THE UNIVERSITY OF HELSINKI

Measuring Digital Game Experience

Response Coherence of Psychophysiology and Self-Reports

Simo Järvelä
Master's thesis
Cognitive Science
Department of Modern Languages
The University of Helsinki
April 2017

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty Humanistinen tiedekunta		Laitos – Institution – Department Nykykielten laitos
Tekijä – Författare – Author Simo Järvelä		
Työn nimi – Arbetets titel – Title Measuring Digital Game Experience: Response Coherence of Psychophysiology and Self-Reports		
Oppiaine – Läroämne – Subject Kognitiotiede		
Työn laji – Arbetets art – Level Pro gradu	Aika – Datum – Month and year 3.4.2017	Sivumäärä– Sidoantal – Number of pages 32

Tiivistelmä – Referat – Abstract

Tässä työssä tutkitaan psykofysiologisella menetelmällä ja itseraportoinneilla saatujen tulosten vastauskoherenssia pelattaessa pitkiä pelijaksoja digitaalisia pelejä. Emootioiden valenssia ja virittyneisyyttä voidaan mitata molemmilla menetelmillä, ja niillä saatujen tulosten on todettu mm. kuvankatselukokeissa korreloivan melko voimakkaasti. Digitaaliset pelit ovat kuitenkin stimuluksena ja tutkimuskohteena huomattavan erilaisia kuin staattiset kuvat, sillä ne ovat luonteeltaan interaktiivisia, nopeatempoisia, kompleksisia, ja lisäksi niitä pelataan tavoitteellisesti ja päämäärähakuisesti. Psykofysiologisessa pelitutkimuksessa on aiemmin huomattu kuinka vastauskoherenssi näiden menetelmien välillä on melko vaihteleva, mutta sitä ei alalla kuitenkaan ole toistaiseksi systemaattisesti lähdetty tarkastelemaan.

Työssä analysoidaan neljän eri digitaalisen pelin osalta kuinka vahvasti psykofysiologisella menetelmällä mitatut kasvonlihasaktiviteetit sekä kämmenistä mitattu ihon sähkönjohtavuus – vakiintuneet tavat arvioida valenssia ja virittyneisyyttä – korreloivat puolen tunnin mittaisten pelijaksojen päätteeksi täytettyjen valenssia ja virittyneisyyttä mittaavien itseraportointien kanssa. Aineiston otoskoko on 36, ja koehenkilöt ovat ikäväliltä 18-34 (ka = 24,0, s = 4,35) ja he ovat kaikki miespuolisia aktiivisia peliharrastajia. Psykofysiologian ja itseraportointien vastauskoherenssin lisäksi työssä tarkastellaan sitä ovatko itseraportoinnit vahvemmin yhteydessä välittömästi vastaushetkeä edeltäviin ajanhetkiin psykofysiologisesta mitta-aineistosta kuin koko mittausjakson keskiarvoon.

Tutkimuksen tulokset osoittavat, että vastauskoherenssi on systemaattinen ja oletetun suuntainen, mutta efektikoot ovat huomattavasti kuvankatseilukokeita alhaisempia riippumatta siitä mihin ajanjaksoon fysiologisesta mitta-aineistoista vertailu tehdään. Johtopäätöksenä esitetään, että psykofysiologisia mittareita pelitutkimuksessa käytettäessä on syytä toisaalta olla hyvin harkitsevainen saatujen tulosten tulkinnan suhteen, sekä toisaalta pyrkiä tiukkaan kontrolliin koeasetelmasuunnittelussa, jotta mahdollisuus mielekkäiden tulkintojen tekemiseen säilyy.

Avainsanat – Nyckelord – Keywords
Digitaaliset pelit, psykofysiologia, kognitiotiede
Säilytyspaikka – Förvaringställe – Where deposited
Keskustakampuksen kirjasto
Muita tietoja – Övriga uppgifter – Additional information

Table of contents

1 Introduction

“Thus, emotional judgments, physiology, and behavior can present a confusing rock pile that resists a simple classification by specific emotional states.”

– Lang, 1995

The rise of digital games to a dominating position as an entertainment media form (cf. ESA, 2016) demands thorough scientific research on their effects on consumers. Psychophysiology (Cacioppo, Tassinary, & Berntson, 2007) has been one method that is utilized in studying digital games in increasing amount during recent years, especially focusing on studying the game experience playing digital games elicits. In psychophysiology various physiological signals are measured and inferences regarding psychological states are conducted based on them (Cacioppo et al., 2007). In psychophysiological experiments the various components of emotions (see chapter Psychophysiology for more details) are typically assessed by both measuring a selection of physiological signals and by self-reports (Cacioppo, Tassinary, & Berntson, 2000; Scherer, 2005). This multi-method measurement approach aims at a more reliable assessment of a particular theoretical construct by utilizing more than one method of measuring it (Scherer, 2005). Facial electromyography (fEMG) (Fridlund & Cacioppo, 1986) and electrodermal activity (EDA) (Dawson, Schell, & Filion, 2000) have been established as reliable methods for assessing emotional valence and arousal respectively e.g. when viewing still images with strong emotional content (Lang, Greenwald, Bradley, & Hamm, 1993). The studies examining the validity of these methods in game studies conducted so far carefully recommend their usage, and in general have found a varying degree of response coherence between self-reports and physiology (for a review, see Kivikangas et al., 2011). However, despite some accumulation of research data, a systematic examination of the coherence of tonic physiological activity and self-reports is lacking.

The relation of physiological signals and psychological constructs has always been many-fold (see e.g. Cacioppo et al., 2000; Kreibig, 2010), and the groundwork studies in this field have been conducted on a vastly different type of stimulus than digital games of today. In order to be truly useful method in examining the game experience, the validity of the very fundamentals of using psychophysiological measurements in assessing it must be examined. Even though there is an increasing amount of studies conducted using these mixed methods, such groundwork is still sorely missing. Consequently, it is quite unclear how the various results achieved with this method should be interpreted in gaming context. The whole plausibility of the method is at stake when there is no solid and established way of interpreting the results acquired. This is especially worrying as the many-to-many relationship of physiological signals and psychological constructs already establishes certain innate ambiguity to the method (Cacioppo et al., 2000). Ultimately, combined with publication bias (Kühberger, Fritz, Scherndl, Haug, & Hoey, 2014) it could lead to researchers reporting only those results that are in-line with the established theory from other research areas or those that support their own hypothesis even though the rest of the results would also be entirely solid provided there was a theory how they should be interpreted.

It is entirely possible that psychophysiology as a method is not suitable for all studying all aspects of game playing. For example, digital games can simply be too complex to be studied with a method that basically requires a well-controlled experimental setup and strict stimulus control (see, chapter Digital Games as Stimulus, cf. Järvelä, Ekman, Kivikangas, & Ravaja, 2013). Yet, the psychophysiological method has potentially substantial benefits over many other methods when studying gaming experience (Kivikangas et al., 2011; Mandryk, Inkpen, & Calvert, 2006; Ravaja, 2004; Yannakakis, Martinez, & Garbarino, 2016), such as providing data regarding subconscious phenomena that are difficult to assess objectively with self-reports, and being able to collect data in real time with high temporal accuracy without interrupting the game play flow. The true capabilities and limitations of psychophysiological methods for studying games are still unclear and finding them is a laborious task for which this work aims to contribute to.

