

emotional reaction. But with this instruction subject needs to activate the concepts related to the stimulus, and this activation is expected to produce emotional reactions.

The emotional reaction is not expected to be directly connected to the stimulus, e.g. the subject doesn't feel "pain" or "sadness" if exposed to these stimuli words. On the other hand, it is expected that the level of evoked personal discomfort or emotions associated with a stimulus are related to its valence rating. Larger physiological reactions are observed with negative stimuli, which suggests that the discomfort is more likely to activate physiological reactions than the pleasant associations activated by positive stimuli.

It is possible that the activation of emotion-related concepts is done simply on cognitive or linguistic level, i.e. subject responds with an antonym ("fun-boredom") or a frequent collocation ("fun-laugh"). Such task does not require activation of personal concepts or memories, and only a relevant cognitive workload is observed. The cognitive workload is always present (the word must be processed), but the emotional workload can be missing – it can attribute to the fact that larger effects of cognitive workload were observed in this study.

The pictures could be also used as stimuli material, but it would not suspend a cognitive workload. Some studies with pictures reported that EDA reactions were related to attentional processes. Flanagan (1967) presented pictures to his subjects and asked them to report the intensity of their emotional reactions. Later, they were asked to rate the "attention-getting" of the pictures. The correlations between the attention-related scale and EDA was higher (r = 0.61) than with emotion-related scale (r = 0.33).

A possible extension of the current study would be to use stimuli of dif-

ferent intensities in each condition. The intensities are usually obtained from different subjects, who are asked to rate stimuli on various scales (e.g. intensity, valence). The stimuli with high intersubject agreement are later used in experiment and they are characterized by mean intensity values. Bernat et al. (2006) showed that EDA reactions on pictures with emotional content are different based on the content. In erotic and threat content the intensity of EDA reactions was increasing with higher content intensity. But adventure and victim conditions the EDA reactions didn't differ across the content intensity. The physiological methods differed in sensitivity in each content category.

### 5.3 Stability of results

The current study didn't inspect the test-retest reliability of the results for several reasons. First, this issue has already been inspected by Šlechta (2002b) in very similar design. Second, if WAT is used as a screening method, only one result is available for interpretation.

Slechta found that test-retest correlations after one month are very weak (Spearman  $R \doteq 0.35$  for RTs,  $R \doteq 0.10$  for EDA). This difference can be interpreted that the cognitive worload is similar in retest situation, but the emotional workload is very different. The negative stimuli are probably less threatening or surprising, and the subject's attitude toward the test and the test anxiety are different, which influences the results.

Considering low sensitivity of physiological responses in WAT found in this study, the stability of results in time is probably not necessary.

### Chapter 6

### **Conclusion and implications**

In this chapter, first we will review the results and compare them to the Šlechta's study (2002b). Second the main question of the project will be addressed: can we rely in EDA interpretation on single changes in EDA? Finally, the potential of SST testing within WAT procedure will be described.

#### 6.1 Results replication

The original study (Slechta, 2002b) which we further developed in this project, included both healthy subjects and patients tested with one WAT protocol. Here we can compare only the results for healthy subjects, but larger sample was tested (N = 80 compared to N = 36) and alternative protocols differing in content and presentation order were used.

Basically, the effect of cognitive workload on RTs and the effect of emotionally workload on EDA reactions were confirmed (see Figure 6.1<sup>1</sup>). Both effect sizes were compared and we found that the cognitive workload effects show more consistent differences in data (Cohen's d = 0.51).

<sup>&</sup>lt;sup>1</sup>Reprint of Figure 4.3, see page 55 for more details.



Figure 6.1: Effects of cognitive and emotional workload on reaction time (RT), electrodermal activity peak values (EDA) and pupillary responses (PR). The significant differences (tested with t-test) are marked with \*\* (p < 0.01) or \*\*\* (p < 0.001).

Statistically significant but low correlations between RTs and EDA reaction were observed (Pearson r = 0.23), which is similar to the previous results by Šlechta (2002b, p.39).

