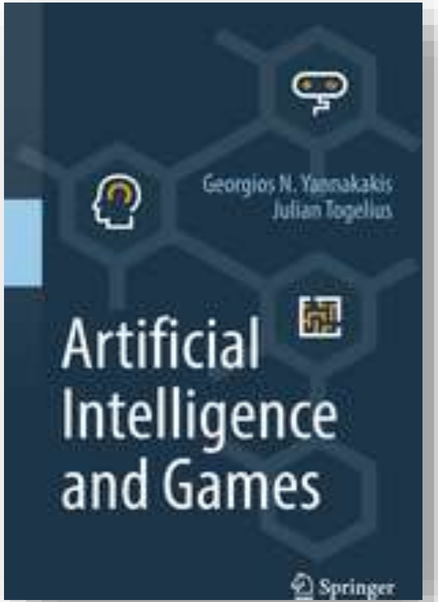
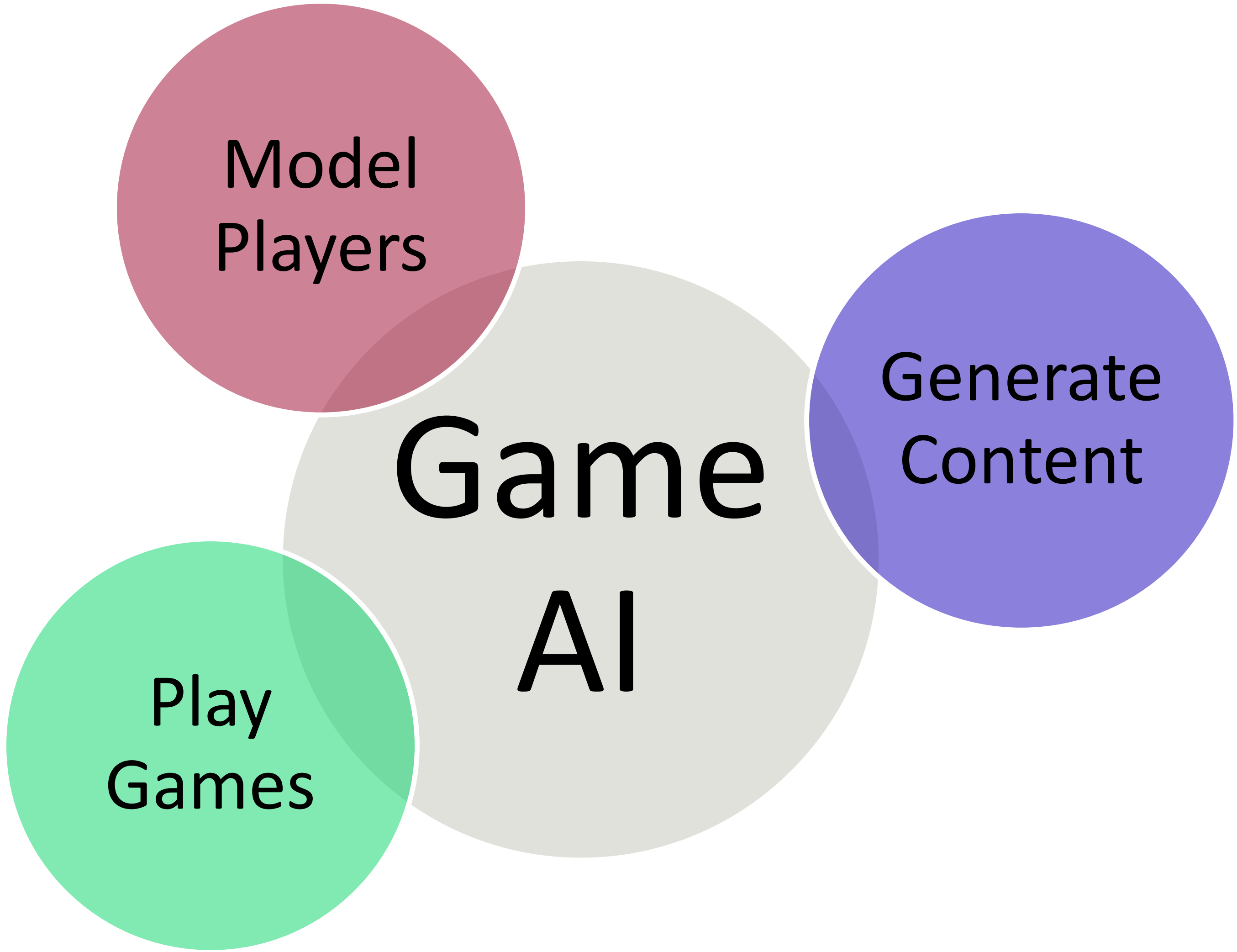
Artificial Intelligence and Games
Modeling Players

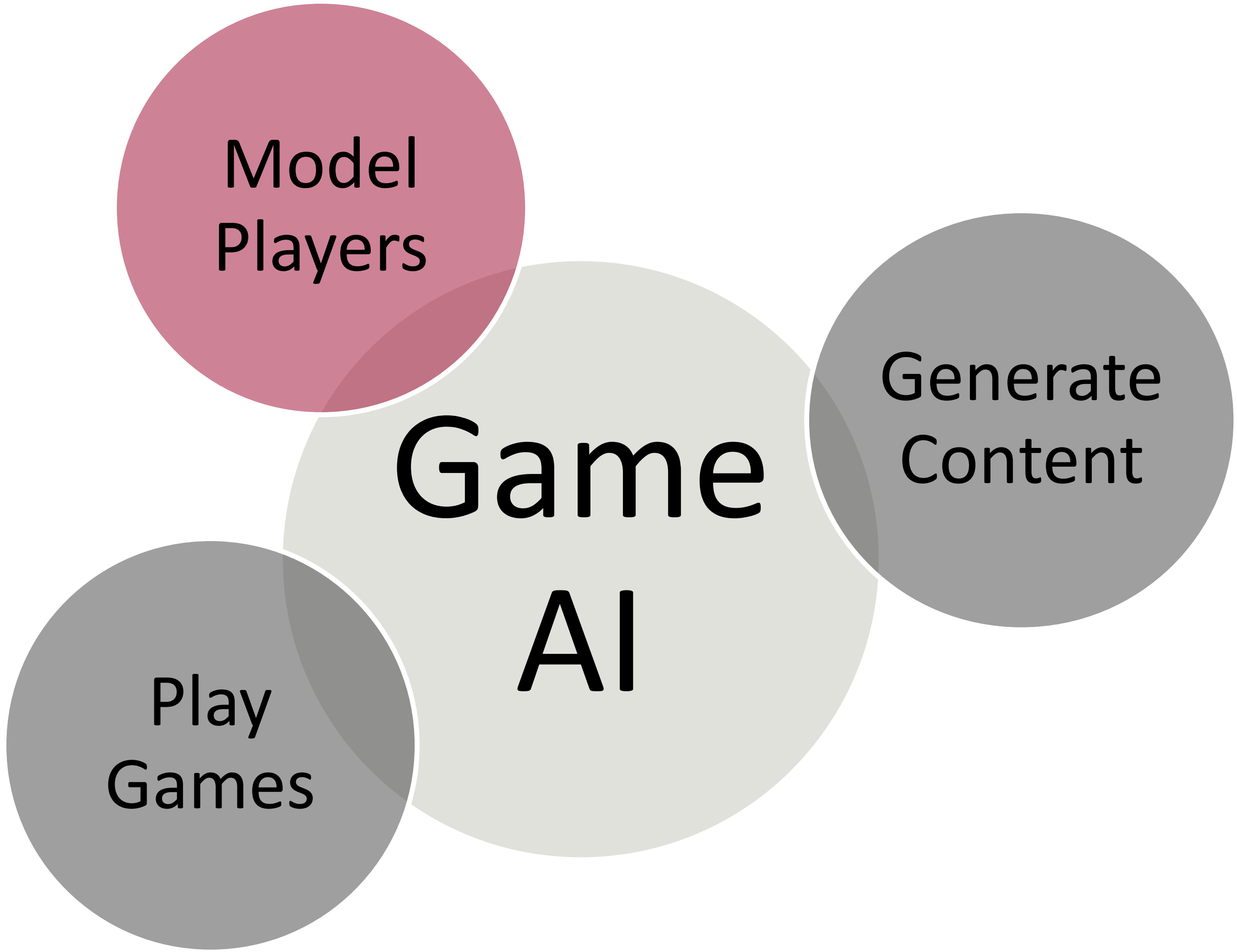
Georgios N. Yannakakis
@yannakakis

Julian Togelius
@togelius

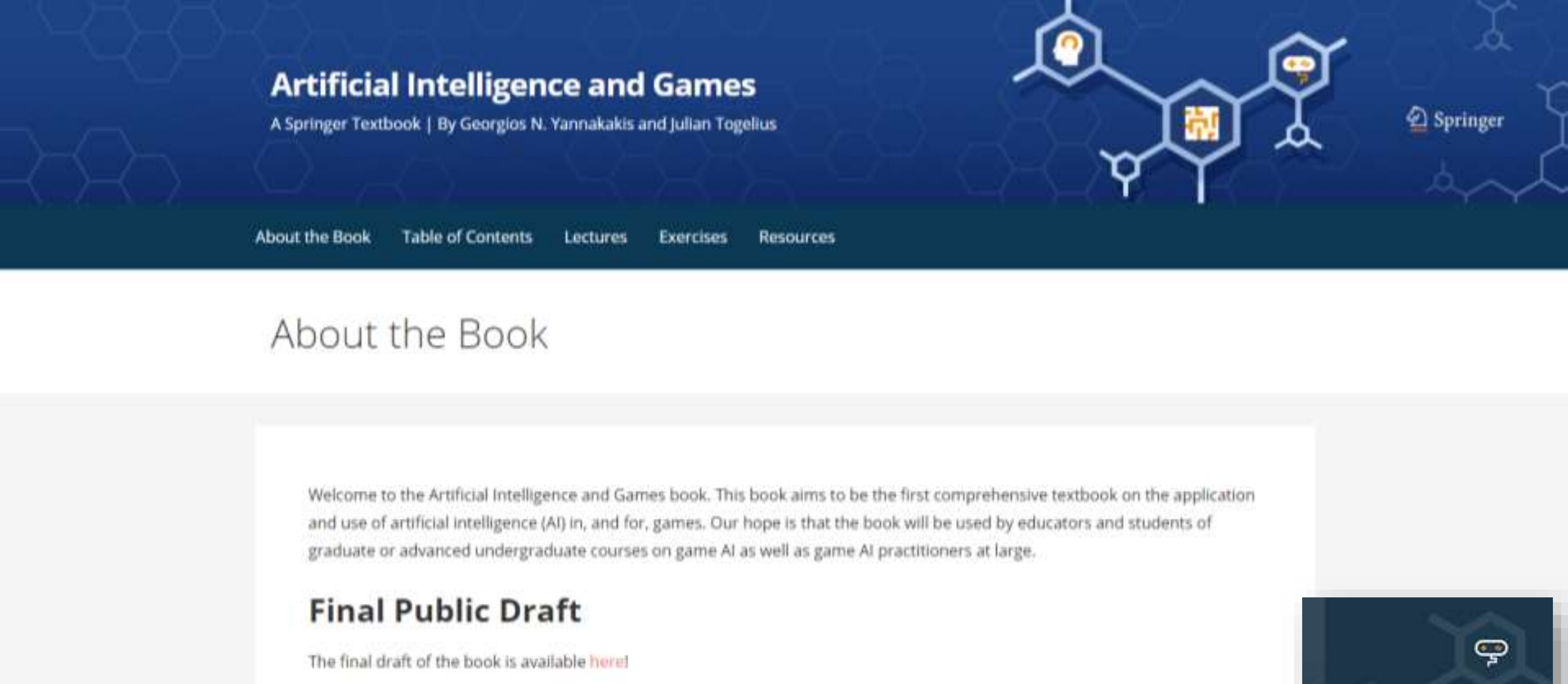




G. N. Yannakakis and J. Togelius, “**Artificial Intelligence and Games**,” Springer Nature, 2018.

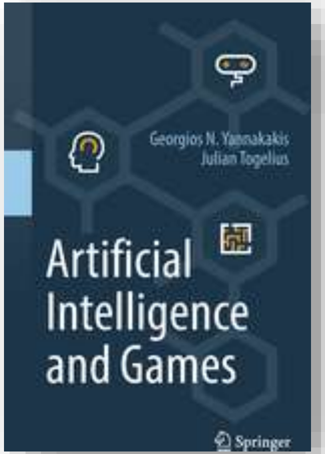


G. N. Yannakakis and J. Togelius, "**Artificial Intelligence and Games**," Springer, 2018.



Your readings from **gameaibook.org**

Chapter: 5



Overview

- Player Modeling
  - Why?
  - How?
- Main tasks for AI/Machine Learning
- Holistic view
  - From Affective Computing to Game Studies to Game Analytics
- Examples
  - Player Behavior
  - Player Experience

What is Player Modeling?

Player Modeling

- Player Modeling
  - Non human players
  - **Human players**
- Player Modeling vs Opponent Modeling
- Player Modeling vs Player Profiling
  - Modeling: complex dynamic phenomenon
  - Categorization of players based on static information

Player Modeling

The study of *computational* means for the modeling of a player's **experience** or **behavior** which is based on *theoretical frameworks* about player experience and/or *data* derived from the interaction of the player with a game

Why Model Players?

Goals and Aims

Goal

"…*understand how the interaction with a game is experienced by players*."

Why use AI for Player modelling?

- Understanding player experience
- Understanding player behavior

Why Model Players?

- Why not?
- Machines understand numbers
- Player Experience is the holy grail for design and development
- But most importantly because…

Why Model Players?

- The perfect game is tailored to you!
- We are different (and many more than before)
- If you learn to play…. it is only fair that the game **learns you**

Core Player Modeling Tasks

Supervised/Reinforcement Learning

Imitation

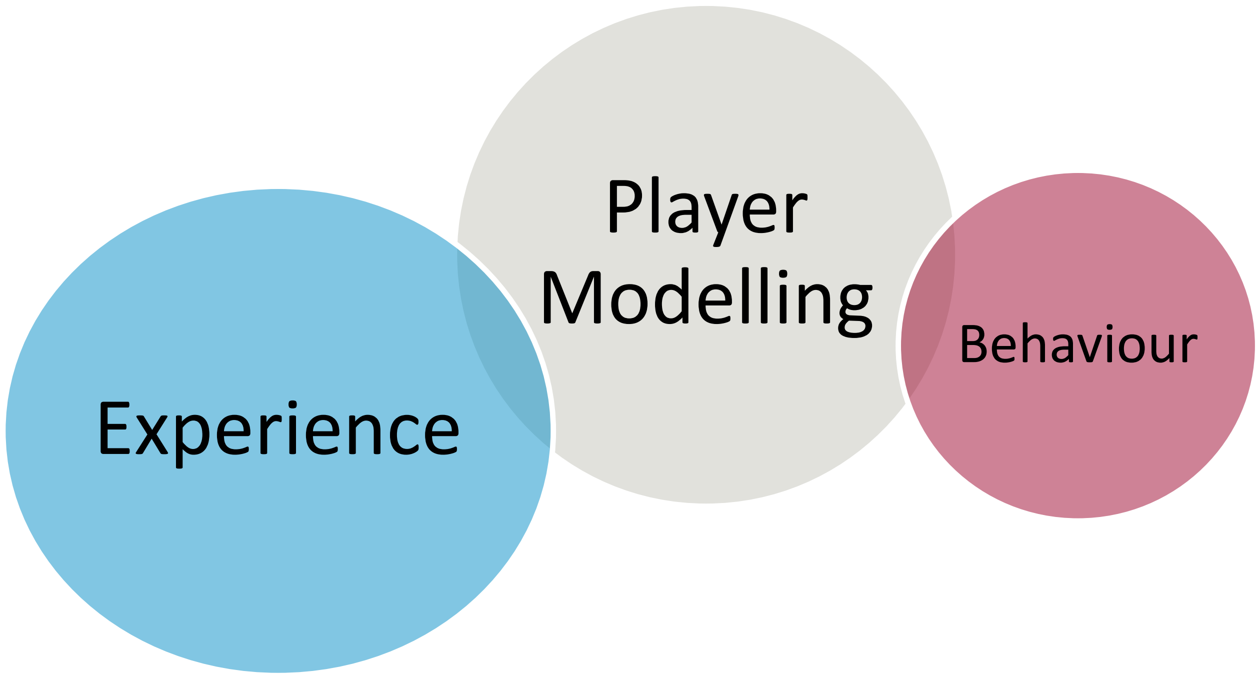
Prediction

Unsupervised Learning

Clustering

Association mining

Player Modeling Examples

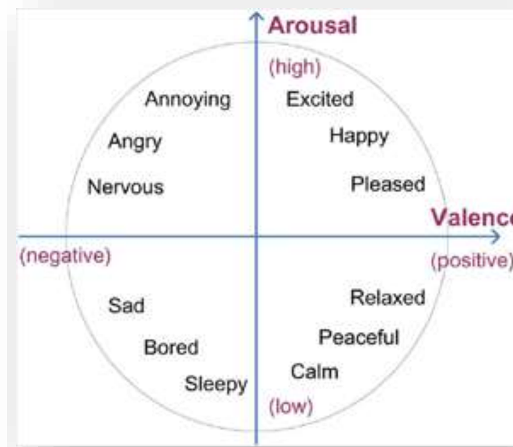


G. N. Yannakakis and J. Togelius, “**Artificial Intelligence and Games**,” Springer Nature, 2018.

Player Experience vs Player Behavior

Experience: how you **feel during play**

- A set (a synthesis) of affective, cognitive and behavioral states
- Or else *user states*
- Emotions: Appraisal theory, …
- Cognition/Behavior: several models (e.g. BDI,…)



Behavior: what you **do during play**

A High-Level Taxonomy


Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, …)
Model Free [Bottom-Up]
(Data Science, Machine Learning)

Player Model

Model-Based [Top-Down]

(Psychology, Cognitive Science, Game Studies, …)

Model-Based

Psychology and Affective Sciences

Psychology and Affective Sciences

"**Affective computing** is the study and development of systems and devices that can recognize, interpret, process, and simulate human **affect**" (Wiki)

Term coined by R. Picard (1999)

Emotion

Emotion is a subjective, conscious experience characterized primarily by psychophysiological expressions, biological reactions, and mental states. (Wikipedia)

NB: Emotion (and affect) is a **construct** with **fuzzy** boundaries!

Cognitive Appraisal

Cognitive appraisal

You see a shark!