In this work the response coherence (see chapter Response Coherence) of emotional valence and arousal measured with psychophysiological methods and self-reports are examined. Presuming both methods in fact measure the same theoretical constructs – such as valence and arousal dimensions of emotion – in a reliable and valid manner when studying digital games, they should be highly correlated even though physiological measures are not unambiguous. The main research question of this work is:

How strongly physiological signals and self-reported valence and arousal are correlated when examining digital game play over extended playing times?

As physiological signals are measured continuously during the playing period and self-reports are collected after it, it can be hypothesized that the response coherence of physiological measures and self-reports vary depending on the chosen time segment of physiology, e.g. that the final moments of play time would dominate the self-reported experience. Also, it can be hypothesized that the relation of physiology and self-reports depends on the stimulus game used and that it would not be the same for all games. For these reasons the research question will be examined by analysing the response coherence of self-reports and different segments of physiology in four different games. These analysis' aim to provide understanding how physiological signals can be interpreted when measuring extended periods of digital game play. While certainly not enough to entirely falsify or validate existing theory, this work will contribute to mapping out the parts of established theory that seem suitable for more complex media forms and where there seems to be contradictions that need more work in order to be solvable.

2 Emotion Theories

Considering emotions have been scientifically studied for over a century, starting from the founding fathers of psychological science Wundt (1897) and James (1884), there is surprisingly little consensus as to what emotions really are. There is no single agreed upon definition, but dozens of different theories and a several schools of theory (Dixon, 2012; Izard, 2010a, 2010b).

Probably the most well-known are the discrete emotion theories and especially the basic emotion theory (cf. Ekman & Cordaro, 2011) that suggests that there is a modest number – typically from five to seven depending on the theory – of basic emotions (or emotion families) that are discrete and universal. They emphasize how this set of basic emotions is limited, and other emotions are based on these or are variations of them. Basic emotions are seen as at the core being evolved beneficial adaptions to certain conditions, but that there is additionally an ontogenic quality to them so that they are adapted to the life experiences and surroundings of each person.

Dimensional models of emotions follow in Wundt's footsteps and emphasize how all emotions have a number of shared basic dimensions such as pleasantness – unpleasantness, excitement – calm, and relaxation – strain (Wundt, 1897), or arousal and cognitive label (Schacter & Singer, 1962), or valence, arousal, dominance (Lang, 1980), etc. Some dimensional theories posit that emotions such as fear or joy can be placed in dimensional space, such as circumplex (e.g. Posner, Russell, & Peterson, 2005; Russell & A., 1980; Yik, Russell, & Steiger, 2011). Valence and arousal are related to withdrawal/approach motivation where positive valence increases approach motivation and negative valence increases withdraw motivation (Elliot & Covington, 2001; Lang, 1995) and arousal is in essence the power of that emotion. However, there also arguments that valence cannot be regarded as a single bipolar scale from negative to positive valence, but two independent unipolar positive and negative affect scales instead (Tellegen, Watson, & Clark, 1999). A more recent dimensional model is the Evaluative Space Model (ESM) that sees withdrawal and approach as the output of the affective system, and that these are dependent on various factors that are different for positive and negative affects (Norris, Gollan, Berntson, & Cacioppo, 2010). Dimensional/circumplex models are commonly used as the background framework with psychophysiological methods (see chapter Psychophysiology).

Appraisal theories (see e.g. Roseman, 2013; Scherer, 1999; Smith & Lazarus, 1990) see emotions as collections of functions that appraise stimuli and provide behaviour guiding motivations, e.g. a snake is appraised to be dangerous and through fear and withdrawal behaviour is activated. Central to appraisal theories is how emotion as a construct is divided into a set of appraisals that are clearly defined and specific, which also allows them to be studied separately. For example, Roseman (2013) and Emotion System model present several separate appraisals, such as motive inconsistency vs. consistency, high vs. low control potential and instrumental vs. intrinsic problem types. These appraisals form the subjective significance and assessment of the appraised situation, event, or object, and produce behavioural motivation depending on the motivation in that context. While affective and cognitive processing have been historically considered to be separate, appraisal theories very strongly emphasize the traditionally cognitive aspects in emotional processing, and similar elements are included in most modern emotion theories.

Emotion constructionists (see e.g. Barrett, 2009, 2013) claim that what we call emotions are not in fact emotions but constructed from more primitive cognitive processes and consequently very much cultural and language based. Constructionists often underline how much difficulties other emotion theories have in explaining the complexity and abundance of different emotional experiences. According to Barrett (Barrett, 2013) several different theories can be included within the constructionist approach, such as OCC model of emotion (Clore & Ortony, 2013) where emotions are embodied and situated representations of situations experienced by the subject that retain their structural characteristics, the conceptual act theory (e.g. Lindquist, 2013) that has the same basic tenets as OCC but further emphasises the contextual aspect in the specific situation and how it affects the processing, and the iterative reprocessing model (Cunningham, Dunfield, & Stillman, 2013) that underlines how the situational component is iteratively reprocessed and changes the emotion representations over time. Constructionist emotion theories have been rarely used in experimental game studies as they are quite hard to operationalize and they do not offer clear predictions to base hypothesis on.

Then there are theories such as LeDoux's (2000, 2012)) survival circuits and Panksepp's (1982, 2005) primal processes that take on very evolution biological perspective and underline how emotions are part of circuitry that have a very clear function why they have developed in evolution. These theories emphasize how primitive the core emotional processing is, and that is shared by all primates (or vertebrates, perhaps even non-vertebrates), and the complexity of human emotional experiences is built on that basis. These more neuroscientifically informed theories and their basic claims are quite widely accepted among different families of emotion theories, though it seems common to interpret them as supporting their own theories. It would be tempting to label these as being basic emotion theories as they posit that separate discrete neural structures exist for certain emotions, however, they approach the whole question from a different direction and discuss mainly existing neural structures that most likely are behind some cognitive features incorporated in other emotion theories, and thus they do not posit themselves as being part of any traditional schools of emotion theory.

In the end, surprisingly few integrative models that try combine the different features of various theories into a single model currently exist (cf. Russell, 2014), and existing theories are widely considered competing and mutually exclusive. For a recent rare attempt at integration see Kivikangas's Affect Channel Model of Evaluation (ACME) (Kivikangas, 2016)

Despite the wide variety of theories, what is widely agreed on though is that emotion is not just the subjective phenomenological part to which a lay person commonly refers to when talking about emotions. In fact, there are different components to emotion, and that subjective component is typically called 'feeling'. Roseman (2013) separates five different components: *phenomenological* (feelings, thoughts), *physiological* (neural and muscular activation patterns), *expression* (facial expressions, posture), *behavioral* (actions); and *emotivational* (goal directed motivations). Several other views on components of emotion exist, but in all of them the essence is roughly the same: emotions are not merely the phenomenological feeling a person has, but other components such as physiological or expressive are just as essential. A crucial aspect of this multicomponent nature of emotions is that it additionally allows them to be assessed through other measures besides self-reports; mainly the physiological component can be precisely measured with proper equipment. That method is called psychophysiology, see the following chapter for more details on the psychophysiological method.