The pupillary reactions showed similar pattern to RT effects, and therefore it seems to be more related to the manifestation of cognitive workload. It is not possible to generalize this observation since it is tightly related to WAT design, in which the pupillary reactions show greater differences for concrete/abstract neutral words. In a different design, pupillary response could reflect emotional reaction. For example, (Bradley, Miccoli, Escrig, & Lang, 2008) used a picture–viewing task to assess the effect of picture valence on EDA, PR, and heart rate. They found similar patterns in EDA and PR, with high reactions in both pleasant and unpleasant stimuli.

For clinical applications, the distinction between neutral and negative is of

the biggest interest. Here, measuring EDA reactions proved to be the method indicating largest differences between these two stimuli groups (Cohen's d = 0.24). Still the effect is small, and it doesn't allow simple interpretation of individual results as it is shown in Section 6.2.

### 6.2 Interpretation of physiological responses

EDA reactions (SRRamp.) proved to be the parameter with best relation to emotional valence. If we consider an application example, after the WAT administration, the psychologist gets data like in Figure 6.2. We can observe some high EDA peaks in neutral words and low peaks in negative words, and the question is, whether we can reliable interpret them.



Figure 6.2: Example EDA data from WAT.

The small effect found unfortunately doesn't allow a straightforward in-

terpretation. If the negative and neutral stimuli were classified using EDA reaction, the sensitivity and specificity of this decision is very low (see Figure 6.3 or Figure 4.4 on page 57 for ROC). The variance within the stimuli category groups is too large compared to the group difference.



Figure 6.3: Sensitivity/specificity plot for negative/neutral classification based on EDA data.

Several factors were identified, which could influence the EDA results, but the relationships even statistically significant explain only small portions of variance. The simple or trivial responses (responding with a frequent word) were associated with shorter RTs, lower EDA and pupillary responses (see Section 4.3.6). The responding with a frequent stimulus-response pair was associated only with shorter RTs. Each of these effects explains at most 1.3% of variance<sup>2</sup>.

Because the emotional valence of the stimuli doesn't necessarily match the subjective valence, the subjective valence based on SST was assessed. The subjective valence differed slightly from the experimental condition (Kendall  $\tau = 0.64$ ), but the results for subjective valence were similar to the first findings. The subjective valence used finer scale and revealed that there is a linear trend for EDA reactions in negative part of scale, with EDA reactions increasing with more negatively rated stimuli. In the positive part the reactions did not show consistent differences (see Figure 4.8 on page 63).

The similarities between stimuli were inspected to reveal whether the stimuli perceived as similar evoke similar reactions (see Section 4.4.3). Especially it was interesting whether stimuli with high RT or EDA reactions will demonstrate also higher reactions in stimuli in their SST proximity, but no significant match was found.

It means that either WAT and SST are largely independent, or the reaction similarities manifest on a different level of similarity. In each part of this analysis we identified 25% of stimuli and compared them with 25% similar stimuli and 50% other stimuli. A different approach would be to study whether various stimuli subgroups are significantly more likely to be grouped together in SST projection. Probably some measure of mean stimuli distance could be used and compared with the mean distance distribution for subgroups of the same size within the same projection. This approach requires more advanced mathematical methods beyond the scope of this project. More simple methods could be confounded by comparing the mutually dependent measures (distances) like when the distance correlations were compared in

<sup>&</sup>lt;sup>2</sup>Example based on determination coefficient for stimulus–response frequency and RT (Pearson r = -0.113).

Section 4.3.1.

To conclude, based on the described data and methods the physiological reactions observed in EDA can't be automatically intepreted as having some emotional or personal relevance. This relevance would likely appear in other methods used here, but it didn't happen. For clinical applications it means that the psychologist must rely on inquiry in the cases with surprisingly high reactions (for neutral stimuli) or low reaction (e.g. for negative stimuli). In this approach the possibility of automatic decision making based on stimulus reactions was assessed, but its specificity and sensitivity didn't prove useful.