Physiological Manifestation(s)

High adrenaline, increased HR, pupil dilation, increased skin conductivity

Action(s)

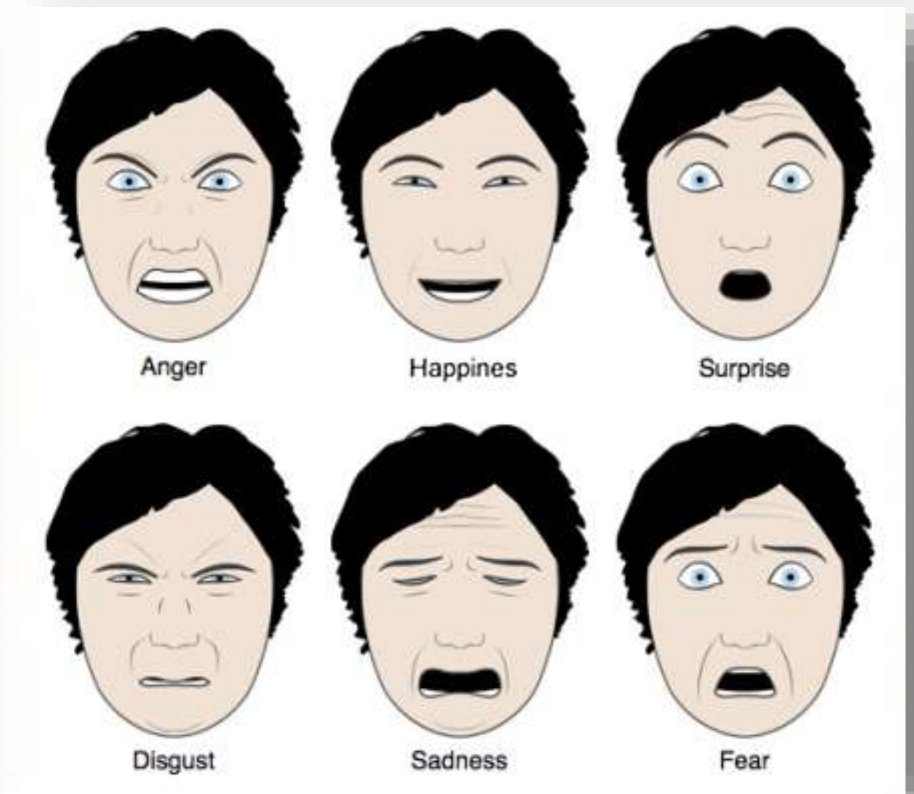
You scream, swim away, fight (?!)

Emotion Representation – Eckman

Emotions

- Are **discrete** (states)
- Are measurable
- Can be identified through their physiological or bodily manifestations

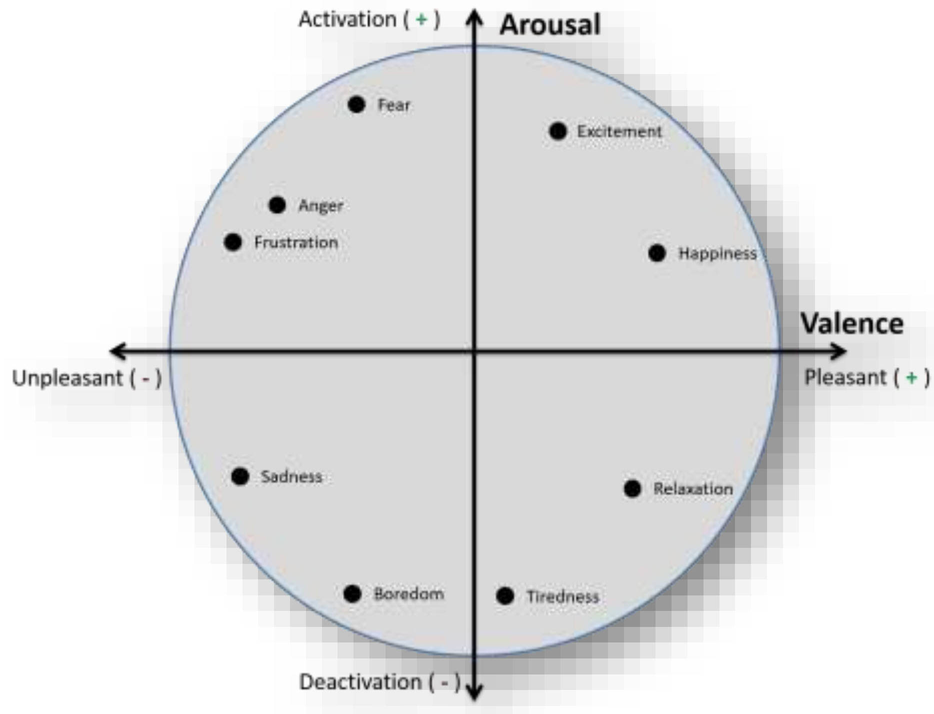
The 6 basic emotions are culture-independent



Emotion Representation – Russell

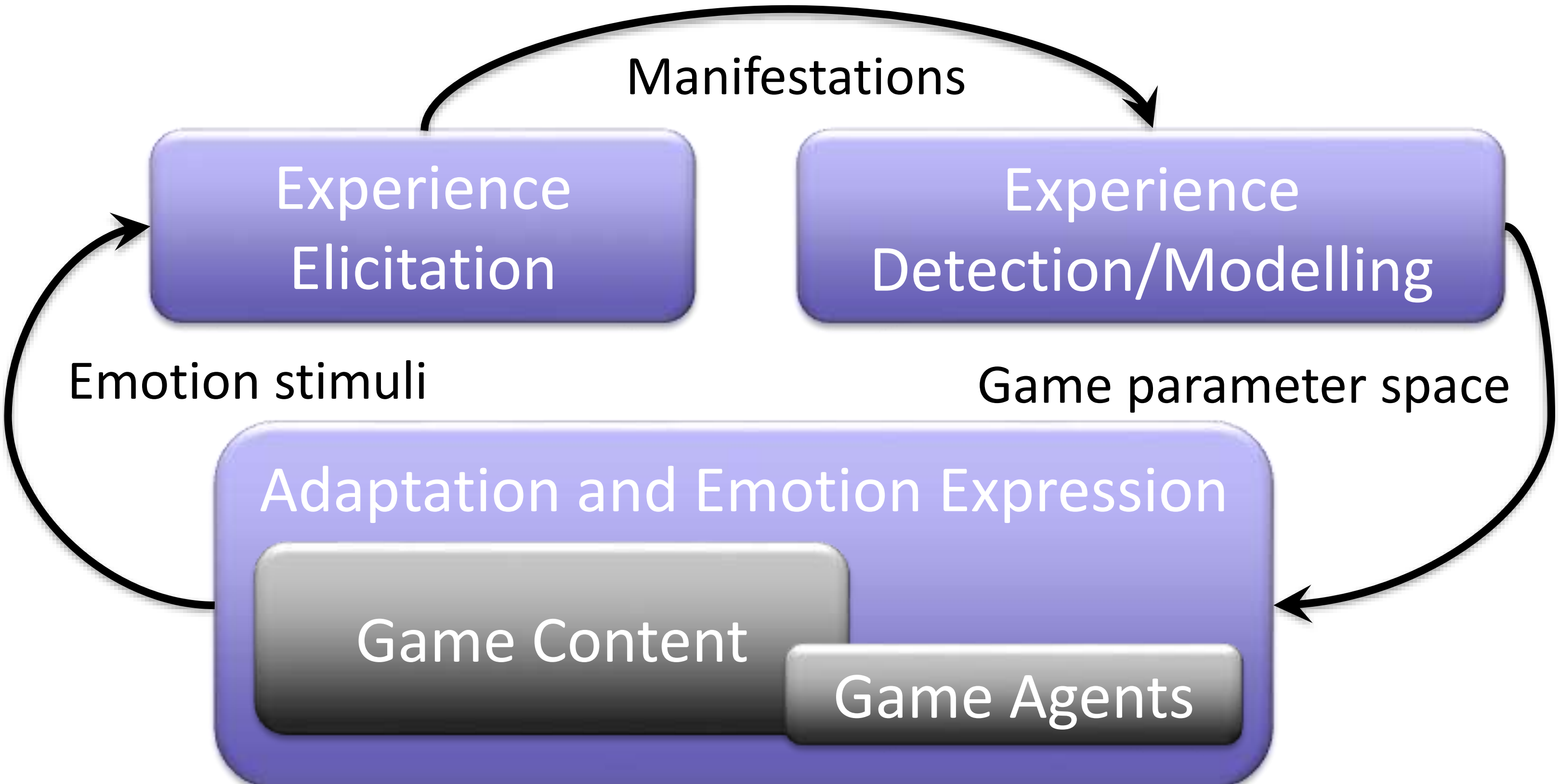
Emotions

- Are **scalar values**
- Represented in a 2-d plane: **arousal** vs. **valence**
- Are measurable
- Can be identified through physiological or bodily manifestations



Games Best Realizes the Affective Loop

Yannakakis and Paiva, **Emotion in Games**, in *Handbook of Affective Computing*, 2013

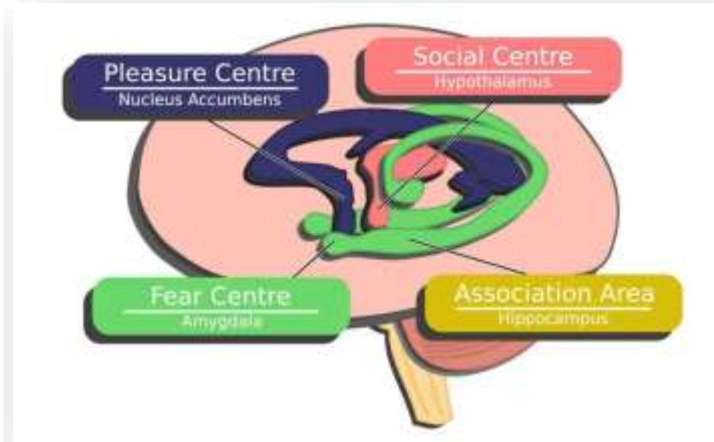


Neuroscience

Neurobiology of Play

Neuroscience – via fMRI

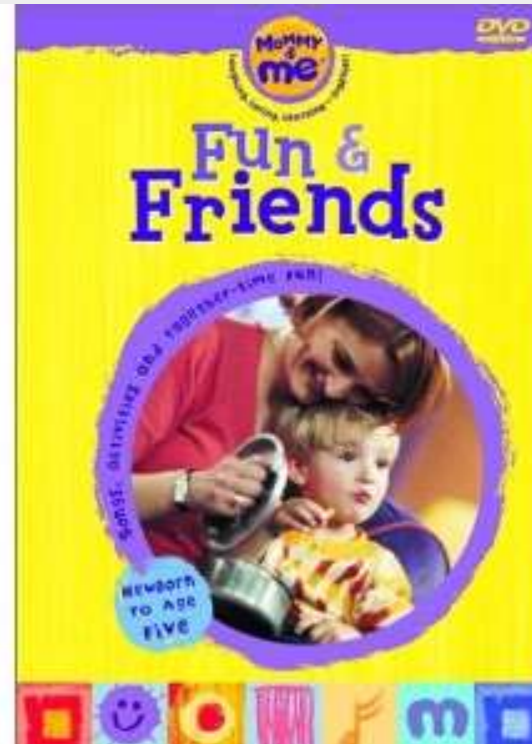
- Interest (or curiosity): **visual cortex** (hippocampus) and **endomorphin.**
- Explorer (Barter) / Easy Fun (Lazzaro) / Curiosity (Malone) → Wonder: **endomorphin (pleasure center)**
- Achiever (Bartle) → Achievement: **dopamine (pleasure center).**
- Hardcore gamers / Hard Fun (Lazzaro) → Fiero: **dopamine (pleasure center)**
- Serious Fun (Lazzaro) → Excitement: **adrenaline (fear center)**
- Socialiser (Bartle) / People Fun (Lazzaro) → Enjoyment: **oxytocin (social center) and dopamine (pleasure center)**



Bateman, Chris, and Lennart E. Nacke. "**The neurobiology of play**" *Proceedings of the International Academic Conference on the Future of Game Design and Technology*. ACM, 2010.

“Fun”

- “Fun” and friends
  - Learnability
  - The right difficulty level
  - Challenge – Curiosity – Fantasy
  - Hard Fun / Easy Fun / **X** Fun / ...
  - Adrenalin / Dopamine
  - Arousal / Positive Emotions
  - Flow
  - ...

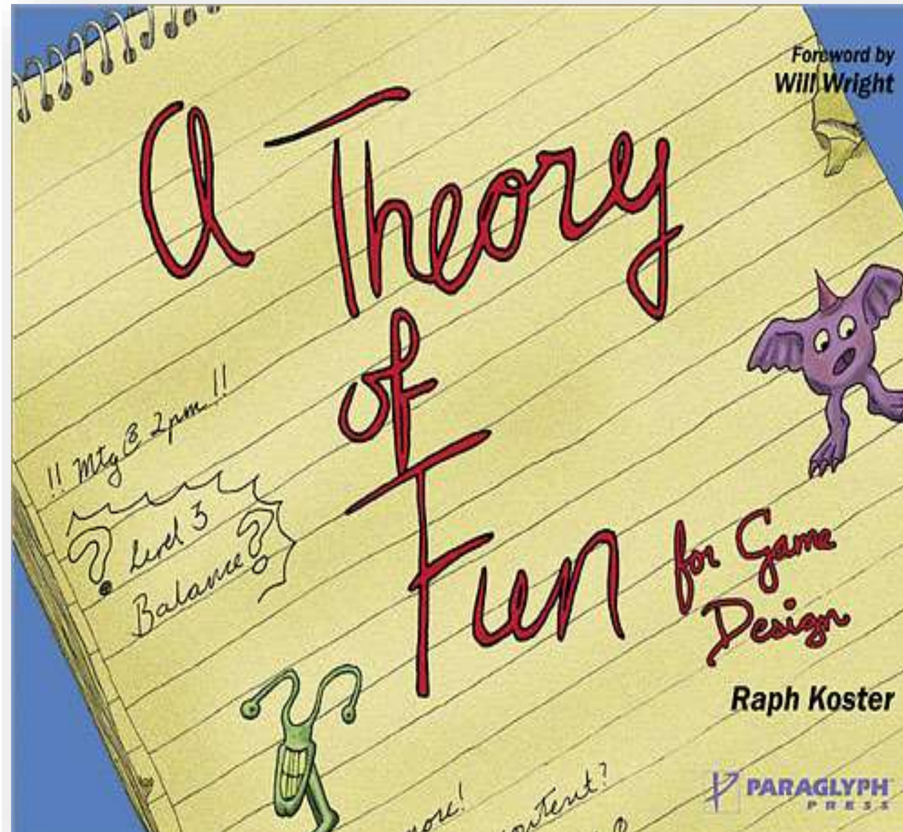


A Theory of Fun (Koster, 2005)

Fun = Learning

Complex training signals / patterns → Learning Difficulty → Frustration → No Fun!

Seen (and mastered) training signals / patterns → Boredom → No Fun!

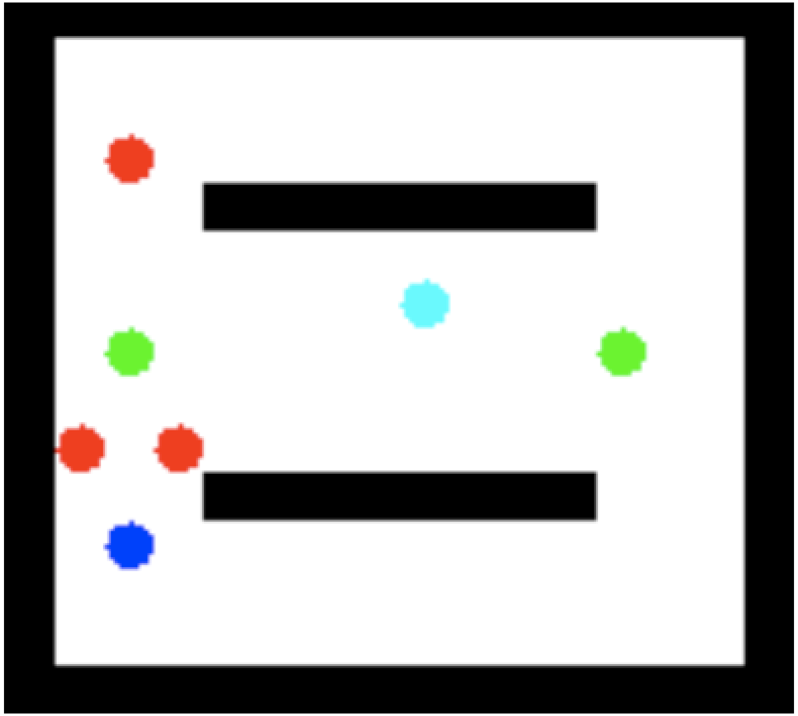


Evolving Game Rules Based on Koster

Fun of an AI agent

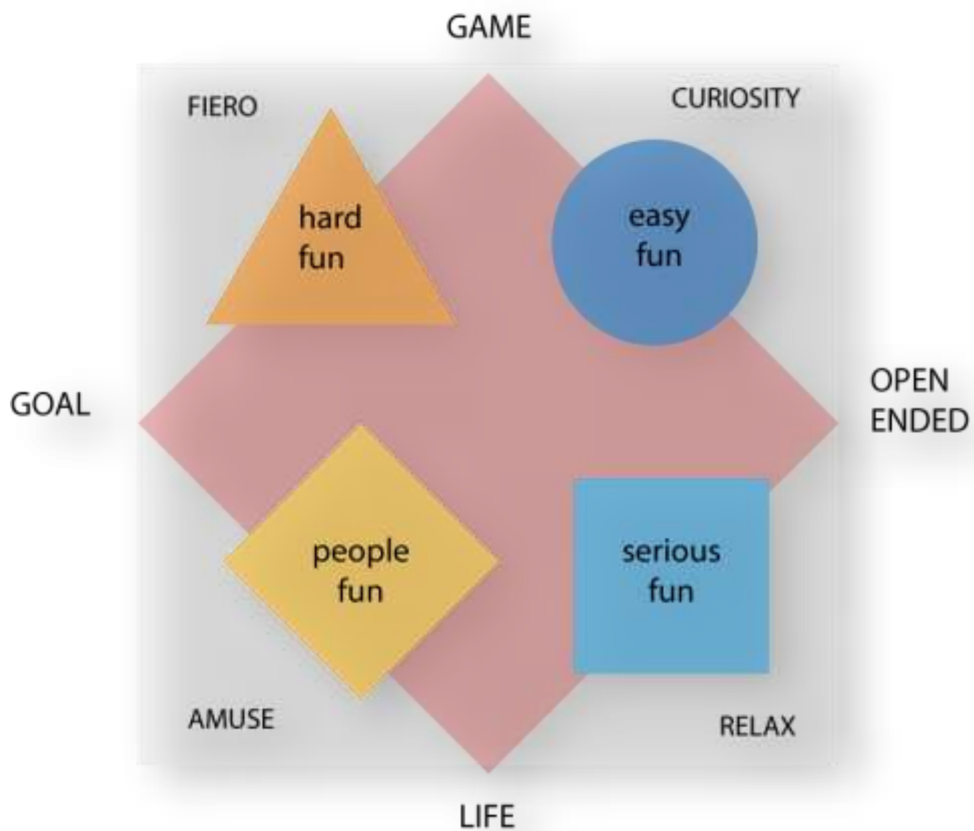
=

Learnability during the game (dynamic “fun” fitness function)



Togelius, J., & Schmidhuber, J. (2008, December). **An experiment in automatic game design**. In *Computational Intelligence and Games, 2008. CIG'08. IEEE Symposium On* (pp. 111-118). IEEE.