3 Measuring Emotions

The complexity and multifacetedness of emotions as phenomena and theoretical constructs make measuring them quite challenging (Scherer, 2005). Different components of emotions require their own measurement methods, for example self-reports for subjective feelings and psychophysiology for physiology and expressions. They have different strengths and weaknesses as methods, and rarely in they have a clear cut one-to-one relation to the theoretical construct that they are supposed to assess, and often the interpretation is less than straightforward. In this work, the focus is on the response coherence between psychophysiological measurements and self-reports when they are supposedly measuring the same emotional state.

3.1 Psychophysiology

3.1.1 Introduction to Psychophysiology

Psychophysiology is a method where psychological phenomena – such as emotions or attention – is assessed by measuring various physiological signals like heart rate (HR), facial electromyography (fEMG), electroencephalography (EEG), or electrodermal activity (EDA) (Cacioppo et al., 2000). It has a long tradition, and in recent years it has been utilized in studying media experience (Ravaja, 2004) including digital games (see chapter Psychophysiology in Games Research for more details). The psychophysiological method's main benefits are that it is a continuous measurement (compared to e.g. post-stimulus ratings), it has high temporal resolution (milliseconds), and it is to a large degree immune to various biases such as social desirability or wish to please the experimenter, as people are not typically aware of their physiological states and not able to change them consciously, and its reliability is not dependent on the experimenter. Additionally, it allows the examination of very short term reactions depending on the signal measured [EEG reacts within a hundred milliseconds, but it takes a couple of seconds for the sweat glands in the palms to activate (Davidson, Jackson, & Larson, 2000; Dawson et al., 2000)], but can also be utilized in measuring extended periods of time without interrupting natural behaviour. However, utilizing psychophysiology is not trivial; it requires well designed experimental setup, precision in conduction and following the protocol, a lot of data processing, often complex statistical analysis, and a solid background theory both on the psychophysiological method itself and the topic of investigation. Yet, the most challenging aspect of psychophysiology is that the signals are not in one-to-one relation to certain psychological phenomena, but the relations are complex many-to-many relations where a single signal can be interpreted to tell something about various constructs, and sometimes even contradictory interpretations exists within the literature. See Kreibig's review (Kreibig, 2010) for more details on the complexity of autonomous nervous system (ANS) signals. This in practice means that interpreting what the measured signal actually tells us is far from simple.

Naturally the interpretation of various physiological signals requires an established background theory of emotions that has both the rationale why and how physiological signals are connected to the psychological phenomenon, and the empirical research to back it up. It is commonly agreed that emotions have physiological changes as one component, but making interpretations about them based on for example constructionist theory of emotion would be quite challenging. With psychophysiology, basic emotion theory is occasionally used, especially in such experimental setups where facial expressions are under scrutiny (see e.g. Ekman, 1993). However, dimensional models of emotions are most commonly utilized as the dimensions of valence and arousal are particularly suitable theoretical constructs for psychophysiological interpretations, and it has been established quite robustly that certain physiological signals are strongly connected to them. Circumplex model (see chapter Emotion Theories) in particular is widely used. Other theorists make a case that instead of a single valence dimension, there are separate dimensions (and systems in the brain) for positive affect (PA) and negative affect (NA), which are not mutually exclusive (Larsen, McGraw, Mellers, & Cacioppo, 2004; Tellegen et al., 1999).

Electrodermal activity (EDA), or skin conductance, has been established as a reliable and valid method for assessing physiological arousal (Dawson et al., 2000). EDA is measured from the palms of the hands where special eccrine sweat glands activate when autonomous nervous system increases the bodily arousal. Sweat increases the conductance of the skin, and so by driving a tiny steady current to the skin and measuring the conductance between two points, the level of arousal can be assessed.

Facial electromyography (fEMG) measures the activity of facial muscles that are related to various facial expressions. When assessing valence, muscles that are related to smiling (for positive valence/affect) and/or frowning (for negative valence/affect) are typically measured. The large cheek muscle, zygomaticus major, is in central role in smiling expressions, as is the muscle that closes the eyelids, orbicularis oculi, and these are often both measured. The brow frowning muscle, corrugator superscilii, in turn typically activates in many negative valence expressions such as anger, frustration or disgust. When measuring muscle activation, it is the small electrical current created by the muscle that is measured. (Fridlund & Cacioppo, 1986) As the muscles related to positive and negative affect are separate, they can be easily measured separately and mapped to different theoretical constructs if the background theory allows it.

3.1.2 Psychophysiology in Games Research

For the last ten years or so, research on digital games have utilized the psychophysiological method to understand both digital games as a medium and the gaming experience, and also in studying more general psychological phenomena using games as the activity that brings out the effect under scrutiny (Kivikangas et al., 2011; Mandryk et al., 2006). The field is scattered, and psychophysiology has been applied in quite a variety of ways, as is natural when applying an established method to a new field. Consequently, so far, while the results obtained from psychophysiological games research provide new insight to a variety of game experience related phenomena, they do not yet paint any clear holistic picture, but are rather scattered and often hard to compare (for a review, see Kivikangas et al., 2011). This is partly due to the specific expertise required by the psychophysiological method, but also due to the broader issue where experimental study of digital games is quite new also. So far, only a few generic guidelines how to use games as stimuli in experiments exist (see Järvelä et al., 2013; McMahan, Ragan, & Leal, 2011). Following good practices of experimental research requires a good understanding of the nature of the stimulus; and in the case of digital games that nature is exceedingly complex. Next we will discuss the unique nature of digital games as stimuli and how they differ from stimuli typically used in psychological experimental studies.

3.1.3 Digital Games as Stimulus

Digital games are a very unique entertainment media form. While not delving in-depth to what is the definition of a game, it can be said that the range of different games is vast, and in different ends of the spectrum it can be hard to distinguish them from other media forms such as movies, interactive storytelling etc. The very complex nature of digital games places hard challenges for anyone wishing to study them, especially with methods that require well designed experiments and strict stimulus control. In this chapter we examine digital games’ nature as a stimulus, and what aspects of it differentiate them from more traditional media forms.

First of all, digital games are highly dynamic and their content is constantly changing. This is evident when comparing them to for example pictures – a form of stimulus often utilized in psychological studies – that are mostly static. Videos, moving pictures, are more dynamic stimulus with its content in constant change.

However, some games are even more dynamic with the tempo of changes being far higher than in most movies.

Secondly, many digital games are very complex visually, with a dizzying amount of visual elements actively present on the screen and often all of them containing more or less relevant information regarding the game state and not merely eye candy. Consider for example a first-person-shooter game (FPS) or a massively multiplayer online (MMO) game and their combat scenes with multiple avatars, numerical indicators of various factors such as ammo, health, mana etc., and the coordination and complexity of multiplayer cohort's actions.