### 6.3 Using SST within WAT procedure

The Semantic Selection Test was used in this project as a measure of subjective similarity, and both methods could be also used in clinical situations. When the semantic similarity and the similarity of reactions were assessed, no significant relationships were found, which indicates that both methods provide on some level independent results. On the other hand subjective valence dimension is usually found in SST projections as one of the main axes. The advantage of SST projection is a simply readable output (see example in Figure 4.12 on page 70), which could be used either by a psychologist to inspect the data, or together with a client to discuss the results. This discussion approach can be a useful start for inquiry inspecting also specific findings in WAT.

WAT and SST share the advantage that the psychologist can adapt the stimuli based on the situation requirements. Both methods can be used as largely independent information sources.

#### 6.4 Summary

Data support the original findings that RTs and EDA reactions reflect in WAT different situation requirements: cognitive workload and emotional workload, respectively. The effect of cognitive workload is larger and manifests also in EDA reactions, while EDA demonstrated the best sensitivity to emotional valence condition from all employed methods (RTs, EDA, pupillary reactions). The effects are unfortunatelly too small to allow a reliable classification of responses.

When the influence of other factors was inspected (subjective valence, corpus frequency of responses), the effects were significant but low. Thus they can't be used to refine the reaction data, but still they can be considered in other WAT protocol designs as important factors.

The semantic proximity measured with SST didn't help to reduce the variance in WAT physiological reactions. As both WAT and SST can be easily adapted to specific assessment situations, both methods can be used as largely independent sources of assessment data.

### Appendix A

## Appendix

### A.1 Word stimuli

In this section you can find tables containing information about stimuli used in the study (Tables A.1 and A.2). In each table following information can be found:

Czech	Czech form used in the experiment				
English	English translation, possible other meanings				
	are mentioned in footnotes				
Valence	Emotional valence of the stimulus – positive				
	(+), neutral $(0)$ or negative $(-)$				
Concr.	Level of concreteness – concrete stimulus (C)				
	or abstract (left blank)				
Percentile	Frequency percentile of the lemma in				
	SYN2005 corpus				
Log frequency	The logaritmized number of word occurences per million words in SYN2005				

The numbers on the left side indicate the presentation order. In the alternative order condition first the stimuli 21 - 40 were presented and than the stimuli 1 - 20.

	Czech	English	Valence	Concr.	Percentile	Log freq.
1	kámen	stone	0	С	99.22	2.00
2	ztráta	loss	-		99.35	2.05
3	rozhovor	conversation	0		99.28	2.02
4	pochvala	praise	+		89.74	0.66
5	hala	hall	0	$\mathbf{C}$	98.64	1.77
6	nemoc	illness	-		99.06	1.92
7	událost	event	0		99.42	2.09
8	hrozba	threat	-		97.73	1.51
9	hodiny	clock	0	$\mathbf{C}$	99.82	2.51
10	radost	joy	+		99.43	2.10
11	pořadí	order	0		97.53	1.47
12	vražda	murder	-		98.85	1.84
13	fáze	phase	0		98.83	1.83
14	úžas	$surprise^1$	+		95.45	1.14
15	povaha	personality	0		98.74	1.80
16	štěstí	happiness	+		99.27	2.01
17	$\operatorname{prkno}$	board	0	$\mathbf{C}$	96.47	1.29
18	úzkost	anxiety	-		97.23	1.41
19	výměna	change	0		98.44	1.70
20	legrace	fun	+		97.08	1.39
21	brada	chin	0	$\mathbf{C}$	97.22	1.41
22	$\operatorname{smutek}$	sadness	-		97.25	1.42
23	měsíc	$\mathrm{moon}^2$	0	$\mathbf{C}$	99.82	2.51
24	hnus	disgust	-		88.38	0.59
25	kladivo	hammer	0	$\mathbf{C}$	94.18	1.00
26	polibek	kiss	+		96.60	1.31
27	$\operatorname{smlouva}$	contract	0		99.68	2.31
28	dárek	present	+		97.75	1.52
29	dveře	door	0	$\mathbf{C}$	99.82	2.50
30	vztek	anger	-		97.42	1.45
31	oblast	area	0		99.84	2.55
32	důvěra	$\operatorname{trust}$	+		98.31	1.67
33	$\operatorname{minulost}$	past	0		99.27	2.02
34	strach	fear	-		99.49	2.14
35	kilometr	kilometer	0	$\mathbf{C}$	99.13	1.95
36	bolest	pain	-		99.18	1.98
37	průběh	progress	0		99.41	2.08
38	fantazie	fantasy	+		97.46	1.46
39	hektar	hectare	0	$\mathbf{C}$	58.48	-0.34
40	láska	love	+		99.69	2.32