Lazzaro’s 4 Fun-factor Model

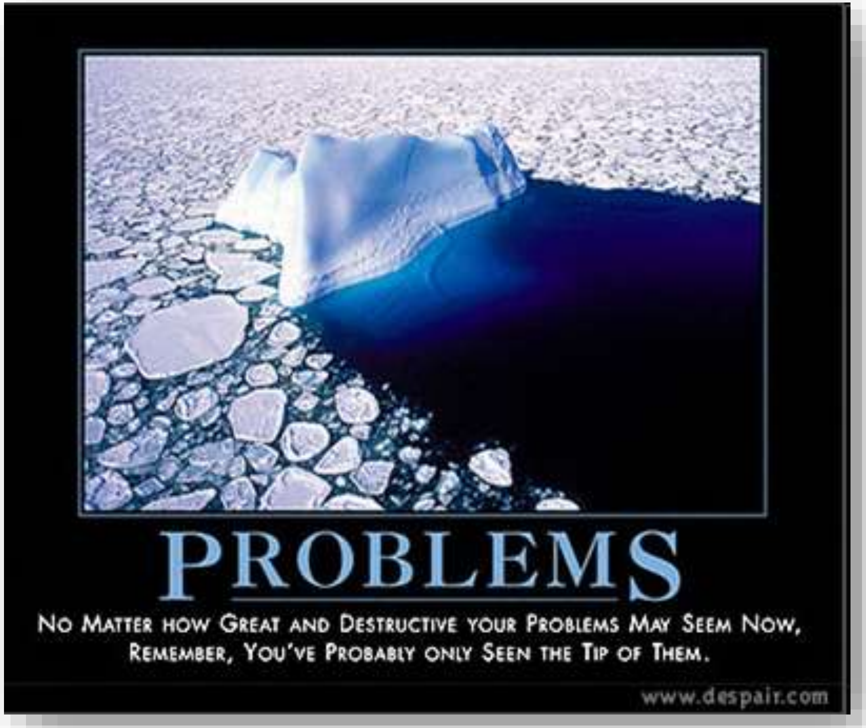


(c) Nicole Lazzaro, www.xeodesign.com

Problems with "Fun"

Way too many

- Complex – composite of other things…
  - Engagement, interest, immersion, …? Who knows?
- Quasi/Fake Emotion (not real!)
- Nowhere in any model of emotion
- No device can measure "fun"

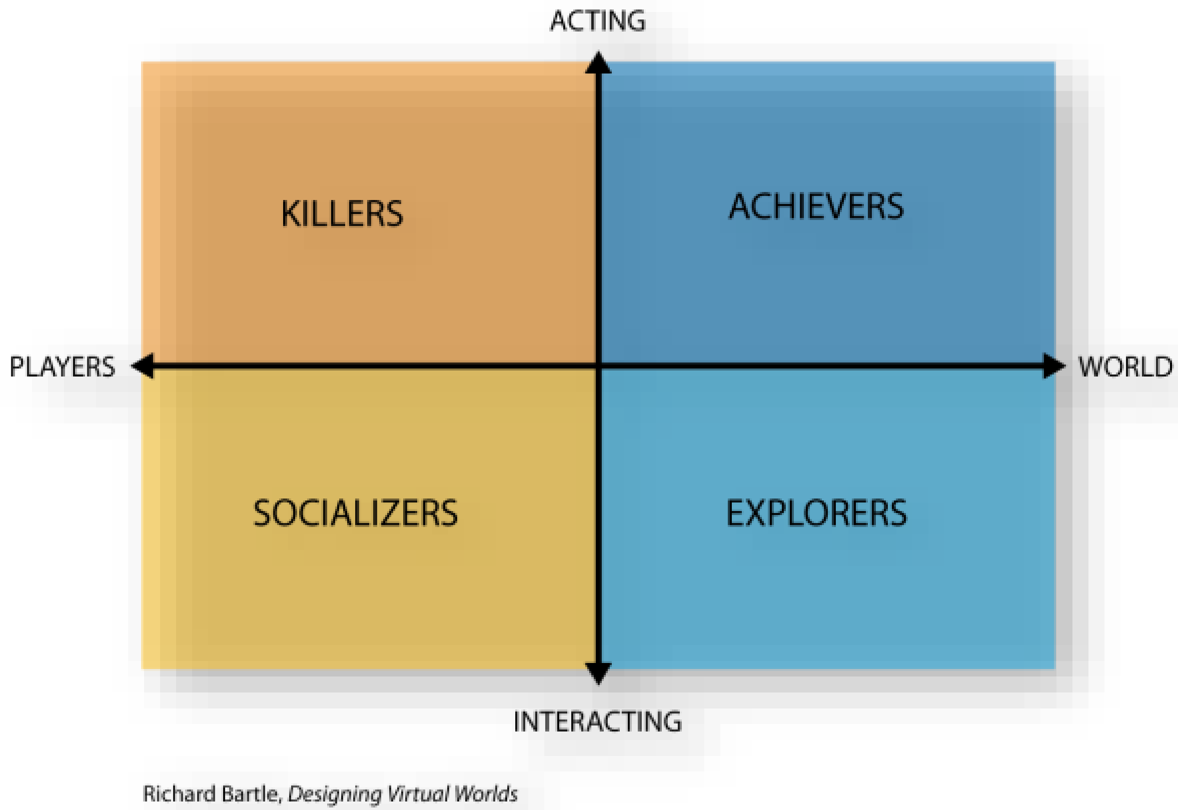


But

- People can still express it…
- Ask 4-yo children if they are immersed….

Player Archetypes

Bartle, Richard A. **Designing virtual worlds**. New Riders, 2004.



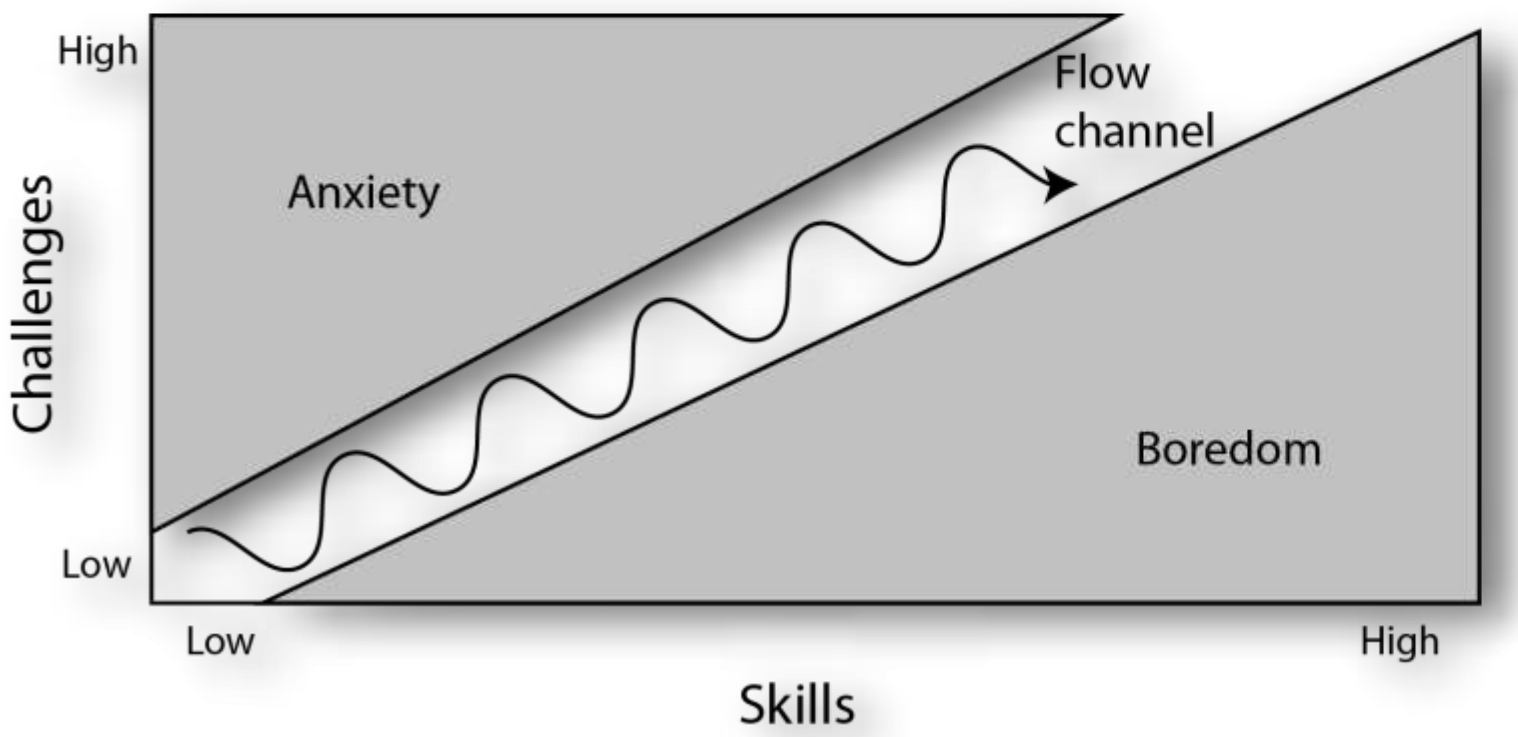
Flow (Csikszentmihalyi, 1975)

Flow is the state of “happiness”

When in *Flow*:

1. focused concentration on the present moment
2. loss of reflective self-consciousness
3. sense of personal control or agency over the situation or activity
4. a distortion of temporal experience (subjective experience of time is altered)
5. experience of the activity as intrinsically rewarding

Flow (Csikszentmihalyi, 1975)



Malone (1981) : Challenge, Curiosity and Fantasy

Challenge

- Uncertainty of achieving a goal due to e.g. variable difficulty level, multiple level goals, hidden information, and randomness.

Fantasy

- Show (or evoke) images/situations/context not present
  - Extrinsic: depends on the skill used in a game
  - Intrinsic: related to the use of that skill.

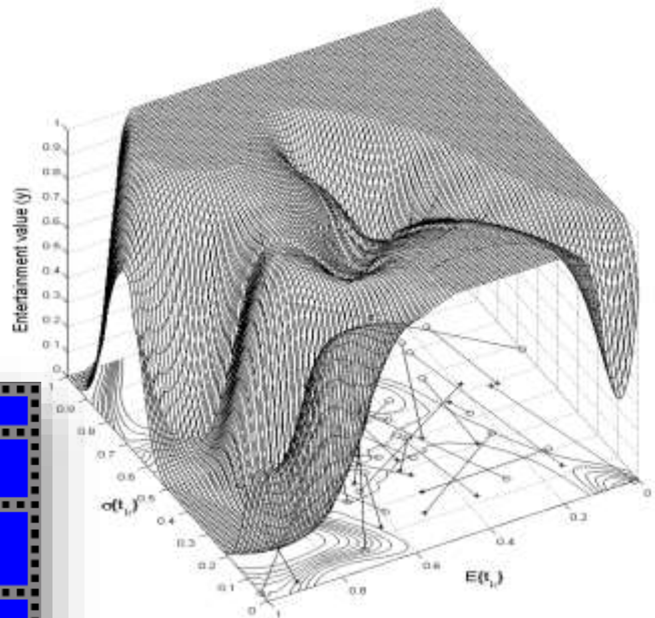
Curiosity

- What will happen next to the game?
  - Sensory curiosity
  - Cognitive curiosity – it can be aroused via incomplete/inconsistent knowledge structures

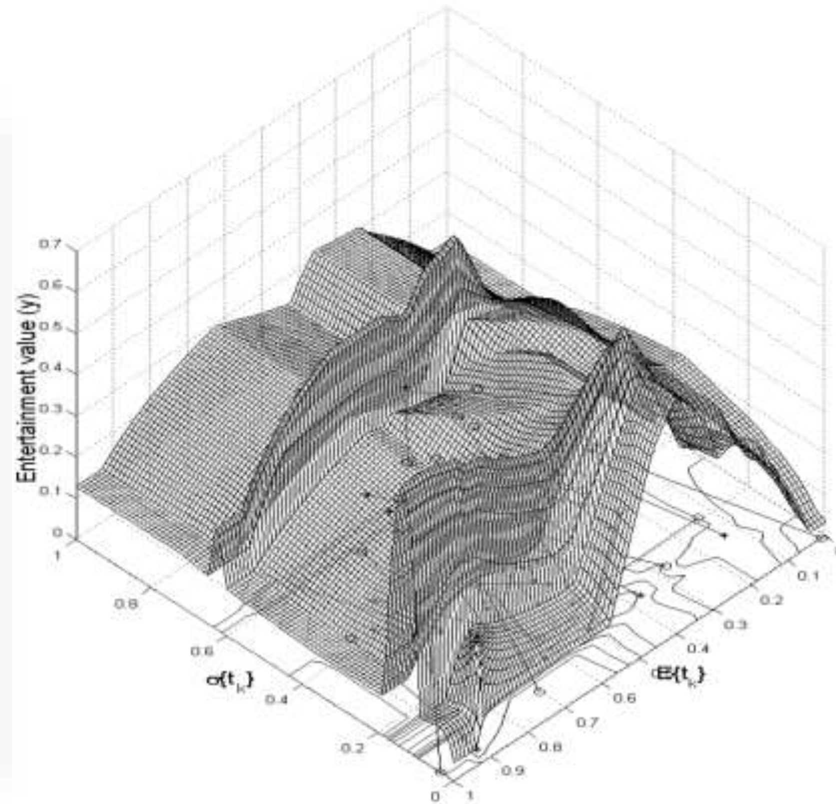
T. Malone, **Toward a Theory of Intrinsically Motivating Instruction**, Cognitive Science, 1981

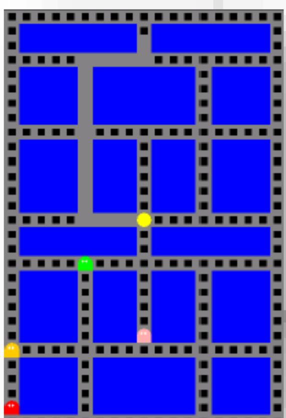
Entertainment Modelling based on Malone

Prey/Predator Games



(a) The fittest feedforward NN solution ($f = 29.95$).





G. N. Yannakakis, and J. Hallam, “**Modeling and Augmenting Game Entertainment through Challenge and Curiosity**,” *International Journal on Artificial Intelligence Tools,* vol. 16, issue 6, pp. 981-999, December 2007.

Entertainment Modelling based on Malone

Playware Playground

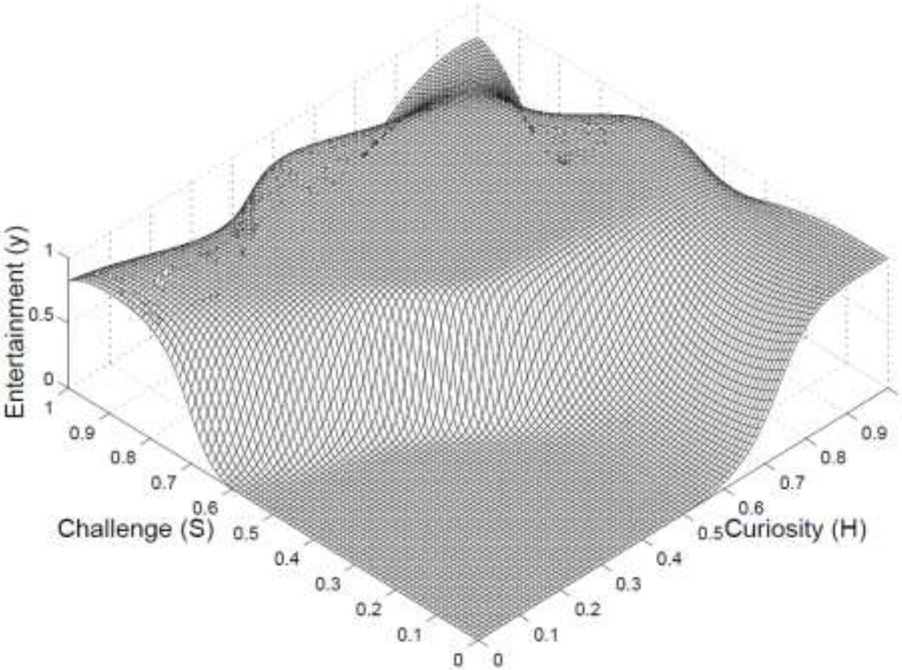
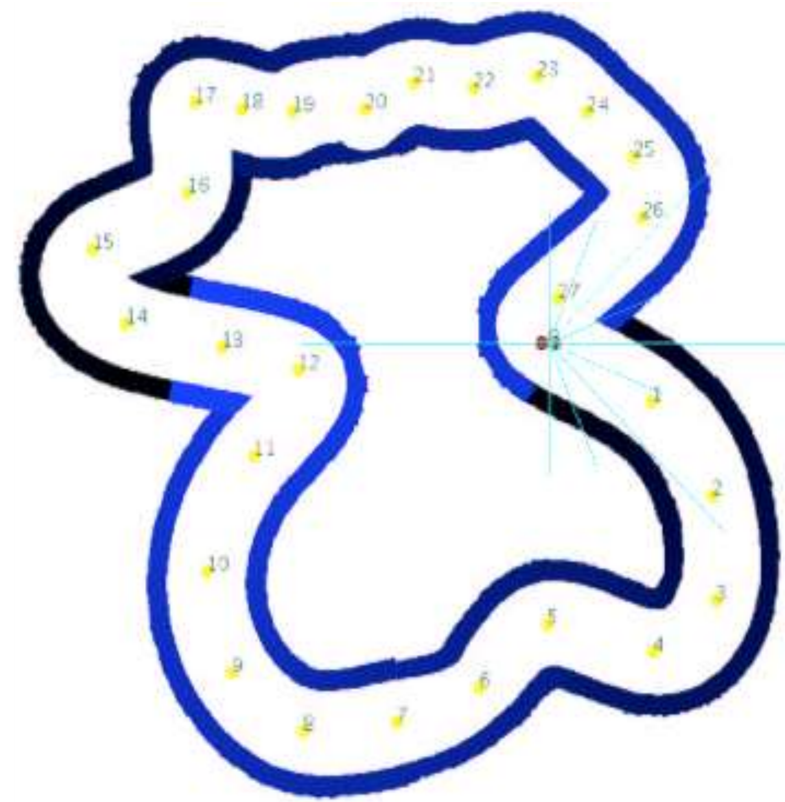


Fig. 4. Fittest feedforward NN ($f = 22.82$).

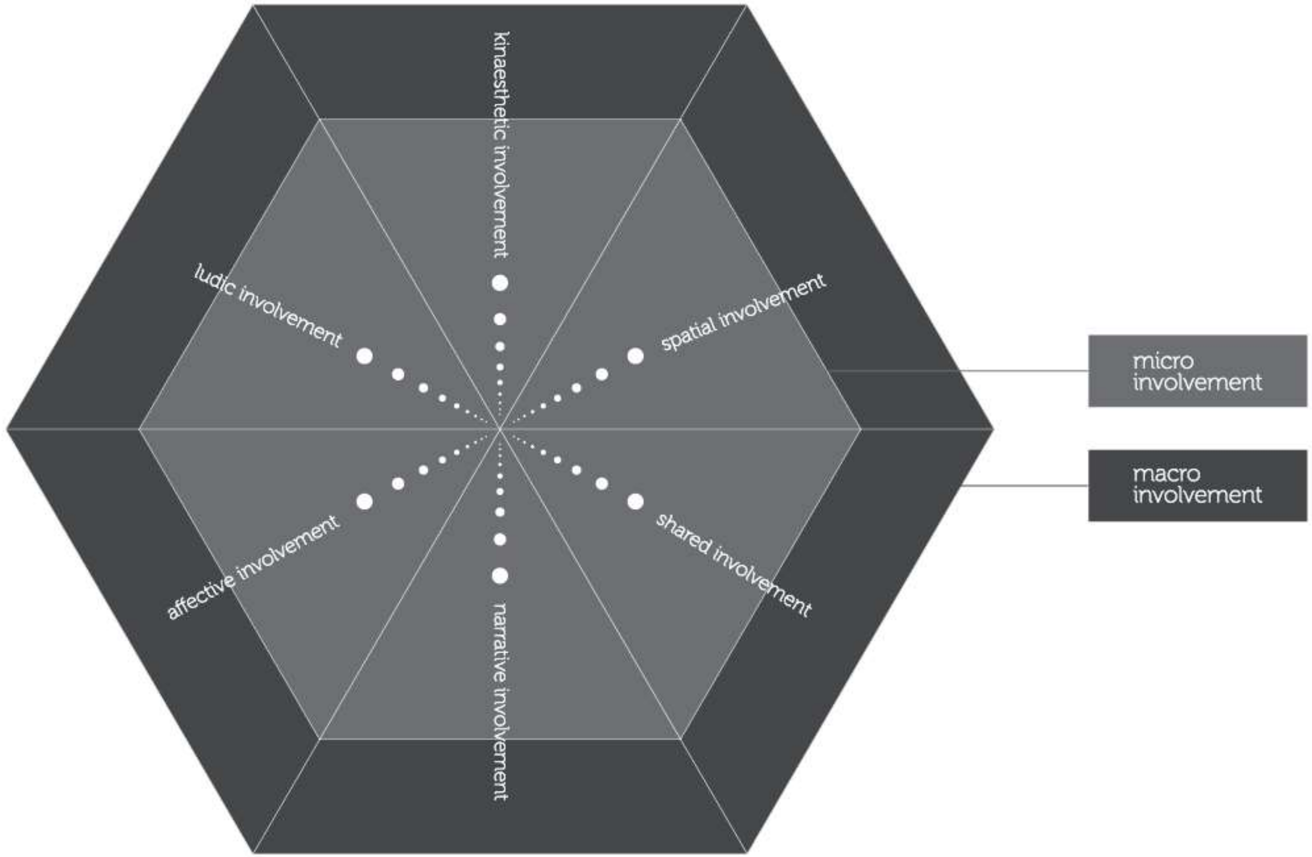
G. N. Yannakakis, and J. Hallam, "**Modeling and Augmenting Game Entertainment through Challenge and Curiosity**," *International Journal on Artificial Intelligence Tools,* vol. 16, issue 6, pp. 981-999, December 2007.