Digital games are also interactive by nature, that is, the content changes according to player actions. While there are some interactive video installations and choose-your-own-adventure type of books, interactivity is a rare feature in media. In digital games it's a defining feature, which makes games quite unique in this regard. This changes the whole relation of the media form and the consumer as the player has at least a degree of control over the content. Interactivity of games changes the whole process of media consumption from passive to active. Playing digital games is always an intentional goal directed activity where the player is actively trying to achieve something within the game (Suits, 1967). Consequently, a set of details become affordances (Deterding, 2011; Linderoth, 2012) for a certain action within the game that can be taken to forward one's goals, whereas a similar detail in passive media might still be meaningful and relevant, it would not be a behavioural cue.

Digital games can also be multiplayer games from two person games to massively multiplayer online games with thousands of players playing at the same time in an interactive game world, e.g. one of the largest battles in online games to date, the Bloodbath of B-R5RB in EVE Online (CCP Games, 2003) that took over 20 hours to play, involved over 2500 players and 7500 player characters at the same time (see https://en.wikipedia.org/wiki/Bloodbath_of_B-R5RB). The multiplayer aspects adds a layer of social complexity to the game that is not easily present in traditional media forms.

The dynamic, complex and interactive nature of games as stimulus is one aspect that make them and the emotions elicited by them hard to study experimentally, but even more challenging is how the whole activity is framed as *playing a game* (Deterding, 2009; Montola, 2012; Stenros, 2015). The play approach, depending how it is conceptualized – e.g. rather simply as paratelic state (Apter,

1984) where the activity is taken for its own sake, or as more holistic approach where games are seen as systems of endogenous meaning where that affects all interpretations of meaning regarding the activity (cf. Costikyan, 2002) – potentially thoroughly affects how we see the whole fundamental relation of digital games as stimuli and the emotions elicited by them. So far, this issue is rarely discussed and there lacks consensus in regard how this element should be taken into account in experimental games research.

All in all, studying emotional responses to digital games using psychophysiological methods is hugely challenging due to the limitations of the method and the nature of the stimulus itself. Particularly the combination of complexity of the game stimulus and how nearly impossible it is to precisely control them in an experimental setting, and how most physiological signals are connected to more than one psychological phenomena, makes it very strenuous to draw exact inferences regarding them.

3.2 Self-Reports

Self-reports primarily measure phenomenological feeling component of emotions, that is, emotional states (cf. e.g. Scherer, 2005). A vast array of different types of emotion state questionnaires exist, they mostly consist of a number of scales with individual items making statements about feeling like something and Likert scale to mark to what extent the statement applies to subject's current state. These questionnaires are clearly language based and ask the subject to reflect her own state and rate the statements accordingly. Other type of emotional state self-reports exist also, such as the pictorial Self-Assessment Manikins (Lang, 1980).

Self-Assessment Manikins (SAM) are pictorial bipolar self-report scales for assessing valence, arousal and dominance dimensions of experience (Lang, 1980). It was designed to measure the same dimensions that e.g. 18-item Semantic Differential Scale (Mehrabian & Russell, 1974) but more quickly and efficiently, and in a manner that is not as language dependent so it's more suited to speakers of different languages and aphasics (Bradley & Lang, 1994). SAM's have been validated with a variety of stimulus types, e.g. International Affective Pictures System (IAPS) (Lang et al., 1993), but also sounds (Bradley, 1994) etc.. The cross-validity of semantic multi-item questionnaires and SAM's have been established (Bradley & Lang, 1994) and found to be high especially regarding valence and arousal ratings. Also, the correlation of physiological measures and SAM's has been established (e.g. Lang et al., 1993). However, all these studies have been conducted using short term static non-interactive stimuli of different types, and their usability in assessing emotional reactions to interactive stimuli, such as digital games, for extended durations, have not been established even though they are quite often utilized in studying them.

3.3 Response Coherence

Though often left unstated, the basic assumption is that the different methods of measuring emotions are ultimately measuring the same theoretical constructs, such as valence and arousal, and consequently there should a high degree of coherence between the different measures. That is, for example assessing valence by using psychophysiology should provide very similar results as when assessing valence by self-reports, provided that the measures are both valid and reliable of course. This is assumed to be true also when measuring various components of emotions, e.g. Feeling, physiology or behaviour components, as the idea is that the emotion comprises of these components (Evers, Hopp, Gross, & Fischer, 2014). This assumption has strong roots as old as psychological emotion theories, as already William James saw emotions being fundamentally based on the physiological state that gives rise to the phenomenological component (James, 1884).

This relation has been validated in static picture viewing experiments where the level of response coherence between self-reports and psychophysiology is examined. For example in Lang et al. study (Lang et al., 1993) that utilized the standardized IAPS emotional picture set as stimulus (Lang, Bradley, & Cuthbert, 2008), the reported correlations were remarkably high ($.7 < r < .9$) between valence and fEMG, and arousal and EDA. Such studies provide strong support for the idea that they are indeed different components of the same theoretical emotion construct, and that emotional state can be measured with these two methods quite interchangeably regarding validity. In addition, most sources (e.g. Mandryk et al., 2006; Scherer, 2005) seem to recommend utilizing several different measuring methods simultaneously to ensure most reliable assessment of subject's emotional state instead of relying solely on one method of measuring. However, despite the strong correlations reported in specific studies, often in research papers the level of response coherence is extremely varied, and the evidence so far cannot be considered entirely solid (Evers et al., 2014).

To account for the non-uniform evidence supporting for the coherence hypothesis, further theoretical models have been developed to better explain the variety of empirical data. One class of such theories are the dual-processing models, which in the style of Kahneman's system 1 and system 2 (Kahneman, 2013) posit that instead of a single emotion system and various of methods of measuring its components, there are actually two separate systems (for a broader review on dual-processing models of cognition, see Evans, 2008). Evers et al. refer to these two systems as automatic and reflective systems (Evers et al., 2014).

"Dual-process frameworks assume that psychological responses are a joint function of two largely independent systems, one automatic and the other reflective. Automatic responses are relatively unconscious, fast, and efficient, while reflective responses are relatively conscious, deliberate, and effortful. Both systems are thought to play in concert to promote adaptive behavior, including emotions." (Evers et al., 2014)

They present empirical data to support the idea that the coherence of different measures of emotional components is not uniform, but greater within a single system and minimal across the systems. That is, different measures of automatic system would correlate, as do different measures of reflective system, but they do not correlate with each other because they are processed by two separate systems.

Another theory type that provides a framework for taking the variance in response coherence into account are what could be called the time frame theories of emotion. Their main idea is that emotional processing starts as rather simple stimulus appraisal and further develops into highly complex forms proceeding in stages within a certain time frame. A time frame theory, such as Affect Channel Model of Emotions (ACME) (Kivikangas, 2016), opens up the possibility that self-reports and psychophysiology are measuring different components of the same emotional appraisal process but at a different point in time within the process. That is, self-reports mostly measure the later stages and closer to the end result when the processing has advanced to conscious level, and psychophysiology can already catch the early on physiological reactions in milliseconds timeframe. As the processing advances, the later stages add complexity to the he early physiological reactions to the stimulus, e.g. by predicting its impact and by framing it in the current social context, and consequently the self-reported feelings might differ significantly from the measured physiological component. Due to the complexity of digital games as stimulus, the framing of the activity as play, and the long playing periods and consequent extended meaning structures, it is plausible that different time frames of emotional processing, and how different measurement methods measure different parts of that process, constitute to lower response coherence when studying digital games compared to static emotional pictures. The dual-processing theories and time-frame theories are not mutually exclusive and should be considered as complementary instead.