Table A.1: Word stimuli and their linguistic properties (version A)

Note. <sup>1</sup> or astonishment <sup>2</sup> or month

	Czech	English	Valence	Concr.	Percentile	Log freq.
1	kraj	region	0		99.67	2.30
2	$\operatorname{smutek}$	sadness	-		97.25	1.42
3	židle	chair	0	$\mathbf{C}$	98.69	1.78
4	naděje	hope	+		99.23	2.00
5	voda	water	0	$\mathbf{C}$	99.90	2.72
6	zdraví	health	+		98.96	1.88
7	podoba	$form^1$	0		99.54	2.18
8	vztek	anger	-		97.42	1.45
9	tužka	pencil	0	$\mathbf{C}$	96.46	1.28
10	výhra	victory	+		97.58	1.48
11	zpráva	message	0		99.71	2.35
12	úžas	$surprise^2$	+		95.45	1.14
13	auto	car	0	$\mathbf{C}$	99.73	2.37
14	slabost	weakness	-		96.32	1.26
15	budova	building	0	$\mathbf{C}$	99.53	2.17
16	krev	blood	-		99.45	2.11
17	reakce	reaction	0		99.13	1.95
18	láska	love	+		99.69	2.32
19	úroveň	level	0		99.64	2.26
20	pohoda	$\mathrm{comfort}^3$	+		97.17	1.40
21	list	$leaf^4$	0	$\mathbf{C}$	99.53	2.17
22	zranění	injury	-		98.47	1.71
23	výsledek	result	0		99.79	2.45
24	vina	$\operatorname{guilt}$	-		98.66	1.77
25	práce	work	0		99.92	2.82
26	$\operatorname{smich}$	laugh	+		98.81	1.83
27	pole	field	0	$\mathbf{C}$	99.45	2.11
28	hnus	disgust	-		88.38	0.59
29	dohoda	agreement	0		99.19	1.98
30	úspěch	success	+		99.52	2.16
31	etapa	phase	0		97.59	1.48
32	strach	fear	-		99.49	2.14
33	rameno	shoulder	0	$\mathbf{C}$	99.47	2.12
34	bezpečí	safety	+		97.09	1.39
35	talír	plate	0	$\mathbf{C}$	97.61	1.49
36	zrada	betrayal	-		95.03	1.09
37	přítomnost	present	0		99.04	1.91
38	závist	envy	-		93.31	0.92
39	okurka	cucumber	0	$\mathbf{C}$	94.98	1.08
40	radost	joy	+		99.43	2.10

Table A.2: Word stimuli and their linguistic properties (version B)

Note.

<sup>1</sup> or appearance <sup>2</sup> or astonishment <sup>3</sup> or peace <sup>4</sup> or sheet

### References

- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276-292.
- Bernat, E., Patrick, C. J., Benning, S. D., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, 43(1), 93-103.
- Boucsein, W. (1992). *Electrodermal activity*. New York: Plenum Press.
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45, 602–607.
- Brown, C. C. (1967). A proposed standard nomenclature for psychophysiologic measures. *Psychophysiology*, 4, 260-264.
- Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind*, 11, 220-242.
- Cattell, J. M., & Bryant, S. (1889). Mental association investigated by experiment. *Mind*, 54, 230-250.
- Cecero, J. J., & Karp, S. A. (1997). Word association profiles of alcoholics. Journal Of Personality Assessment, 69(1), 199–204.
- Český národní korpus. (2005). SYN2005. Ústav Českého národního korpusu FF UK, Praha 2005. (Retrieved August 20, 2007, from

http://ucnk.ff.cuni.cz)