Evolving Racing Tracks Based on Malone



Togelius, J., De Nardi, R., & Lucas, S. M. (2007). **Towards automatic personalised content creation for racing games**. In *Computational Intelligence and Games, 2007. CIG 2007. IEEE Symposium on* (pp. 252-259). IEEE.

Player Incorporation Model (Calleja, 2011)



Gordon Calleja. **In-game: from immersion to incorporation**. MIT Press, 2011.

Summary of Top-Down Player Modeling

One Theory After all?

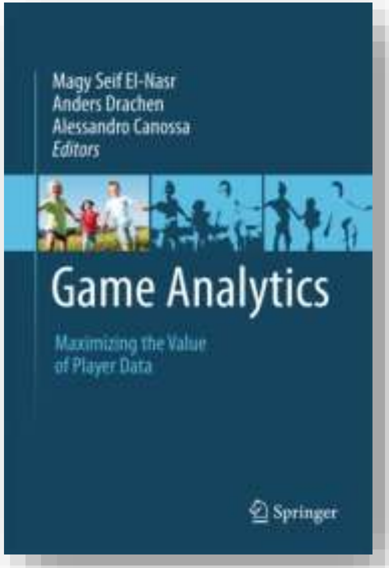
Some Examples

- Explorer (Barter) / Easy Fun (Lazzaro) / Curiosity (Malone)
- Achiever (Bartle) / Serious Fun (Lazzaro)
- Killer (Bartle) / Hard Fun (Lazzaro) / Challenge (Malone) / Flow aspects
- Socialiser (Bartle) / People Fun (Lazzaro) / Shared Involvement (Calleja)

Player Model

Model Free [Bottom-Up]
(Data Science, Machine Learning)

Model-free: From Theory to Data



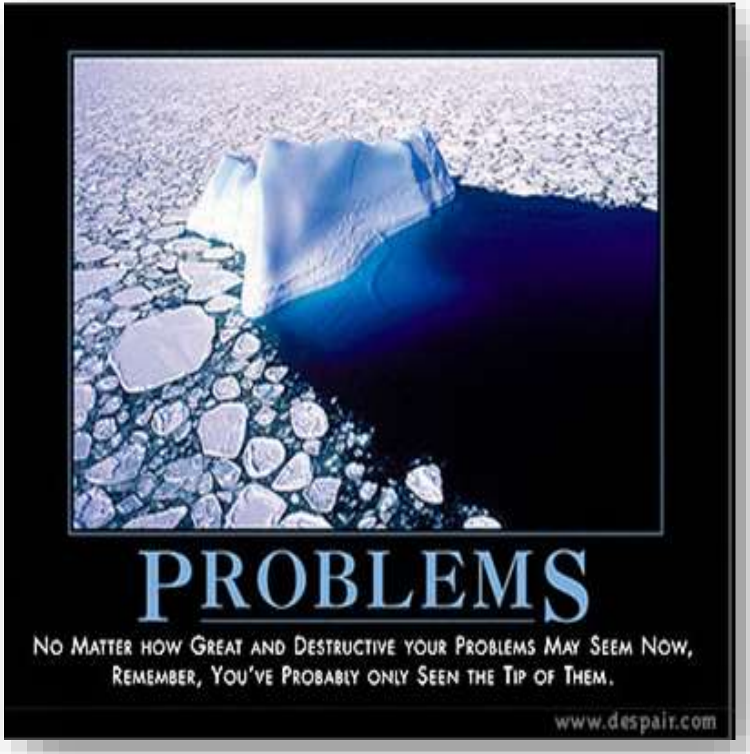
Game Analytics

Basic Definitions

- **Game analytics:** application of analytics to game development and research.
- **Goal:** support decision making, at operational, tactical and strategic levels for design, art, programming, marketing, user research, etc.
- **Game metrics:** interpretable measures of something related to games
  - User/Player metrics
    - User: customer, gameplay community
    - Gameplay: interface, in-game
  - System metrics

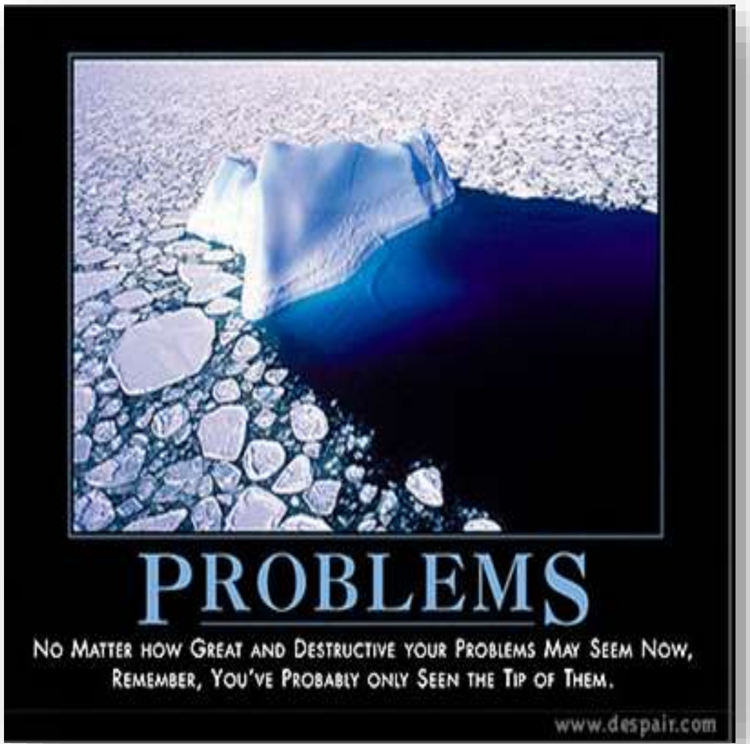
What's the Problem?

- Understanding players
  - Behaviour
  - Player Experience
- How do people play a game?
- Do they play the game as intended?



What's the Problem?

- Debugging costs!
  - Code debugging
  - Experience debugging
- Play testing costs!
- Non-linear games
  - Testing is challenge
- Improve next game release costs!
- …



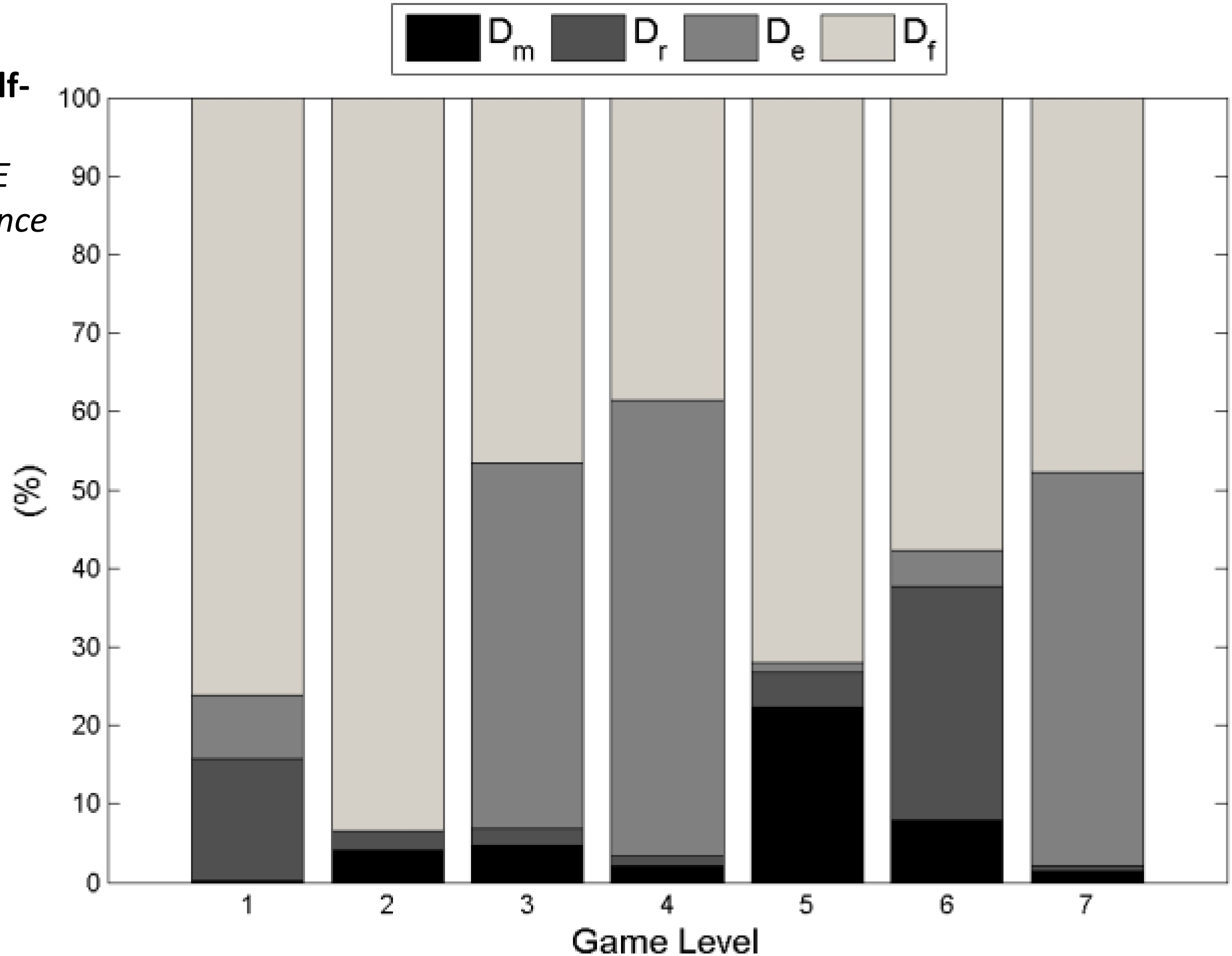
Game Data Visualization

Descriptive Statistics in Games

- Very important to visualise data prior to further processing
- Lots of information is hidden in basic statistics and analytics

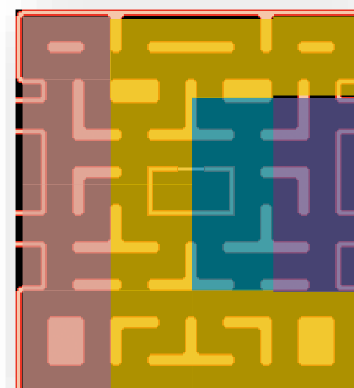
An Example: Tomb Raider: Underworld

A. Drachen, A. Canossa, and G. N. Yannakakis, “**Player Modeling using Self-Organization in *Tomb Raider: Underworld,***” in*Proceedings of the IEEE Symposium on Computational Intelligence and Games*, Milan, September, 2



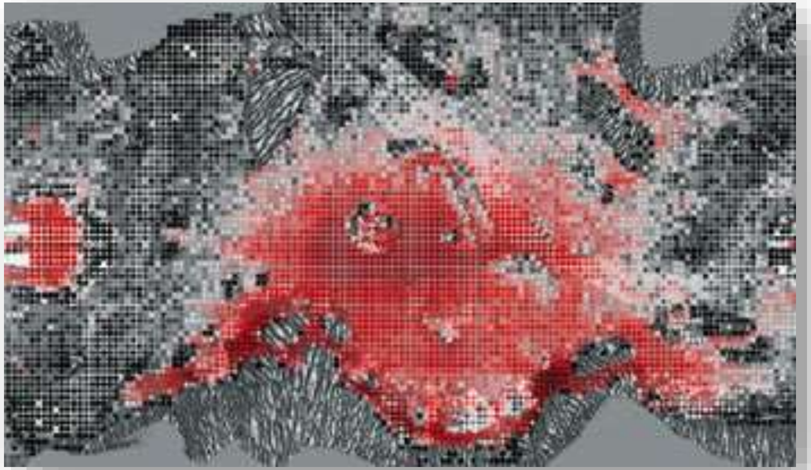
Heatmaps

- Plot specific information over the map of a level
- Typically, counts of a particular event in a map location
- The higher the count, the higher the *temperature*



Heatmaps in *Halo*

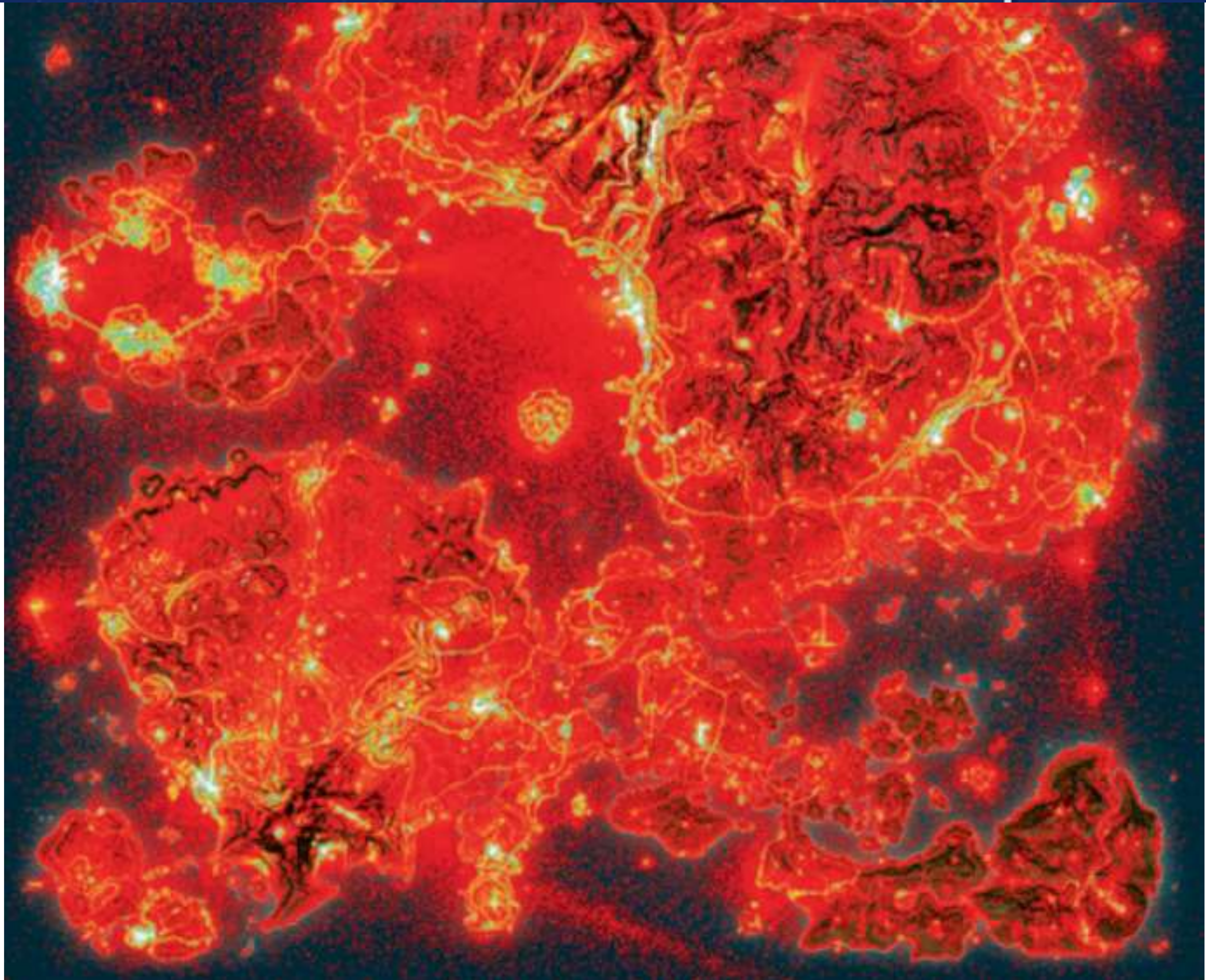
Source: How Microsoft Labs Invented a New Science of Play, Thompson, Wired, 2007



Number of deaths

Player navigation

Heatmap in *Just Cause 2*

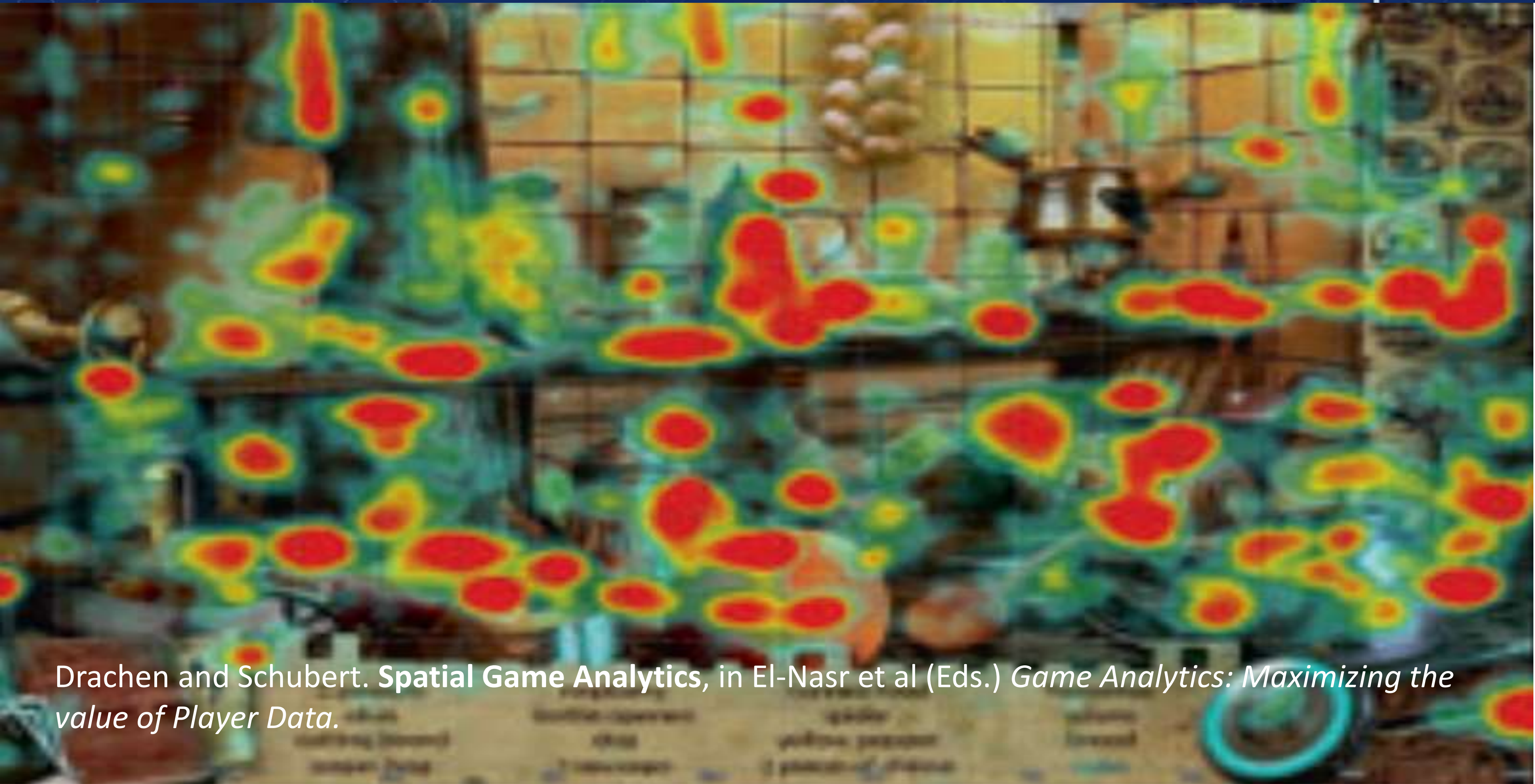


Drachen and Shubert. **Spatial Game Analytics**, in El-Nasr et al (Eds.) *Game Analytics: Maximizing the value of Player Data.*

Menu Heatmap

Drachen and Schubert. **Spatial Game Analytics**, in El-Nasr et al (Eds.) *Game Analytics: Maximizing the value of Player Data.*

Player Click Heatmaps



Drachen and Schubert. **Spatial Game Analytics**, in El-Nasr et al (Eds.) *Game Analytics: Maximizing the value of Player Data.*

Trajectory Analysis

Game Analytics as a Service

Can be viewed as a service to

- **Game Developers** (publishers, designers, programmers etc.)
- **Third Parties** (recommendation systems, adverts, etc.)
- **Players** (own performance, tailored challenges, game aids, etc.)