Within psychophysiological games research, the response coherence has not been specifically studied at length. Consequently, a paucity of proof that the assumptions studied with static picture viewing or sound clip listening experiments would hold true with digital games exists. Considering the lack of solid psychological background theory for game experience through which to interpret the results (cf. Kivikangas et al., 2011), and the challenges of precise experimental control when studying digital games, it would be highly surprising if the response coherence would research the levels of those reported by e.g. Lang et al. (1993). A thorough meta-analysis of published research papers utilizing psychophysiological research methods would be required the see the extent of response coherence, however that is beyond the scope of this work and remains a future research endeavour.

4 Methods & Experimental Setup

4.1 Participants

The participants were 36 predominantly university students, all male, and active gamers from age group 18 to 34 (M = 24,0, SD = 4,35). They were recruited using university mailing lists and various gaming related web forums.

4.2 Procedure / Research Design

Before arriving at the laboratory, the participants had filled out background and trait questionnaires. After signing the informed consent form, the electrodes were attached. The experiment started with an eight minute baseline recording and then the participants played each of the four games for 30 minutes in random order. After each playing period, they filled out the self-reports regarding their gaming experience. In the end of the experiment they could freely choose which game or games they would play for one hour. For the analysis in this work, only the 30 minute playing sessions were included and the data from the one hour free-play period omitted to maximize the comparability of the sessions.

Originally this data set was collected as a predictive validity study. The predictive validity results are reported in elsewhere (Kivikangas, 2015). That is the main reason why the extended play times were originally recorded. After the lab phase there was three week follow-up phase where the participants could play the same games freely. After the lab phase, the four games were given to the participants and after the follow-up period they could keep two of those games as a compensation for their time and effort.

4.3 Stimulus

The four stimulus games used in the experiment were Fahrenheit (FH, Quantic Dream, 2005), Operation Flashpoint: Cold War Crisis (OF, Codemasters, 2001), Painkiller (PK, DreamCatcher Interactive, 2004), and Sam & Max: Season 1 (SM, Telltale Games, 2006). The chosen games were slightly older at the time of the experiment to ensure maximum compatibility and less demanding hardware requirements for the participants own computers during the follow-up period included in the original study. Also this way, none of the game titles used in the experiment were under active marketing or such that would affect the player's perceptions of those titles during the experiment.

The games were chosen so that they represent two different popular game genres, first-person-shooters (FPS; PK & OF) and adventure games (SM & FH), and have a more serious (FH & OF) and a more humorous game (SM & PK) from within those genres. This way they would assumably elicit a wider variety of emotional responses and both positive and negative valence. The four titles were estimated to be approximately equal in quality by comparing a variety of reviews and comparing Metacritic (http://www.metacritic.com) scores which all were in the range of 81-85/100.

4.4 Procedure

Each game was played 30 minutes from the beginning of the game including the titles and tutorials to avoid researcher bias in selecting the scene. This way the participants had less of actual interactive play time, but would be playing the game as the designers intended, and also would go through the tutorials to help them grasp the basics of the game if necessary. Unfortunately, the same procedure could not be utilized with Operation Flashpoint as the intro sequence was so long that the actual play time would have been cut too short to be usable and comparable; consequently a same single mission was selected for that game for all participants.

4.5 Psychophysiological Measures

The physiological signals were recorded at 1024 Hz sample rate with Varioport-B portable recorder systems (Becker Meditec, Karlsruhe, Germany). Facial EMG from three muscles from the left side corrugator supercilii, orbicularis oculi, and zygomaticus major were measured as per guidelines (Fridlund & Cacioppo, 1986). Pre-processing of the physiological data included rectifying, smoothing, and low-pass filtering following the recommendations by Tassinar, Cacioppo & Vanman (2000). Logarithmic transformation was used to correct the natural skewness in EMG data.

Electrodermal activity, or skin conductance, was measured with 32 Hz sample rate from two electrodes attached to the middle phalanges of the ring and little fingers of the participant's left hand, so that the electrodes would interfere with their gaming as little as possible.

4.6 Self-Reports

After each playing period, the participants filled out self-reports. Their emotional state was measured with both Self-Assessment Manikins (SAM) (Lang, 1980) and as part of the broader Game Experience Questionnaire (GEQ) (IJsselsteijn, De Kort, & Poels, 2007). SAM's are pictorial scales designed to measure emotional dimensions of valence, arousal and dominance quickly by choosing from a range of pictures the one that most closely resembles the current state. Game Experience Questionnaire was specifically designed to measure various aspects of digital gaming experience, and it was developed as part of EU funded FUGA - Fun of Gaming research project, and testing it was part of the original study where the data analysed in this work was collected. GEQ measures several other scales also, but for this work the relevant scales are Negative Affect and Positive Affect, consisting of two items per scale; "I felt bored." and "I found it tiresome" for negative affect, and "I felt content" and "I felt good" for positive aspect, respectively.

4.7 Analysis

For the analysis, all the physiological signal averages and self-report values were standardized to z-values within the participant but over the different playing periods to account for individual differences and reduce auto-correlation that comes from repeated measures structure. To analyse if play-time physiology and self-reports immediately after the playing period are connected when assessing valence and arousal, a bivariate correlation between self-reports and the fEMG and EDA physiological measures of different segments of all playing periods was calculated using R (https://www.r-project.org). The segments used in the analysis were: 1) the whole 30 minute playing period 2) the last 5 minutes 3) the last 2 minutes 4) the last 30 seconds 5) and the last 10 seconds, of each playing period. The rationale for the choice of segments analysed was simply that either it is the grand average of the whole playing period that correlates the most with the self-reports, or that some segment right before self-reporting, and for that several segments of different lengths were analysed. As this not hypothesis testing, but illustrative comparison of different statistical models, familywise alpha correction is not needed, though the number of tests should be taken into account when making interpretations. Confidence intervals were bootstrapped with 2000 samples to better assess the true value of r. Another bivariate correlation was calculated to assess the response coherence of self-reported SAM valence, and GEQig positive and negative affect scales.

5 Results

In general, both the physiological measures and the self-reports individuated the different games from each other quite well (see Figures 1 & 2). Notably, the confidence intervals were calculated on one second averages, and consquently they are very small in the physiological data. If the measures would have given similar results for all four games, they could not be regarded as a valid measure for the gaming experience, and further analysis of their relations had been in vain.