- Český národní korpus. (2006). Srovnávací frekvenční seznamy z korpusů SYN2000 a SYN2005. Ústav Českého národního korpusu FF UK, Praha 2006. (Retrieved August 20, 2007, from http://ucnk.ff.cuni.cz/srovnani.html)
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography., 16/1, 22-29. Computational Linguistics, 16(1), 22–29.
- Cohen, J. (1992). A power primer. Psychological Bulletin, 112 (1), 155–159.
- Collins, A. M., & Loftus, E. F. (1975). A spreading activiation theory on semantic processing. *Psychological Review*, 82, 407-428.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Memory, 8, 240-247.
- Crosland, H. R., & Beck, L. F. (1931). Objective measurement of emotion. a symposium of three experimental papers. University of Oregon Publications, 1(3), 133-202.
- Cruz, M. H., & Larsen, R. J. (1995). Personality correlates of individual differences in electrodermal lability. *Social Behavior and Personality*, 23(1), 93–104.
- Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (p. 200-223). Cambridge University Press.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74-118.
- Feré, C. (1888). Note sur les modifications de la résistance électrique sous

l'influence des excitations sensorielles et des émotions. *Comptes Rendus des Séances de la Société de Biologie*, 40, 217-219.

- Filip, M. (2003). Test sémantického výběru jako psychodiagnostická metoda. Unpublished master's thesis, Masarykova Univerzita, Brno.
- Filip, M., & Urbánek, T. (2005). Nové přístupy k analýze dat testu sémantického výběru. Československá psychologie, 49, 563-578.
- Flanagan, J. (1967). Galvanic skin response: Emotion or attention? In Proceedings of the annual convention of the american psychological association (Vol. 2).
- Galton, F. (1879). Psychometric experiments. Brain, 2, 149-162.
- Galton, F. (1883). Inquiries into human faculty and its development. London: MacMillan.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. Journal of Personality and Social Psychology, 74, 1464-1480.
- Griffith, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114, 211-244.
- Hathaway, S. R. (1929). A comparative study of psychogalvanic and association time measures: A new psychogalvanic apparatus. *Journal of Applied Psychology*, 13, 632-646.
- Heřmanská, D., & Severová, M. (1958). O vztahu velikosti elektrokožních reflexů a délky latencí slovních reakcí v asociačním experimentu. Československá psychologie, 1, 18-33.
- Hume, D. (1748). An enquiry concerning human understanding. London: A.Millar. Available from http://www.gutenberg.org/etext/9662
- Hunt, W. A., & Landis, C. (1935). Word-association reaction-time and the magnitude of the galvanic skin response. The American Journal of

Psychology, XLVII(1), 143-145.

- Ivanouw, J. (2006). Stimulus affectivity of the danish word association test as measured by response heterogeneity and rasch scaled number of prolonged reaction times. *Scandinavian Journal of Psychology*, 47, 51-59.
- Jung, C. G. (1907/1991). Über die psychophysischen begleiterscheinungen im assoziationsexperiment. In L. Jung-Merker & E. Rüf (Eds.), Gessamelte werke 2: Experimentelle untersuchungen (p. 503-511). Olten: Walter.
- Jung, C. G. (1910). The association method. American Journal of Psychology, 31, 219-269.
- Kent, G. H., & Rosanoff, A. J. (1910a). A study of association in insanity part 1: Association in normal subjects. American Journal of Insanity, 66, 37-96.
- Kent, G. H., & Rosanoff, A. J. (1910b). A study of association in insanity part 2: Association in insane subjects. American Journal of Insanity, 67, 317-390.
- Klionsky, E. J., Okawa, J. B., Holmstorm, R. W., Silber, D. E., & Karp, S. A. (1998). Word-association patterns of mental hospital patients. *Psychological Reports*, 83, 1419-1424.
- Kondáš, O. (1989). Asociačný experiment: testová příručka. Bratislava: Psychodiagnostika.
- Kraepelin, E. (1892). Uber die beeinflussung einfacher psychischer vorgänge durch einige arzneimittel. Jena: Fischer.
- Locke, J. (1689). An essay concerning human understanding. London: William Tegg. Available from http://www.gutenberg.org/etext/10615

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty

knowledge test: New technique for the detection of deception. Journal of Experimental Psychology: Applied, 2, 164-177.