Ben Medler. **Visual Game Analytics**, in El-Nasr et al (Eds.) *Game Analytics: Maximizing the value of Player Data.*

Mining Data in Games

- Tracking data in games is a common practice
- The amount of data is usually huge
- One does not simply spot patterns in the data
- Data mining provides methods for finding **regularities** and **anomalies**

Drachen, Anders, Christian Thurau, Julian Togelius, Georgios N. Yannakakis, and Christian Bauckhage. "**Game data mining**." In *Game analytics*, pp. 205-253. Springer, London, 2013.

Game Data Mining: What can it do for you?

- Supplementary approach to traditional testing
- Imitate human playing styles
- Identify player profiles
- Spot cheating
- Spot game design flaws (e.g. “sweet spots” and frustrating sections)

Game Data Mining: What can it do for you?

- Tell you when and why players stop playing
- Enable player-driven adaptive games
- Overall, help you make better games
  - Project management
  - Marketing
  - Customer care

Game Data Mining:
What can it do for you?

A few examples

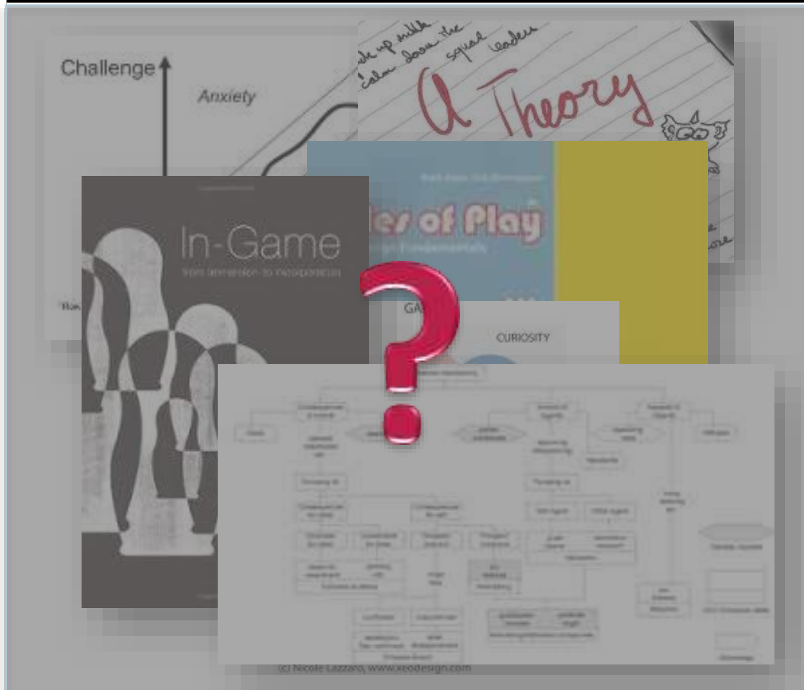
- Find weak spots in the design of game
- Which assets that are not getting used
- Figure out how players spend their time when playing
- Predict when they will stop playing
- Predict what they will do while playing
- Discover gold farmers in an MMORPGs
- Explore and use of social grouping
- Figure out how to convert non-paying to paying users


Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, …)
Model Free [Bottom-Up]
(Data Science, Machine Learning)

Player Modelling: In a nutshell

Yannakakis et al., **Player Modeling**, in *Dagstuhl Seminar on AI/CI in Games*, 2013

Theory (model-based)	**Data (model-free)**



Player Modelling: Limitations

Yannakakis et al., **Player Modeling**, in *Dagstuhl Seminar on AI/CI in Games*, 2013

- **Model-based** (theory-driven):
  - The majority of models have not been tested on or derived from interactive media
  - Models are not cross-validated (they should be!)
  - Agent-based models (OCC, BDI) – rather limited to agent-based interaction (it's limited in games)
- **Model-free** (data-driven):
  - Big data, crowdsourcing, elegant ML (e.g. deep learning / sequence mining), sensor technology
  - Supreme and obvious given the availability of big data and crowdsourcing
  - Problems: quality / quantity of data
  - Problems: garbage in – garbage out

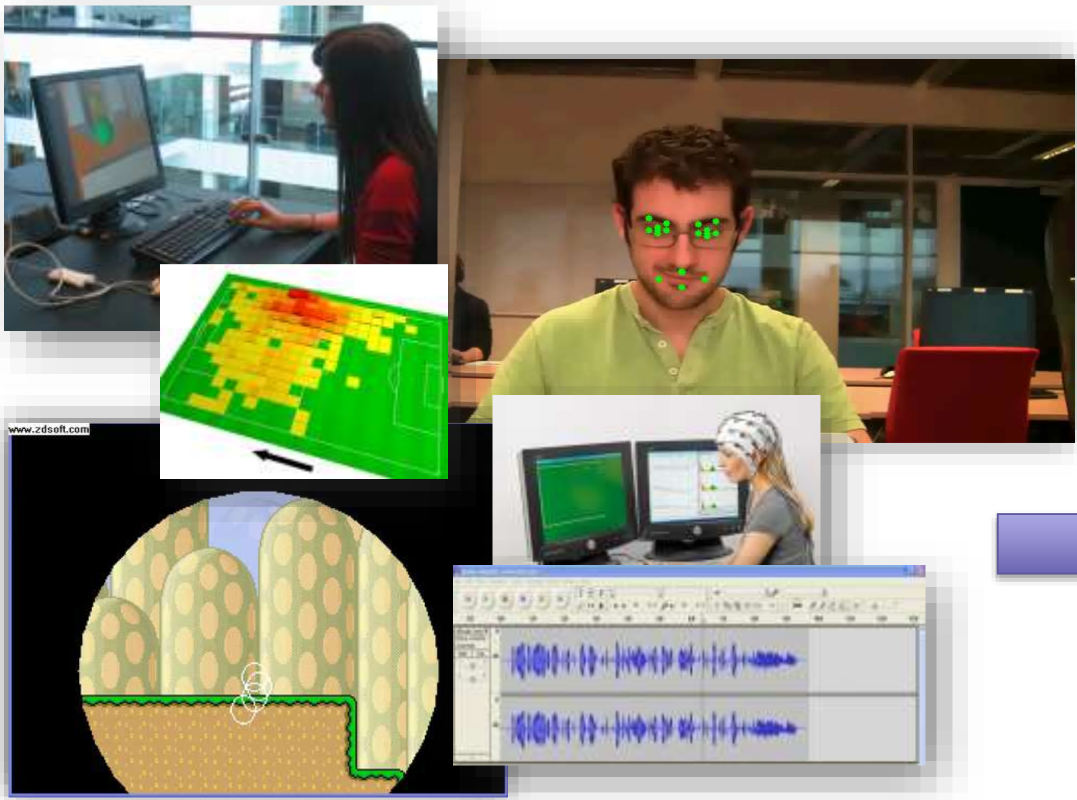
Player Modelling: Hybrid Approaches

Yannakakis et al., **Player Modeling**, in *Dagstuhl Seminar on AI/CI in Games*, 2013

Theory (model-based)	**Data (model-free)**

How – In a Nutshell

Yannakakis et al., **Player Modeling**, in *Dagstuhl Seminar on AI/CI in Games*, 2013



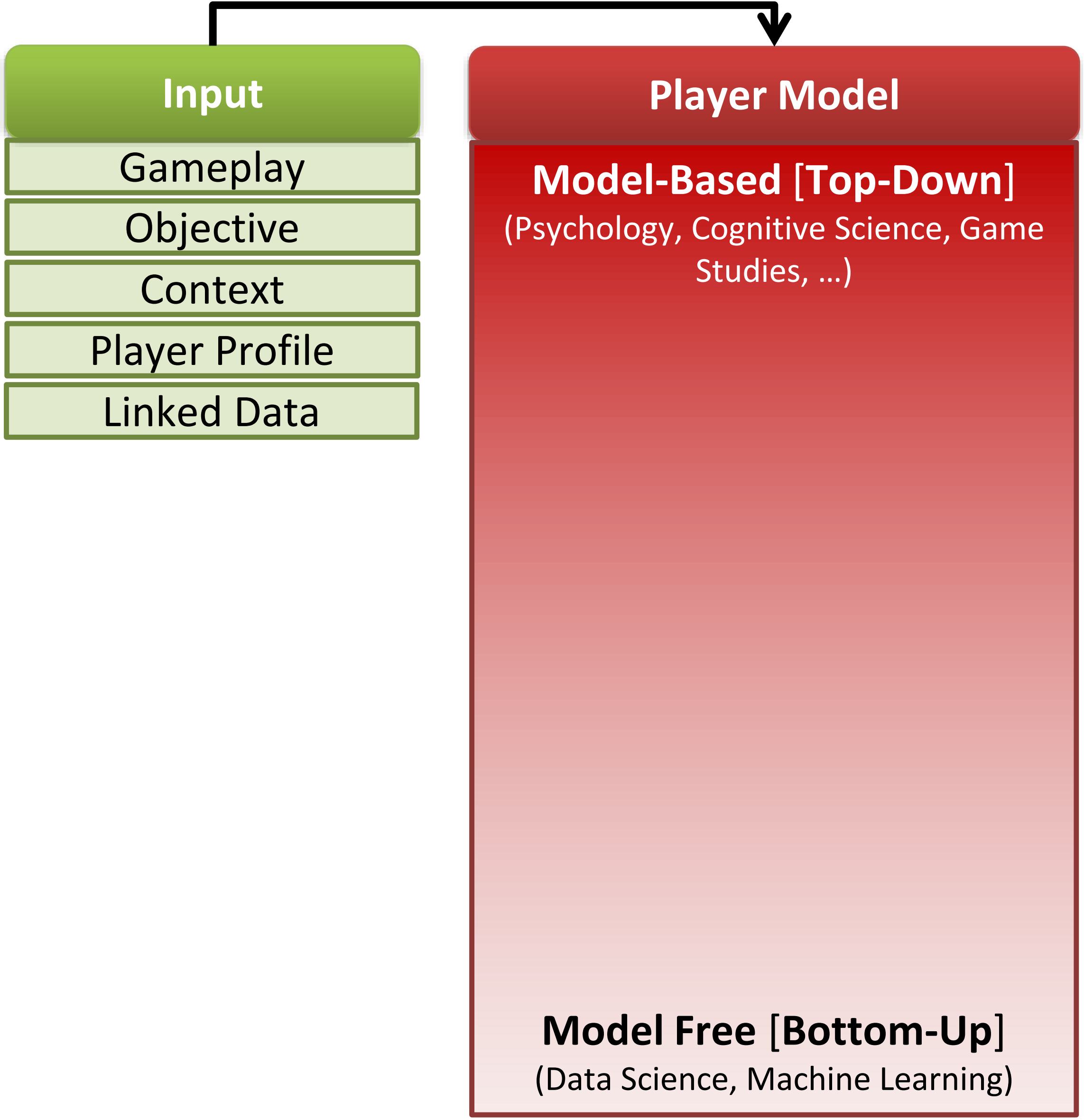
AI

Is X or Y **more** frustrating?

- ☐ X ☐ Y
- ☐ Both are **equally** frustrating
- ☐ **Neither** is frustrating

G. N. Yannakakis, P. Spronck, D. Loiacono and E. Andre, "**Player Modeling**," in Togelius et al., (Eds.) *Dagstuhl Seminar on Artificial and Computational Intelligence in Games*, 2013.


Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, …)
Model Free [Bottom-Up]
(Data Science, Machine Learning)


Input
Gameplay
Objective
Context
Player Profile
Linked Data
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, …)
Model Free [Bottom-Up]
(Data Science, Machine Learning)

Gameplay

Gameplay Input

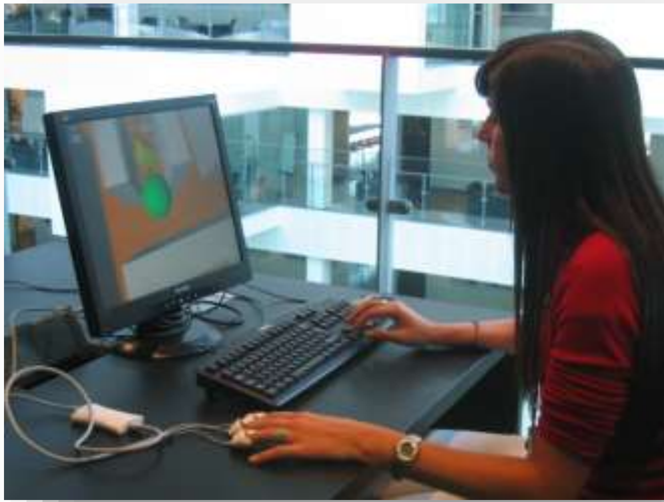
- Player game preferences, behavioral patterns
- Examples: tactics, strategy, play patterns, clickthroughs, deaths, weapon selection, character selection, etc…

- **Common features:** **Micro** vs **macro** actions
- **Pros:** real-time efficiency
- **Challenge:** we can’t tell much beyond player behavior…

Objective Input

- Bodily and physiological manifestations of gameplay
- **Pros:** reliable measures of user experience
- **Challenges:** many; let's see them in more detail

Objective Input – ways



- Cameras
  - Face
  - Body movement
  - Eye movement
- Other sensors/devices
  - Physiology (heart rate, skin conductance, ...)
  - Muscle activation (EMG)
  - Brain activity (EEG)
  - Speech

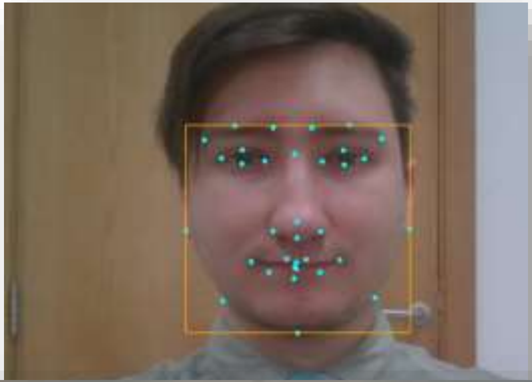
Physiology

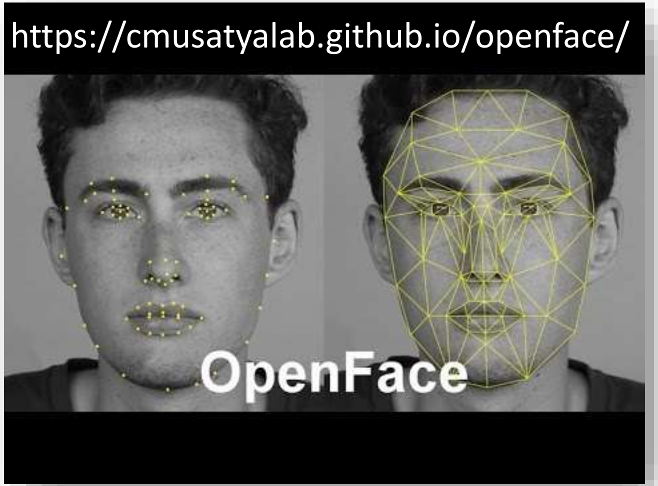
- **Common features:** summarization, time and frequency domain
- **Pros:** directly linked to arousal – immediate response
- **Challenges:** signal denoising/normalization; control for subjectivity of physiological responses, law of initial values, habituation, rebound

Visual Cues

- **Common features:** action unit detection, head pose stats
- **Pros:** every laptop has a camera, off-the-shelf cheap solution, natural interaction
- **Challenges:** do we really express emotions (facially) while playing? Head-pose might be more relevant? Models rely on posed/acted expressions

Tools for Video-Based Affect Detection





Eye-tracking

- **Common features:** blinks and gaze fixation (indicators of attention and engagement); total time spent looking at particular objects
- **Pros:** you know where your player looks at/focuses on/pays attention to
- **Challenges:** practicality, lab conditions (illumination), pupilometry doesn't really work in games

J. Munoz, G. N. Yannakakis, F. Mulvey, D. Witzner, G. Gutierrez and A. Sanchis, "**Towards Gaze-Controlled Platform Games**,"in *Proceedings of 2011 IEEE Conference on Computational Intelligence and Games*, 2011.