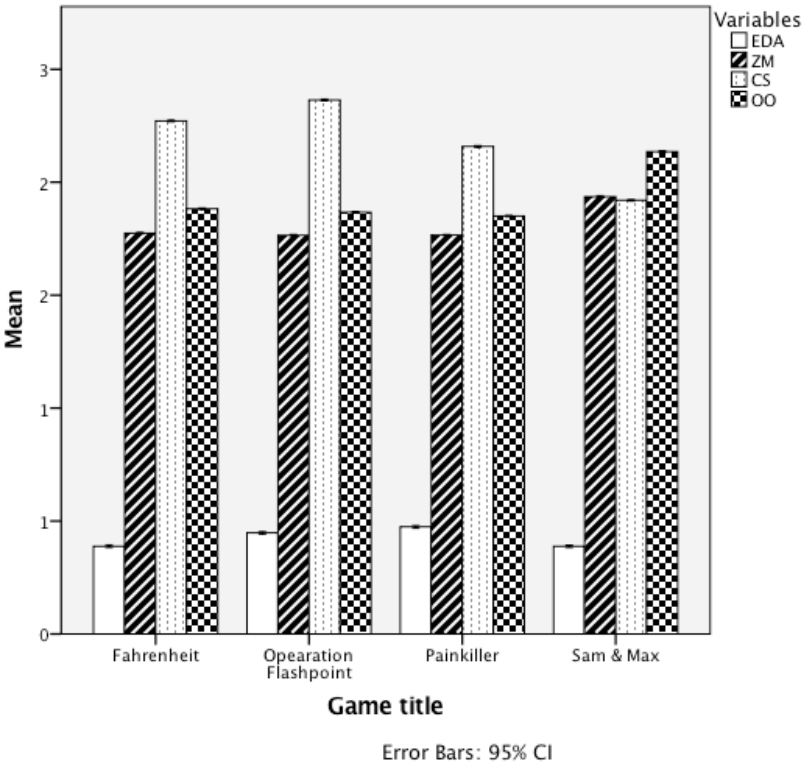

Figure 1. Physiological signal averages per game title.

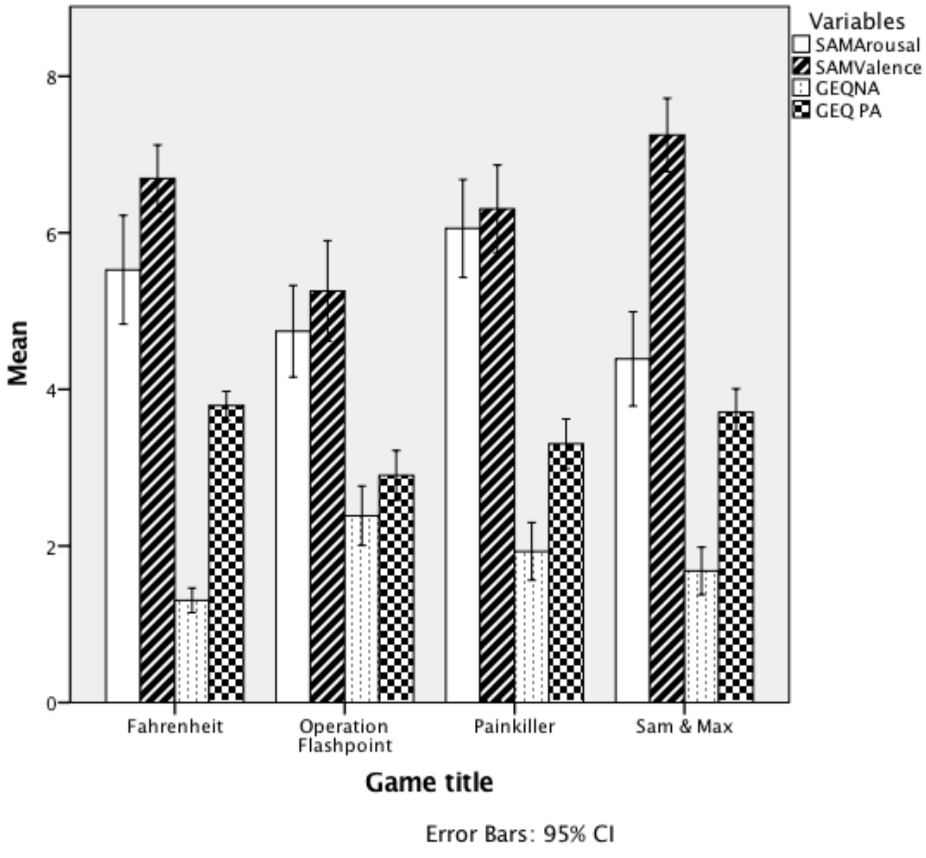

Figure 2. Self-report scores per game title.

When examining the correlation of self-reports and physiological signals, systematic and significant correlations was found (see Tables 1, 2, 3 & 4). Facial EMG ZM and OO were positively correlated with SAM Valence and GEQig Positive Affect, and CS negatively correlated with SAM Valence and positively with GEQig Negative Affect, and these results are in line with the basic assumptions regarding these measures. However, contrary to expectations, EDA did not correlate with self-reported SAM Arousal except occasionally. Overall, when significant, the correlations were rather modest in size, between r = .178 and .406, suggesting a small to medium effect size.

Table 1
SAM valence

			95% CI Bca		
		r	Lower limit	Upper limit	p
30min	ZM	.243**	0.0845	0.3893	.004
	CS	-.406**	-0.5479	-0.2351	.000
	OO	.309**	0.1422	0.4534	.000
5min	ZM	.230**	0.0761	0.3755	.007
	CS	-.315**	-0.4574	-0.1558	.000
	OO	.327**	0.1657	0.4565	.000
2min	ZM	.166	-0.0030	0.3114	.051
	CS	-.309**	-0.4476	-0.1300	.000
	OO	.289**	0.1244	0.4358	.001
30s	ZM	.118	-0.0457	0.2729	.167
	CS	-.301**	-0.4463	-0.1314	.000
	OO	.208*	0.0388	0.3501	.014
10s	ZM	.106	-0.0766	0.2501	.214
	CS	-.305**	-0.4466	-0.1305	.000
	OO	.209*	0.0415	0.3412	.013

* = p < .05, ** = p < .01

Table 2
GEQigNA

		r	95% CI Bca Lower limit	Upper limit	p
30min	CS	.248**	0.0656	0.4054	.004
5min	CS	.130	-0.0516	0.2959	.138
2min	CS	.130	-0.0568	0.3041	.137
30s	CS	.178*	-0.0015	0.3527	.042
10s	CS	.190*	-0.0016	0.3725	.029

* = p < .05, ** = p < .01

Table 3
GEQigPA

		r	95% CI Bca Lower limit	Upper limit	p
30min	ZM	.144	-0.0038	0.2842	.090
	OO	.191*	0.0353	0.3323	.024
5min	ZM	.155	0.0034	0.2972	.069
	OO	.204*	0.0534	0.3406	.016
2min	ZM	.170*	0.0224	0.3228	.046
	OO	.231**	0.0800	0.3819	.006
30s	ZM	.114	-0.0316	0.2640	.182
	OO	.152	-0.0140	0.3014	.074
10s	ZM	.098	-0.0685	0.2536	.252
	OO	.159	-0.0173	0.2970	.062

* = p < .05, ** = p < .01

Table 4
SAM Arousal

		r	95% CI Bca Lower limit	Upper limit	p
30min	EDA	.194*	0.0454	0.3469	.020
5min	EDA	.138	-0.0272	0.2907	.100
2min	EDA	.154	-0.0326	0.3123	.066
30s	EDA	.139	-0.0306	0.2981	.097
10s	EDA	.115	-0.0510	0.2798	.171

* = p < .05, ** = p < .01

When analysing the correlations between different time segments of physiology (the whole 30 minute playing period, the last 5 min, the last 2 min, the last 30s, and the last 10s of physiological activity) and self-reports, quite similar results were found regardless of the segment, but with a slight noticeable trend where longer segments had stronger correlations with self-reports. Self-reported SAM Valence and GEQig Positive and Negative Affect correlated strongly (see Table 5), as is expected from two separate but similar self-report measures, and the correlations with physiology were in the same direction for both self-reports. However, the effect size was in most cases distinctly larger with SAM Valence than GEQig scales regardless of the time segment under scrutiny.