- Lukavský, J. (2003). Vliv temperamentových vlastností na elektrodermální aktivitu při asociačním experimentu. [The temperament characteristics and electrodermal activity during word association test]. Československá psychologie, 47(6), 503–512.
- Lukavský, J. (2004). Afektivní hodnocení testových slov jako rozšíření asociačního experimentu. [Subjective valence of the test words an enhancement of the word association test]. Československá psychologie, 48 (3), 203-214.
- Lukavský, J. (2007). Myšlení v projevech těla. In J. Kelemen, V. Kvasnička,
  & J. Pospíchal (Eds.), Kognice a umělý život VII. Opava.
- Martin, I. (1973). Somatic reactivity: Methodology. In H. J. Eysenck (Ed.), Handbook of abnormal psychology (p. 309-332). Robert R. Knapp.
- McClelland, J., Rumelhart, D., & PDP Research Group. (1986). Parallel distributed processing: Explorations in the microstructure of cognition (Vol. I and II). Cambridge, MA: MIT Press/Bradford Books.
- Metalis, A., Spero, Rhoades, B. K., Hess, E. H., & Petrovich, S. B. (1980). Pupillometric assessment of reading using materials in normal and reversed orientations. *Journal of Applied Psychology*, 65, 359-363.
- Novák, Z. (1992). Volné slovní párové asociace mužů a žen. PÚ JAK ČSAV.
- Novák, Z. (1996). Normy volných slovních párových asociací. pětiletí. sedmiletí. devítiletí. jedenáctiletí. třináctiletí. patnáctiletí. Praha: Idea servis.
- Oka, S., Chapman, C., & Jacobson, R. (2000). Phasic pupil dilatation response to noxious stimulation: Effects of conduction distance, sex, and age. *Journal of Psychophysiology*, 14, 97-105.

- Peterson, F., & Jung, C. G. (1907). Psycho-physical investigations with the galvanometer and pneumograph in normal and insane individuals. *Brain*, 30, 153-218.
- Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. In D. Harris, V. Duffy, M. Smith, & C. Stephanidis (Eds.), *Human-centred computing: Cognitive, social, and ergonomic aspects* (Vol. Vol. 3 of the Proceedings of the 10th International Conference on Human-Computer Interaction, pp. 542–546).
- R Development Core Team. (2007). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from http://www.R-project.org (ISBN 3-900051-07-0)
- Schlegel, M., & Zeier, H. (1982). Psychophysiologische aspekte des assoziationsexperiments und normdaten zu einer reizwörterliste. Analytische Psychologie, 13, 75-92.
- Šlechta, P. (2001). Možnosti psychologické diagnostiky na základě dat kožní vodivosti během testu slovních asociací [Possibilities for the psychological diagnostic based on the skin conductance data during word association test]. Československá psychologie, 45 (5), 460-469.
- Šlechta, P. (2002a). Effectiveness of screening for psychological disturbances using word association test. Československá psychologie, 46(6), 509– 520.
- Slechta, P. (2002b). Effects of stimulus affectivity and concreteness on electrodermal and verbal responses in word association test. Unpublished doctoral dissertation, Charles University, Prague.
- Smékal, V. (1990). Psychosémantické metody. In L. Maršálová & O. Mikšík (Eds.), Metodológia a metódy psychologického výskumu. Slovenské ped-

agogické nakladatelstvo, Bratislava.

- Spitzer, M. (1992). Word-associations in experimental psychiatry: A historical perspective. In M. Spitzer, F. A. Uehlein, M. A. Schwartz, & C. Mundt (Eds.), *Phenomenology, language & schizophrenia* (p. 160-196). New York: Springer-Verlag.
- Spitzer, M. (1994). Key concepts: Associanism. Philosophy, Psychiatry and Psychology, 1(2), 135-7.
- Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, 52(1), 77–86.
- Synek, S., & Skorkovská, S. (2004). Fyziologie oka a vidění. Grada.
- Urbánek, T. (2001). Optimální škálování testu sémantického výběru. Zprávy
   Psychologický ústav, 7(1), 1-26.
- Urbánek, T. (2003). Psychosémantika. Nakladatelství Pavel Křepela.