Speech

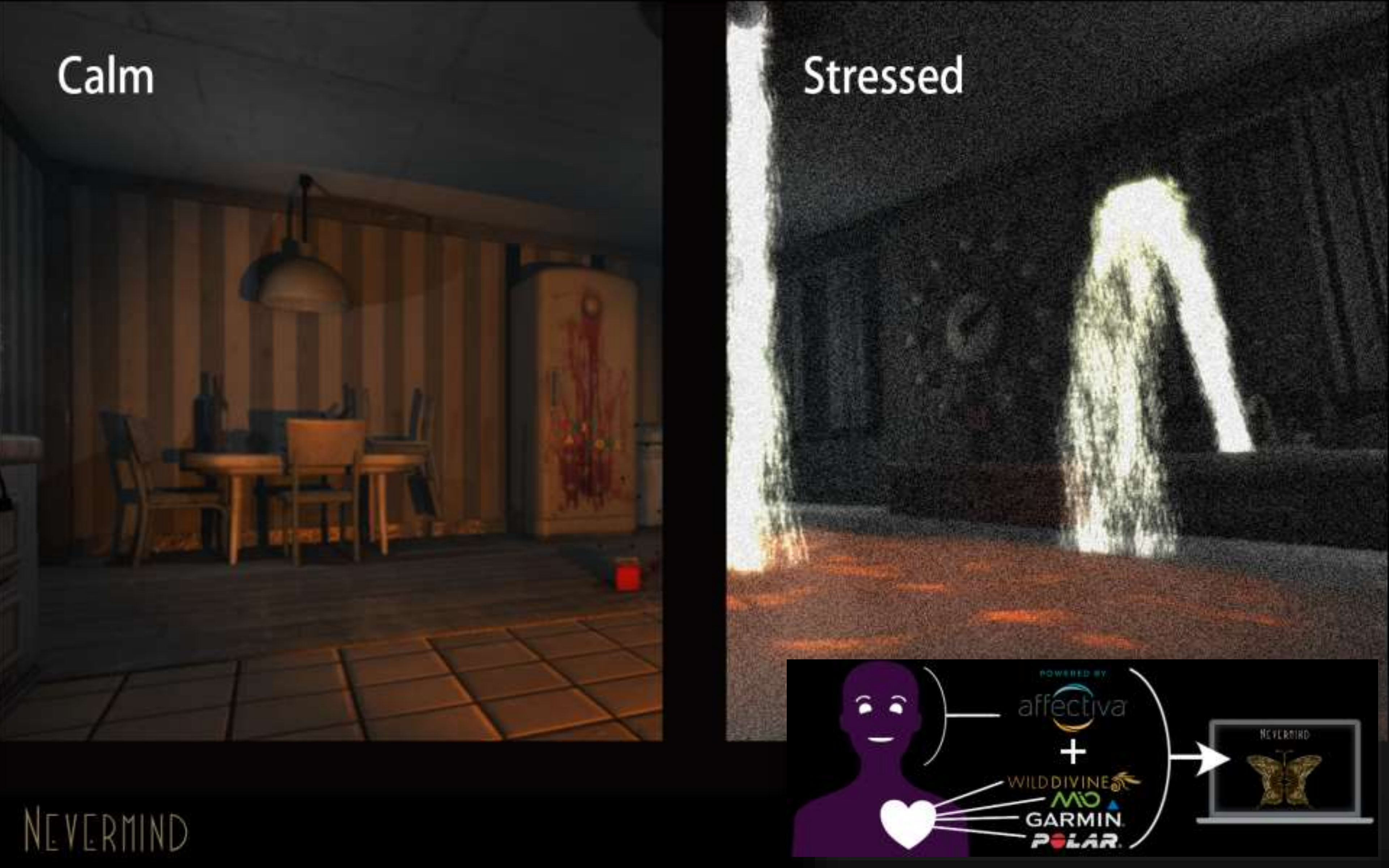
- **Common features:** frequency-based, detection of spoken words and other communication components (e.g. laughter and pauses)
- **Pros:** speech (pitch, loudness, quality) is linked to emotions (arousal/valence); useful in game-child interaction studies
- **Challenges:** verbal cues are rare; environment noise; multi-player games

Tools for Speech/Music-based Feature Extraction

openSMILE:)
by audEERING™

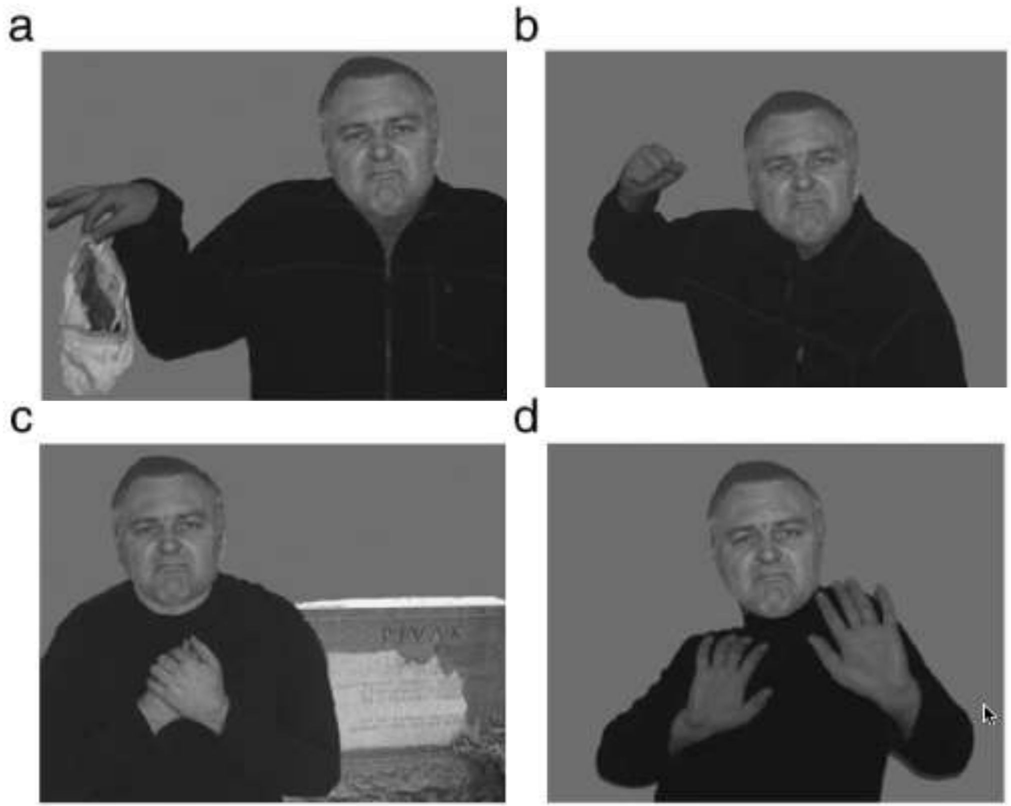
https://audeering.com/technology/opensmile/

Eyben, Wöllmer and Schuller: "**openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor**", In Proc. ACM Multimedia (MM), ACM, Florence, Italy, ACM, ISBN 978-1-60558-933-6, pp. 1459-1462, October 2010.


Calm
Stressed
POWERED BY
affectiva
+
WILDDIVINE
MIO
GARMIN
POLAR
NEVERMIND
NEVERMIND

Game Context

Context Matters!



Player Profile

Player Profile

- Player profile
  - Information about ones' personality, demographics, culture, age, gender, experience with games etc…
  - In general information that does not change due (or not altered via) games – at least not that rapidly…
- A player profile can form additional input(s) to a player model

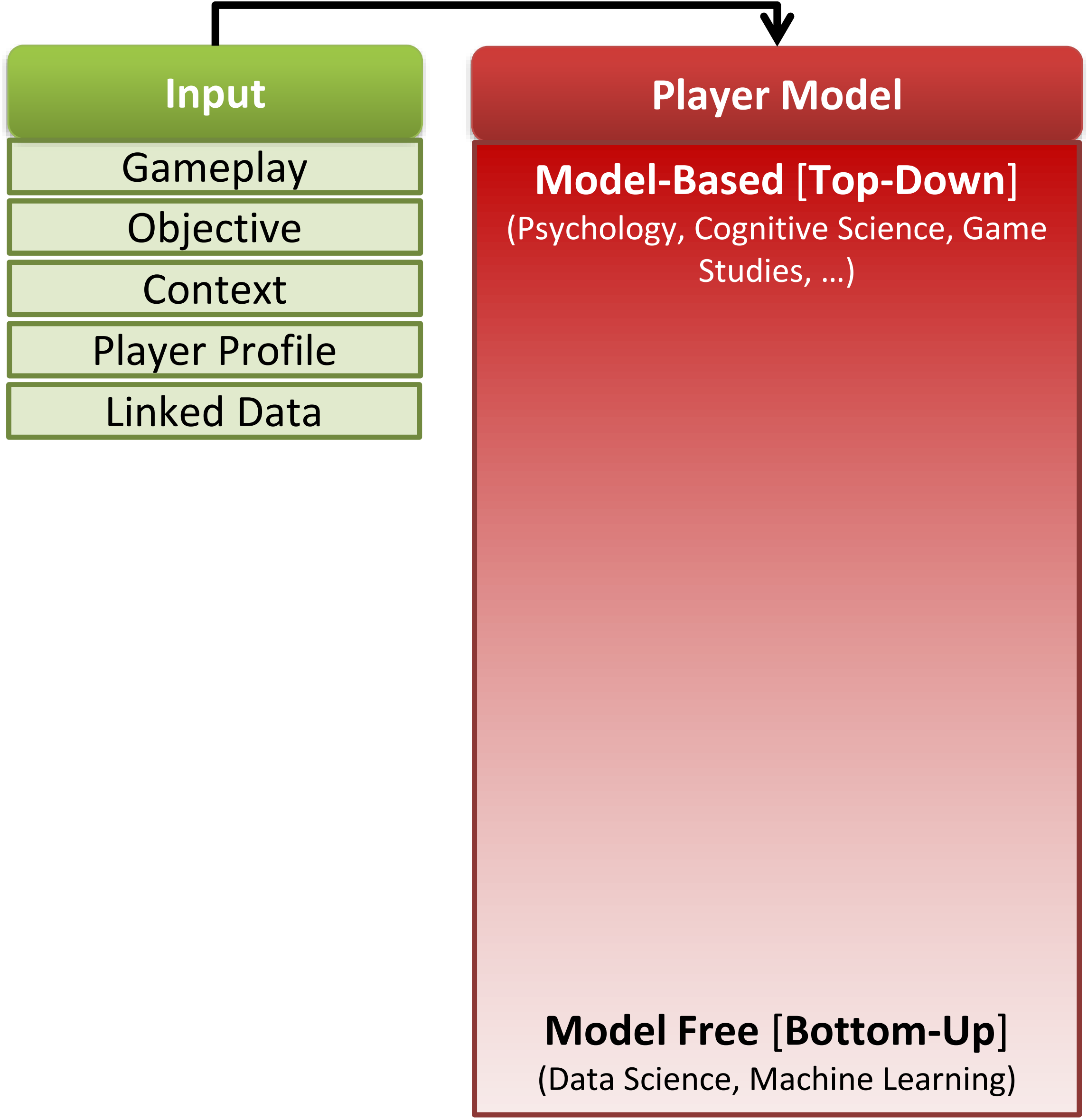
Reminder: Player Profile vs. Player Model

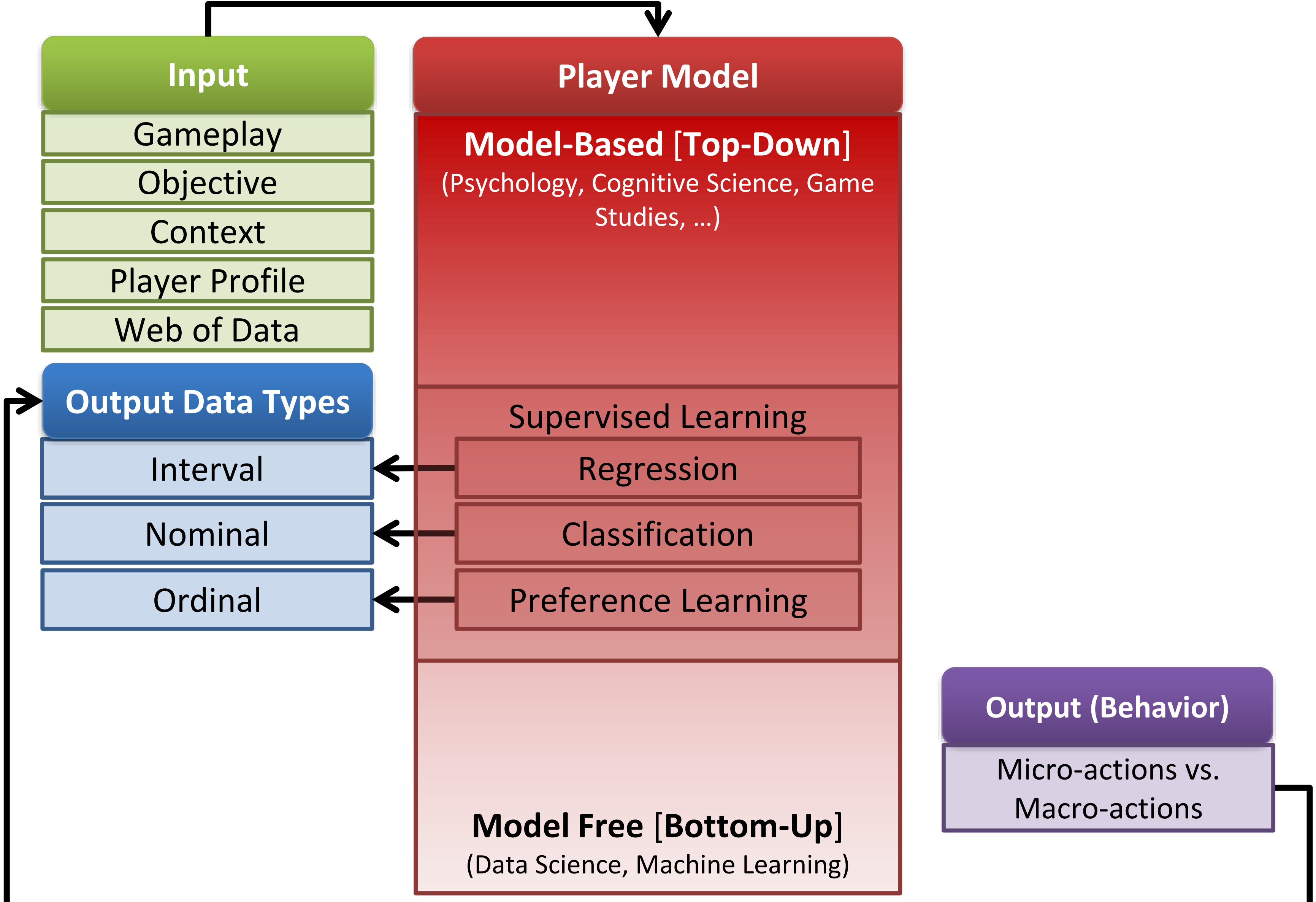
- What are the differences?
  - A **profile** is built on *static* data and not influenced by the game
  - A **model** is built on *dynamic* data from the gaming interaction and is (temporally) influenced by the game

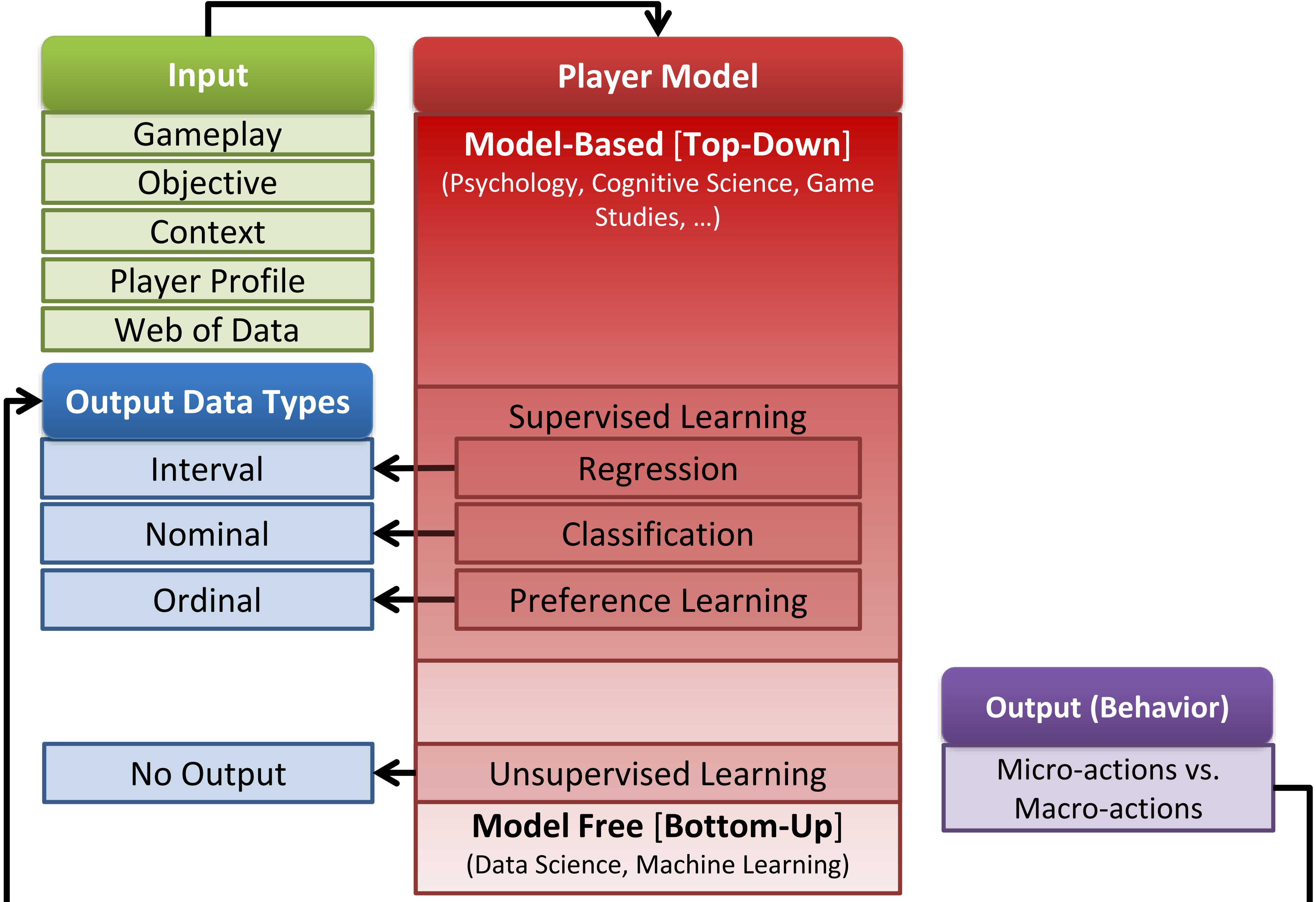
Linked Data

Linked Data

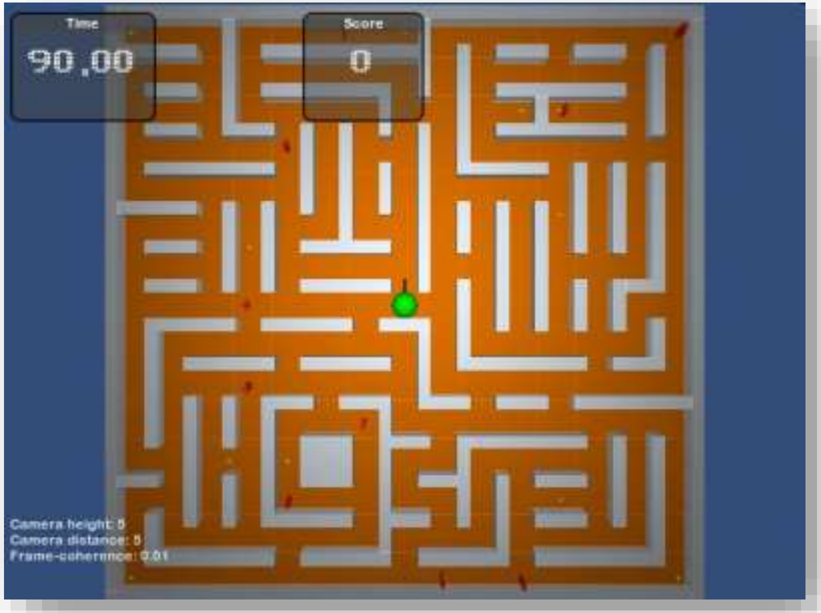
- FB emoticons
  - User daily state, emojis, tags
- Twitter-based – semantic info/analysis
- Game reviews
- …


Input
Gameplay
Objective
Context
Player Profile
Linked Data
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, ...)
Model Free [Bottom-Up]
(Data Science, Machine Learning)