Table 5
Self-report correlations

	GEQig NA		GEQig PA	
	r	p	r	p
SAM Valence	-.692**	.000	.801**	.000

* = p < .05, ** = p < .01

6 Discussion

The results illustrate quite clearly how the self-reported ratings of emotional dimensions and the assessment of those same dimensions utilizing the psychophysiological method do correlate systematically, but not very strongly. Based on these results, it is evident that the two methods assess the same valence and arousal only partially. While the response coherence between physiology and self-reports was mediocre, the response coherence between different self-reports was high, as was predicted by the dual-processing theory (Evers et al., 2014). This data does not allow us to draw inferences regarding whether a dual system theory or time frame processing theory would better explain the lower response coherence than reported in static picture viewing experiments (cf. Lang et al., 1993). The results can be interpreted to support both theories roughly equally; that is, either physiology represents the automatic processing system and self-reports the reflective system, and that theoretical difference explains the lower coherence as predicted by the dual processing hypothesis (Evers et al., 2014), or, the results can be interpreted as the two methods giving us information on different parts of the emotional appraisal process (cf. Kivikangas, 2016) the early on physiological changes and the higher cognitive assessment of the situation. However, as the two theories are fundamentally complementary, and the aim of study was not to assess them as such, it suffices to state that the results obtained here can be meaningfully interpreted in the light of either one – or both – of the theories.

It could be argued based on either theory that the pictorial SAMs that aim to be intuitive, quick, and not language based, could be closer to the physiological signals, while the multi-item GEQig would be definitely require reflective answering. Such an interpretation is supported by the fact that the effect sizes when examining correlation to physiology was considerably larger for SAM Valence than for GEGig PA or NA scales. So, in this data set the GEQig – that was designed specifically for assessing digital games experience – did seem to fare slightly worse than SAMs in this sense. Notably, EDA and SAM arousal correlation was nearly non-existing; a surprising result that merely raises questions and answers none.

In addition, the results do not reveal that the most recent segments of the playing experience would somehow dominate in self-reported emotions; as if that had been the case, the shorter segments from the end would have correlated more with self-reports than physiology form the whole playing period. It can be assumed that more advanced analysis would reveal something regarding the question which part of an extended is most strongly correlated with post-stimulus self-reports, but that is outside the scope of this work.

6.1 Limitations

Even though this work is the first of its kind on an important topic, it is obviously not without its limitations. The results here demonstrate the relation of physiological signals and self-reports, but does not provide any significant insight to why exactly the correlation is rather modest. These results do not help to validate either of the background theories presented, or to make statements about what features of digital games it is exactly that is behind the low response coherence. The study was conducted on only four different games, and while it is a larger number of games than is typical for psychophysiological games research, it is still not sufficient for generalizing the results to all games. While the sample size of the study as such was sufficient for the analysis presented, the participants consisted only of quite limited segment of players; mainly under 30 years old active male gamers. During the time when this data was originally collected, it was immeasurably easier to acquire male participants for game studies, and since then the number of female players have risen considerably. This of course limits the generalizability of the results again a bit more.

The statistical analysis presented here was a simple one. A variety of other analysis could have conducted to test hypothetical scenarios regarding what feature of the physiological data would correlate most with self-reports, but here we tested only if there is a difference how long the measured segment of physiological data is. For example, the maximum of the physiological signal, as to represent the emotional high point of the game, might as well have been the most strongly correlating part of the data. But it was not the point of this work to start exploring the complex nature of physiological signals, or to develop psychophysiological theory or method, but to illustrate what is the current state in this field and to make a statement that we need more advanced theory in the field, and robust systematic testing and validation of the theory in a series of experiments. And for this, a simple correlation analysis was sufficient. So, this work is very much the starting point of a much greater task of finding out the special nature of digital games and forwarding emotion theories, and the psychophysiological method.

6.2 Conclusions

The truly worrying aspect of these results is that they inevitably raise the question that what do the measured physiological signals in reality tell us? The psychophysiological method has been established as a method to assess theoretical constructs such as emotions by showing how those signals and self-repots are to a large degree in line with each other. But if in the case of digital games, they do not correlate to the extent we are accustomed to, how to interpret the physiological measurements?

Delving deeply what type of actions should be taken in forwarding the fundamentals of psychophysiological method and its suitability for studying digital games is beyond the scope of this work, and further effort is needed to even suggest guidelines for the current state of affairs. However, some general suggestions can be brought forward based on these results. First and foremost, when interpreting physiological signals in games research, one should consider explanatory models that take into account the disparity of self-reports and physiological signals. Whether that theory is time scale or dual processing model of emotions does not matter as much as the general idea that in digital gaming context what happens emotionally on physiological level is heavily contextualized so that self-reported emotions can be vastly different. To certain extent they do correlate, but one should not be alarmed if they do not solidly do so. Until further theoretical advances are made, it is advisable to interpret the physiological signals traditionally according to established theory, but add another level of processing on top of it and consider self-reported emotions to be more accurate data on it. Naturally, drawing conclusions from the data will be considerably more challenging as we lose the support for interpretation we gain from response coherence. That increased vagueness and uncertainty can only be battled by extra carefulness when designing and conducting the experiments. On a positive note, by opening up the possibility of separating the two measures, a new window of opportunity is opened for interpretation and perhaps new insights to the fundamental processes of gaming are achieved.

References

Apter, M. J. (1984). Reversal theory and personality: A review. *Journal of Research in Personality*, *18*(3), 265–288. https://doi.org/10.1016/0092-6566(84)90013-8

Association, E. S. (2016). *ESSENTIAL FACTS ABOUT THE COMPUTER AND VIDEO GAME INDUSTRY*.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition & Emotion*, *23*(7), 1284–1306. https://doi.org/10.1080/02699930902985894

Barrett, L. F. (2013). Psychological Construction: The Darwinian Approach to the Science of Emotion. *Emotion Review*, *5*(4), 379–389. https://doi.org/10.1177/1754073913489753

Bradley, M. M. (1994). Emotional Memory: A Dimensional Analysis. In *Emotions Essays on Emotion Theory* (pp. 97–134).

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2000). Psychophysiological science. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 3–26). New York, NY: Cambridge University Press.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of psychophysiology*. New York, NY: Cambridge University Press.

Clore, G. L., & Ortony, A. (2013). Psychological Construction in the OCC Model of Emotion. *Emotion Review*, *5*(4), 335–343. https://doi.org/10.1177/1754073913489751

Costikyan, G. (2002). I have no words & I must design: Toward a critical vocabulary for games. *Computer Games and Digital …*.

Cunningham, W. A., Dunfield, K. A., & Stillman, P. E. (2013). Emotional States from Affective Dynamics. *Emotion Review*, *5*(4), 344–355. https://doi.org/10.1177/1754073913489749

Davidson, R. J., Jackson, D. C., & Larson, C. L. (2000). *Human electroencephalography.* Cambridge University Press.

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* (2nd ed., pp. 200–223). New York, NY: Cambridge University Press.

Deterding, S. (2009). The Game Frame: Systemizing a Goffmanian Approach to Video Game Theory. In *Breaking New Ground: Innovation in Games, Play, Practice and Theory. Proceedings of DiGRA 2009*.

Deterding, S. (2011). Situated motivational affordances of game elements: A conceptual model. In *CHI 2011*. Vancouver, British Columbia, Canada.

Dixon, T. (2012). "Emotion": The History of a Keyword in Crisis. *Emotion Review*, *4*(4), 338–344. https://doi.org/10.1177/1754073912445814

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*.

Ekman, P., & Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, *3*(4), 364–370. https://doi.org/10.1177/1754073911410740

Elliot, A. J., & Covington, M. V. (2001). Approach and Avoidance Motivation. *Educational Psychology Review*, *13*(2), 73–92. https://doi.org/10.1023/A:1009009018235

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evers, C., Hopp, H., Gross, J., & Fischer, A. (2014). Emotion response coherence: A dual-process perspective. *Biological*.

Fridlund, A., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567–589. https://doi.org/10.1111/j.1469-8986.1986.tb00676.x

IJsselsteijn, W. A., De Kort, Y. A. W., & Poels, K. (2007). The Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games. *Manuscript in Preparation*.

Izard, C. E. (2010a). More Meanings and More Questions for the term "Emotion." *Emotion Review*, *2*(4), 383–385. https://doi.org/10.1177/1754073910374670

Izard, C. E. (2010b). The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. *Emotion Review*, *2*(4), 363–370. https://doi.org/10.1177/1754073910374661

James, W. (1884). What is emotion? *Mind*, *9*(34), 188–205.

Järvelä, S., Ekman, I., Kivikangas, J. M., & Ravaja, N. (2013). A practical guide to using digital games as an experiment stimulus. *Transactions of the Digital Games Research Association*.

Kahneman, D. (2013). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kivikangas, J. M. (2015). *Emotion and Social Context in a Digital Game Experience*. University of Helsinki.

Kivikangas, J. M. (2016). Affect Channel Model of Evaluation in the Context of Digital Games (pp. 21–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-41316-7_2

Kivikangas, J. M., Chanel, G., Cowley, B., Ekman, I., Salminen, M., Järvelä, S., & Ravaja, N. (2011). A review of the use of psychophysiological methods in game research. *Journal of Gaming & Virtual Worlds*, *3*(3), 19. https://doi.org/10.1386/jgvw.3.3.181_1

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kühberger, A., Fritz, A., Scherndl, T., Haug, C., & Hoey, J. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PLoS ONE*, *9*(9), e105825. https://doi.org/10.1371/journal.pone.0105825

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Norwood, NJ: Ablex Publishing.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Instruction manual and affective ratings. Technical Report A-8.* https://doi.org/10.1016/j.epsr.2006.03.016

Lang, P. J., Greenwald, M., Bradley, M. M., & Hamm, A. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*, 261–273.

Larsen, J. T., McGraw, A. P., Mellers, B. A., & Cacioppo, J. T. (2004). The agony of victory and thrill of defeat: Mixed emotional reactions to disappointing wins and relieving losses. *Psychological Science*, *15*(5), 325–330. https://doi.org/10.1111/j.0956-7976.2004.00677.x

LeDoux, J. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience*, *23*(1), 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155

LeDoux, J. (2012). Rethinking the Emotional Brain. *Neuron*, *73*(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004

Linderoth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming & Virtual Worlds*, *4*(1), 45–62. https://doi.org/10.1386/jgvw.4.1.45_1

Lindquist, K. A. (2013). Emotions Emerge from More Basic Psychological Ingredients: A Modern Psychological Constructionist Model. *Emotion Review*, *5*(4), 356–368. https://doi.org/10.1177/1754073913489750

Mandryk, R., Inkpen, K., & Calvert, T. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information …*. https://doi.org/10.1080/01449290500331156

McMahan, R., Ragan, E., & Leal, A. (2011). Considerations for the use of commercial video games in controlled experiments. *Entertainment …*.

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. *Cambridge Mass The MIT Press* (Vol. 315).

Montola, M. (2012). *On the Edge of the Magic Circle: Understanding Pervasive Games and Role-Playing*. Tampere University Press.

Norris, C. J., Gollan, J., Berntson, G. G., & Cacioppo, J. T. (2010). The current status of research on the structure of evaluative space. *Biological Psychology*, *84*(3), 422–436. https://doi.org/10.1016/j.biopsycho.2010.03.011

Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, *5*(3), 407–422. https://doi.org/10.1017/S0140525X00012759

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. https://doi.org/10.1016/j.concog.2004.10.004

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–34. https://doi.org/10.1017/S0954579405050340

Ravaja, N. (2004). Contributions of psychophysiology to media research: review and recommendations. *Media Psychology*, *6*(2), 193–235. https://doi.org/10.1207/s1532785xmep0602_4

Roseman, I. J. (2013). Appraisal in the Emotion System: Coherence in Strategies for Coping. *Emotion Review*, *5*(2), 141–149. https://doi.org/10.1177/1754073912469591

Russell, J. A. (2014). Four Perspectives on the Psychology of Emotion: An Introduction. *Emotion Review*, *6*(4), 291–291. https://doi.org/10.1177/1754073914534558

Russell, J. A., & A., J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Schacter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379–399. https://doi.org/10.1037/h0021465

Scherer, K. R. (1999). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 637–663). New York, New York, USA: John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch30

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Smith, C. A., & Lazarus, R. S. (1990). *Emotion and adaptation.* Guilford Press.

Stenros, J. (2015). *Playfulness, Play, and Games: A Constructionist Ludology Approach*. Tampere University Press.

Suits, B. (1967). What is a game? *Philosophy of Science*, *34*(2), 148–156. https://doi.org/10.1086/288138

Tassinary, L. G., Cacioppo, J. T., & Vanman, E. J. (2000). The skeletomotor system: surface electromyography. *Handbook of Psychophysiology*, 267–299. https://doi.org/http://dx.doi.org/10.1017/CBO9780511546396

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the Dimensional and Hierarchical Structure of Affect. *Psychological Science*, *10*(4), 297–303. https://doi.org/10.1111/1467-9280.00157

Wundt, W. (1897). *Outlines of Psychology.*

Yannakakis, G. N., Martinez, H. P., & Garbarino, M. (2016). Psychophysiology in Games (pp. 119–137). https://doi.org/10.1007/978-3-319-41316-7_7

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-Point Circumplex Structure of Core Affect. *Emotion*, *11*(4), 705–31. https://doi.org/10.1037/a0023980