Input
Gameplay
Objective
Context
Player Profile
Web of Data
Output Data Types
Interval
Nominal
Ordinal
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, ...)
Supervised Learning
Regression
Classification
Preference Learning
Model Free [Bottom-Up]
(Data Science, Machine Learning)
Output (Behavior)
Micro-actions vs. Macro-actions



Input
Gameplay
Objective
Context
Player Profile
Web of Data
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, ...)
Output Data Types
Supervised Learning
Interval
Regression
Nominal
Classification
Ordinal
Preference Learning
Output (Behavior)
No Output
Unsupervised Learning
Micro-actions vs. Macro-actions
Model Free [Bottom-Up]
(Data Science, Machine Learning)



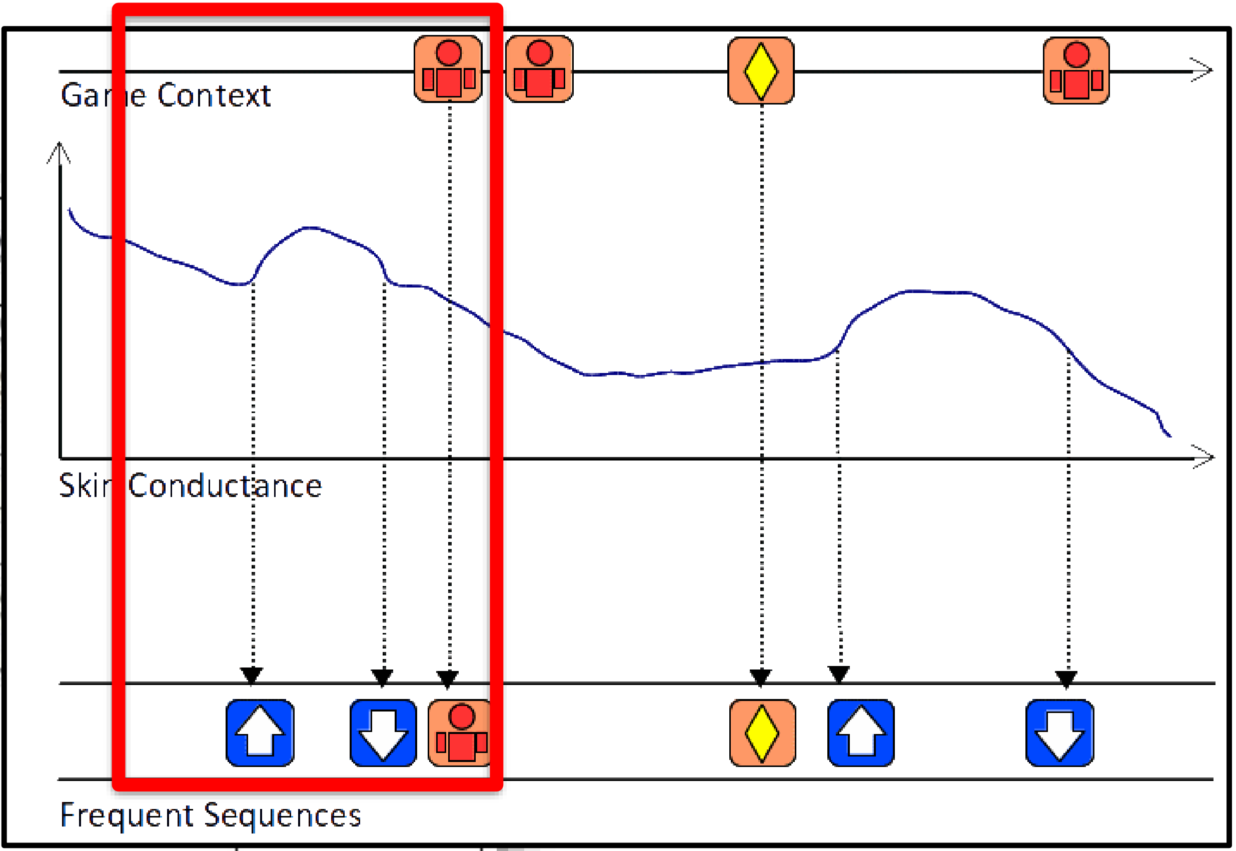
Example (Player Experience Modeling)

Mazeball Dataset – http://hectormartinez.com/

Sequence Mining (General Sequential Pattern)

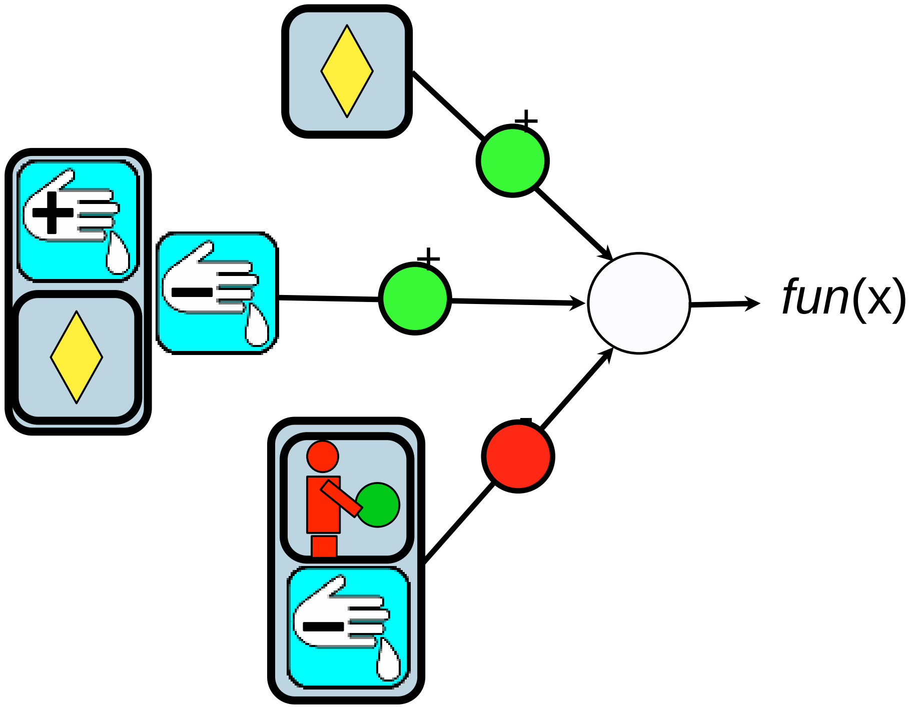
Martinez and Yannakakis, **Mining Multimodal Sequential Patterns: A Case Study on Affect Detection**, ICMI, 2011 **[Outstanding Student Paper Award]**

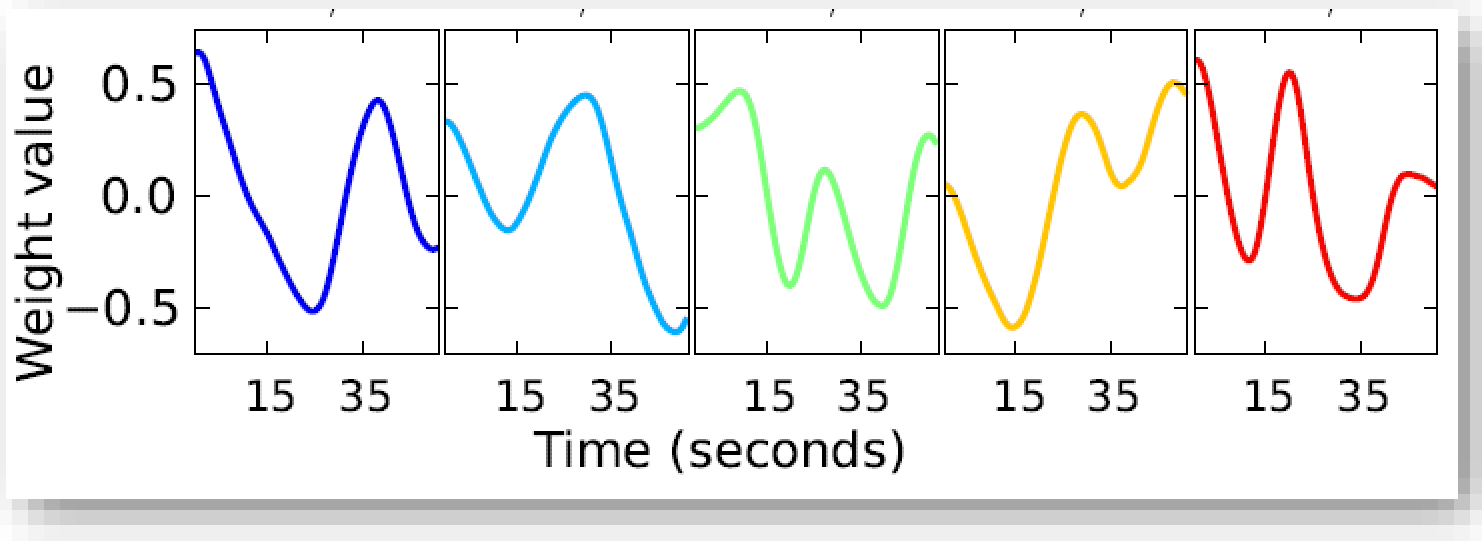
3-sequences	G_{max} 1.5	G_{max} 3
$(\$\ s^{\Uparrow})(s^{\Downarrow})$	141	168
$(E\ s^{\Uparrow})(s^{\Downarrow})$	131	163
$(s^{\Uparrow})(\$)(s^{\Downarrow})$	123	164
$(s^{\Uparrow})(E)(s^{\Downarrow})$	116	164
$(s^{\Uparrow})(s^{\Downarrow})(\$)$	112	181
$(E)(s^{\Uparrow})(s^{\Downarrow})$	109	175
$(s^{\Uparrow})(s^{\Downarrow})(E)$	106	180
$(\$)(s^{\Uparrow})(s^{\Downarrow})$	105	186
$(s^{\Uparrow})(\$\ s^{\Downarrow})$	105	158
$(s^{\Uparrow}\ s^{\Downarrow})(\$)$	102	139



Mazeball's Model of Fun

Martinez and Yannakakis, **Mining Multimodal Sequential Patterns: A Case Study on Affect Detection**, ICMI, 2011 **[Outstanding Student Paper Award]**



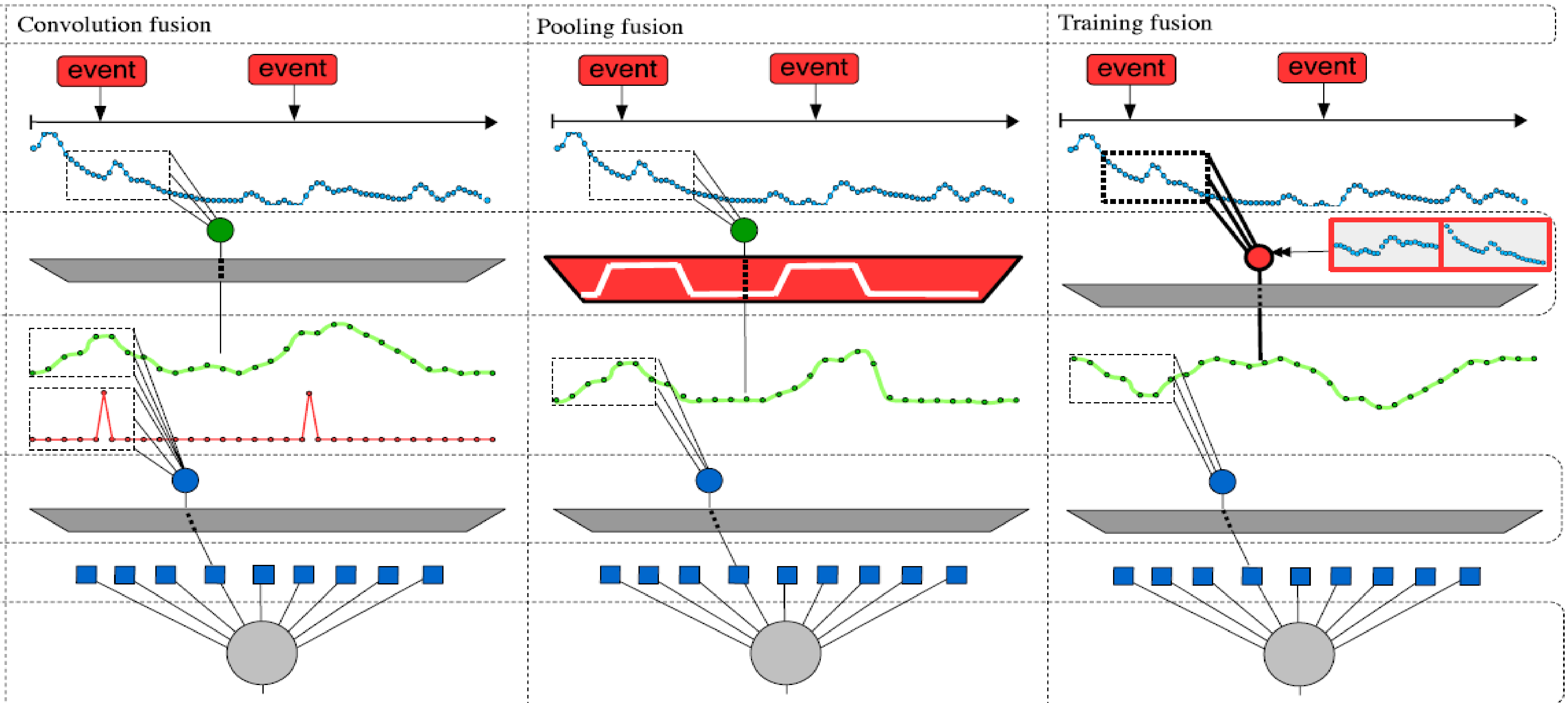


Convolution for Affect Detection

Martinez, Bengio and Yannakakis, **Learning Deep Physiological Models of Affect**, *IEEE Computational Intelligence Magazine*, 2013

Deep Fusion of Events and Signals

Martinez and Yannakakis, **Deep multimodal fusion: Combining discrete events and continuous signals,** *Proceedings of the 16th International conference on multimodal interaction,* 2014



Example (Player Behavior Modeling)
Tomb Raider: Underworld

Drachen, Canossa & Yannakakis, **Player modelling using self-organisation in *Tomb Raider: Underworld*,** IEEE CIG 2009

Motivation

- How do people play TRU?
- User testing: Do people play TRU as intended?
- Player modeling using *metrics* via machine learning → alternative quantitative approach to traditional qualitative approaches of user and playability testing

Clustering in TRU

- Commercial major game test-bed: Tomb Raider Underworld
- Large-scale data collection (1365 players)
- Data is *clean* and *live:* gathered in a natural setup via an industrial logging system and a commercial web service
- First application of SOMs on high-level behaviors of completed games
- Directly addressing game-industry requirements
  - limitations of scalability and commercial-game practicality are minimized
  - bridging the AI gap

Player Data Collection

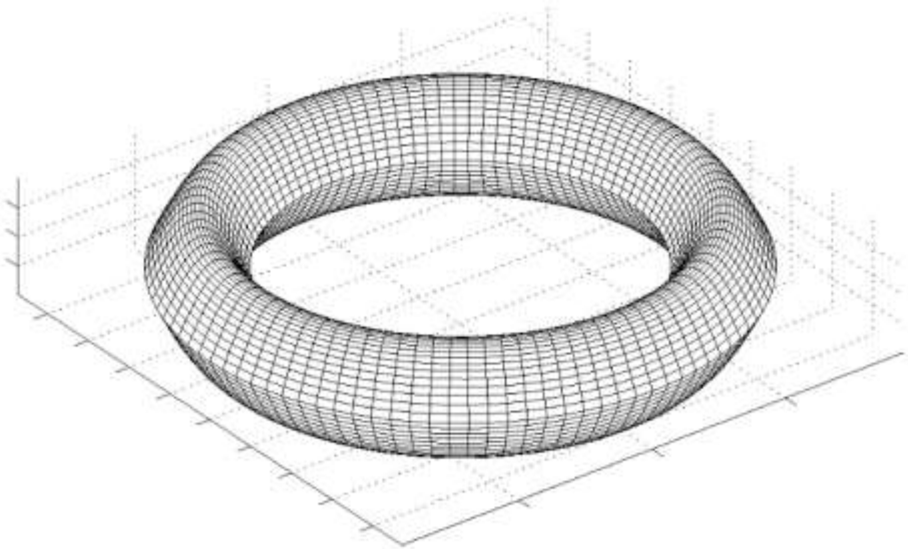
- EIDOS Metrics Suite software
  - Record gameplay data (*metrics*) of EIDOS games
  - Data are stored in an SQL-server via ELT process
  - Several Features are extracted (work done by *Crystal Dynamics*)
    - 3D coordinates, completion time, # deaths etc.
- **Live Data** (xBox Live!)
  - Published version of TRU played in gamers' natural habitats
  - Data free from laboratory bias and experiment expectancy effects
- Data collection during November 2008
  - 25240 Players
  - 1365 of those completed the game
  - Data are stored for all 7 levels (100 map units)

Extracted Features

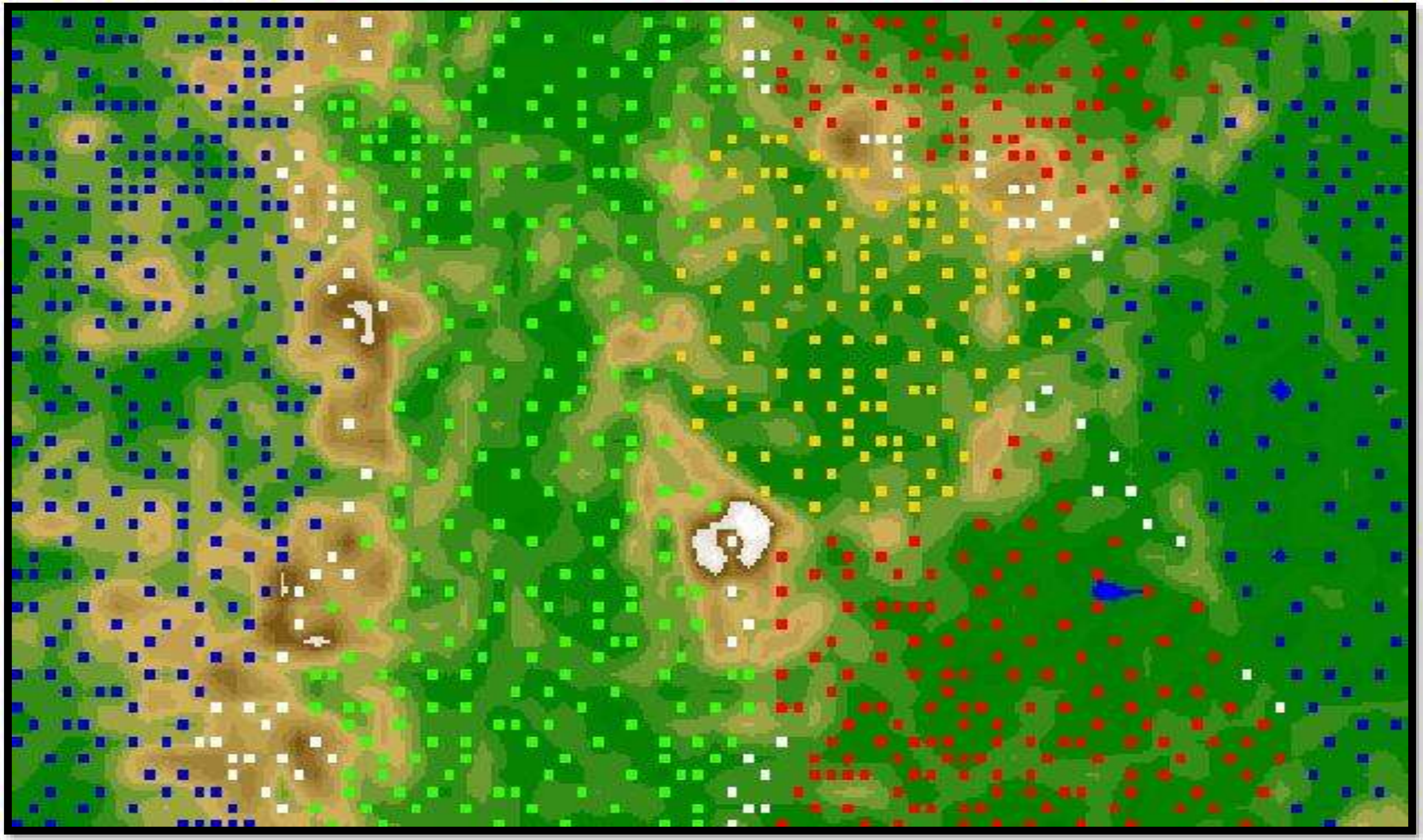
- **Causes of Death (% over total number of deaths)**
  - **Opponent** – 28.9% of all deaths – Min: 6%, Max: 60%
  - **Environment** – 13.7% of all deaths – Min: 2%, Max: 45%
  - **Falling** – 57.2% of all deaths – Min: 27%, Max: 83%
- **Total Number of Deaths** – 140 on average – Min: 16, Max: 458
- **Completion Time** – 550 min on average – Min: 3h, Max: 29h
- **Help-on-Demand (# hint + answer requested) –** 29 on average – Min: 0, Max: 148

Emergent Self-Organizing Maps (ESOMs)

- Unsupervised learning through self-organization of a neuron map
- Dimensionality reduction to 2D
- Training
  - Batch
  - Toroid topology (50X100 neurons)
  - Rectangular grid
- Performance Measures
  - Topographic error
  - Quantization error

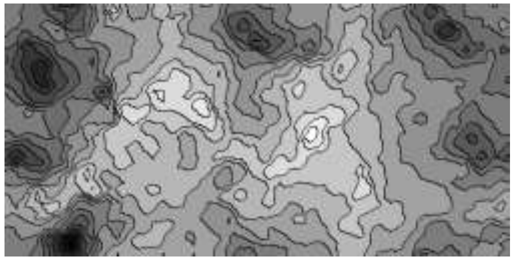


Clustering in TRU (ESOMs)

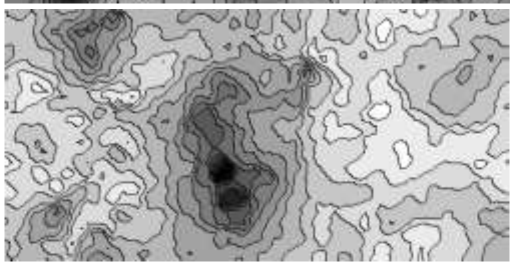


Drachen, Canossa & Yannakakis, **Player modelling using self-organisation in *Tomb Raider: Underworld*,** IEEE CIG 2009

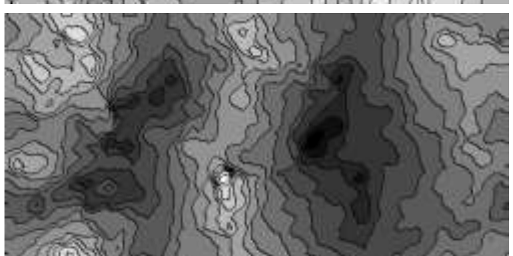
Feature Planes



Cause of Death: Opponent

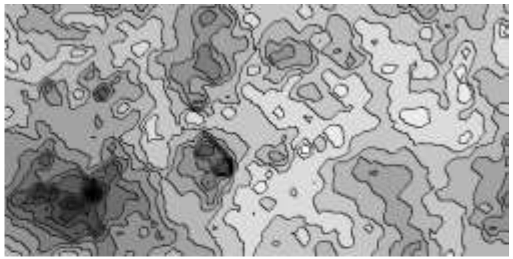


Cause of Death: Environment

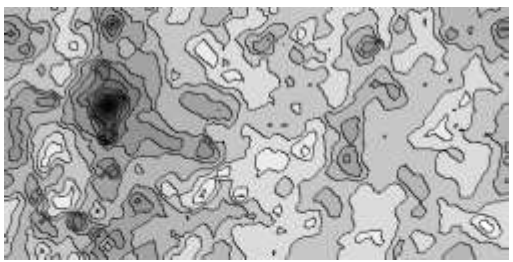


Cause of Death: Falling

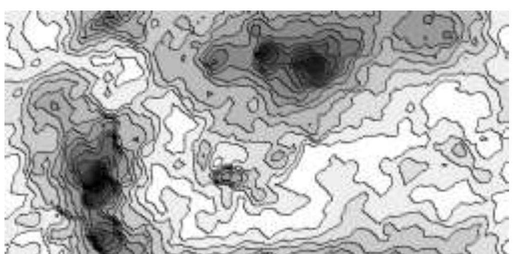
Feature Planes



Number of Deaths



Completion Time



Help on Demand (HoD)

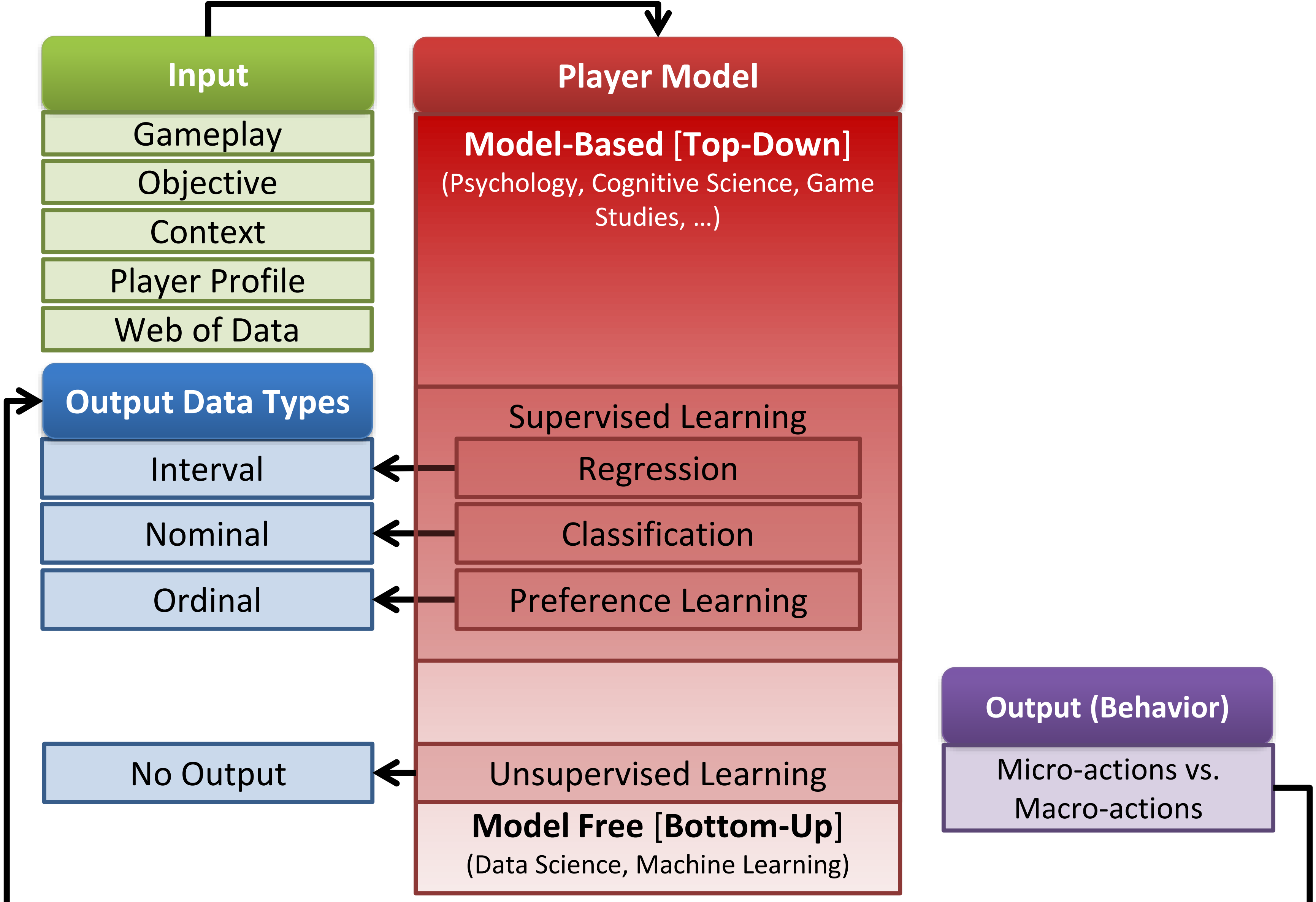
Video: Four Resulting Player Clusters

https://www.youtube.com/watch?v=HJS-SxgXAI4

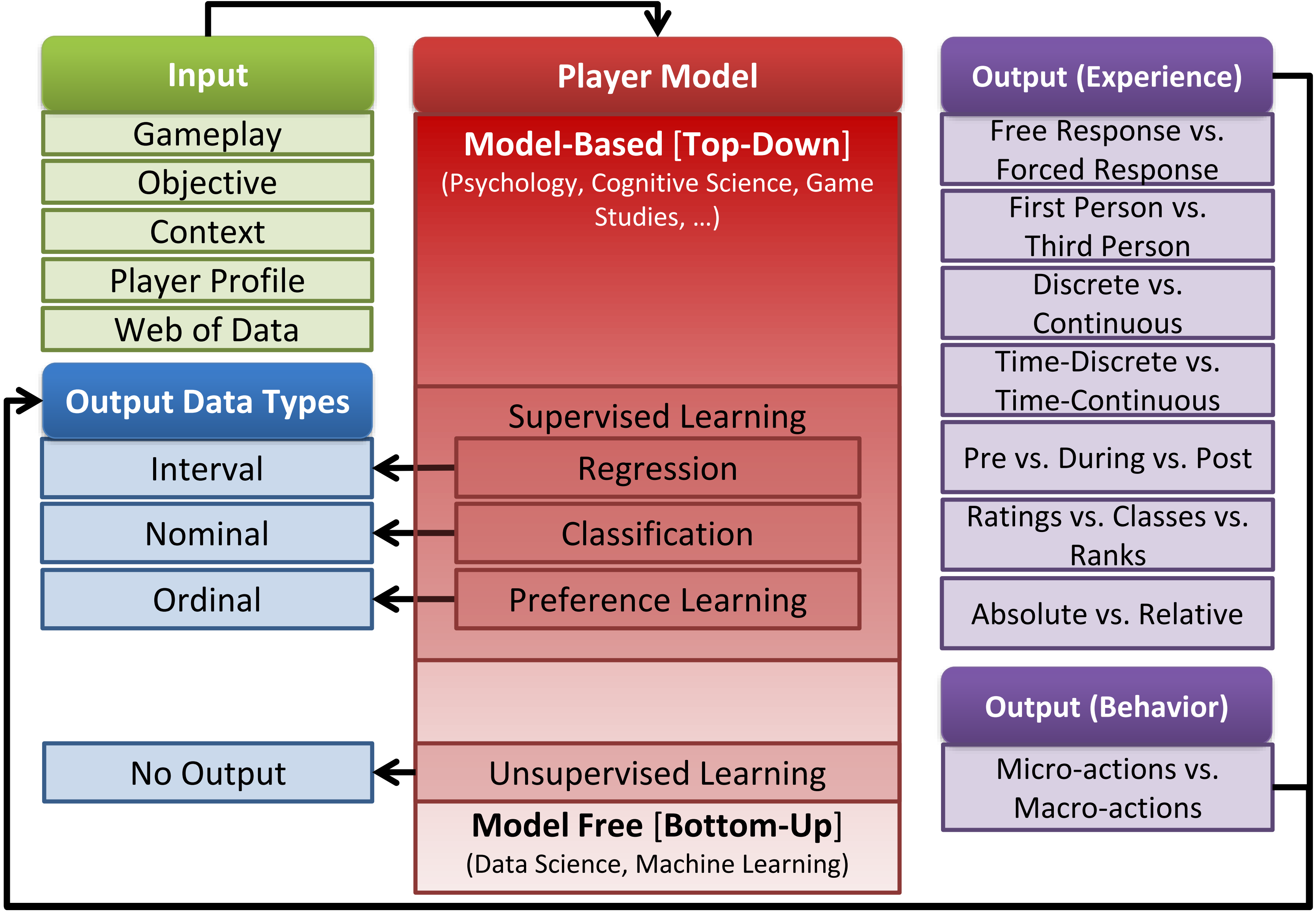
Analysing Player Behaviour in *Tomb Raider: Underworld*

https://www.youtube.com/watch?v=A89ZDjF51Nk

AI and Games



Input
Gameplay
Objective
Context
Player Profile
Web of Data
Output Data Types
Interval
Nominal
Ordinal
No Output
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, …)
Supervised Learning
Regression
Classification
Preference Learning
Unsupervised Learning
Model Free [Bottom-Up]
(Data Science, Machine Learning)
Output (Behavior)
Micro-actions vs. Macro-actions


Input
Gameplay
Objective
Context
Player Profile
Web of Data
Output Data Types
Interval
Nominal
Ordinal
No Output
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, ...)
Supervised Learning
Regression
Classification
Preference Learning
Unsupervised Learning
Model Free [Bottom-Up]
(Data Science, Machine Learning)
Output (Experience)
Free Response vs. Forced Response
First Person vs. Third Person
Discrete vs. Continuous
Time-Discrete vs. Time-Continuous
Pre vs. During vs. Post
Ratings vs. Classes vs. Ranks
Absolute vs. Relative
Output (Behavior)
Micro-actions vs. Macro-actions

Labels are Key!

Why Output (Annotation) is Key?

- Annotation is the labelling of experience
- This is ultimately the *ground truth* (golden standard) of experience
- This is the **training signal** for your computational models

Key Questions of Labelling

- Who annotates?
- When?
- How often?
- How?

Who Annotates?

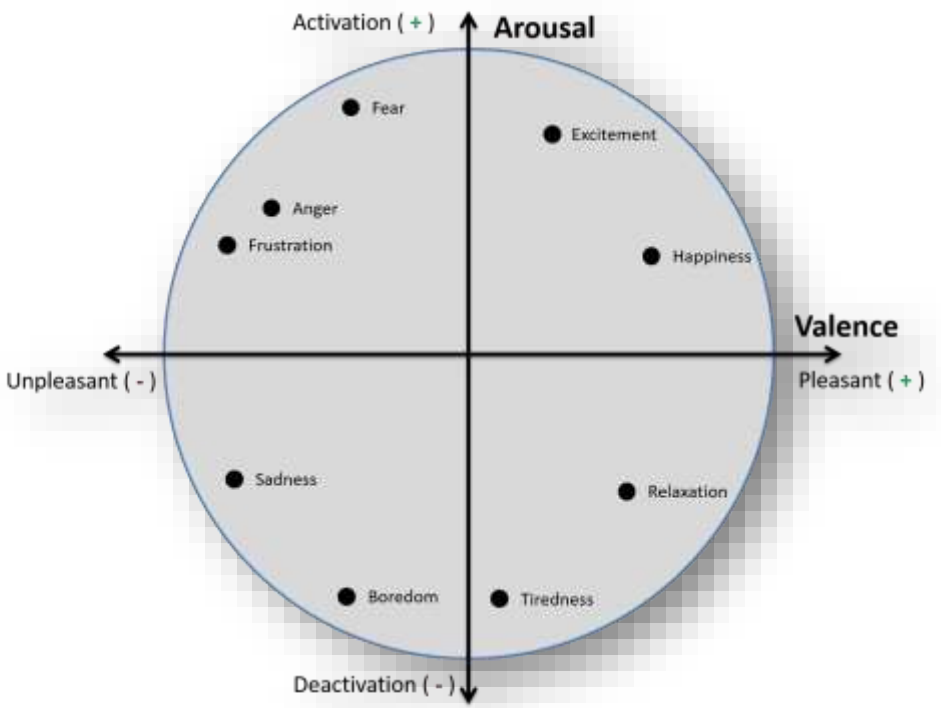
- Third Person
  - Usually a domain expert (game designer) or a psychologist

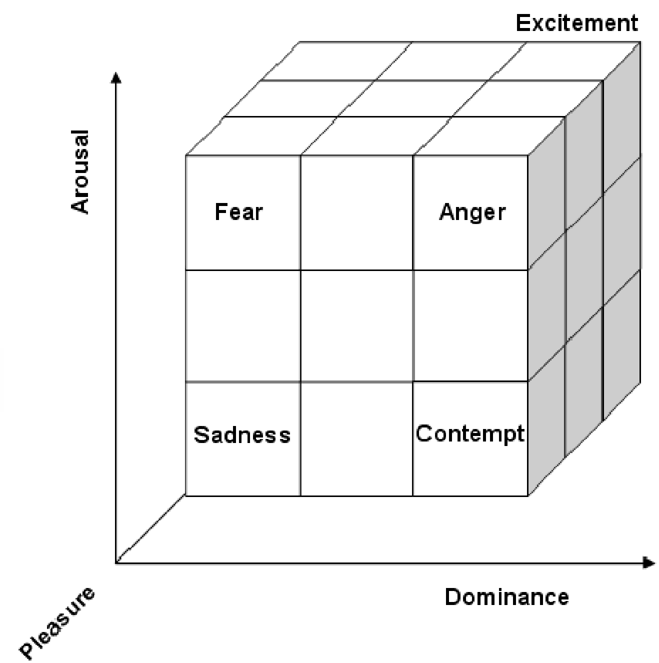
- First Person
  - The person actually experiencing the emotion/affect

	Third person	First person
+	• Expert knowledge	• Reported true experience
-	• Assumptions about the true emotion • Reporting effects	• Self-deception • Reporting effects • No expert knowledge

How is Player Experience Represented?

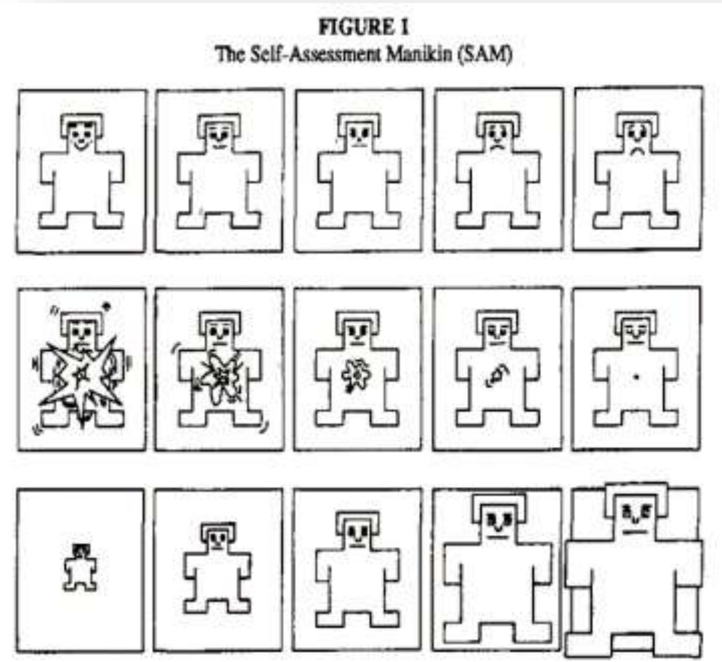
- Discrete states (e.g. fun, engagement, frustration)
- Continuous dimensions (e.g. arousal and valence)

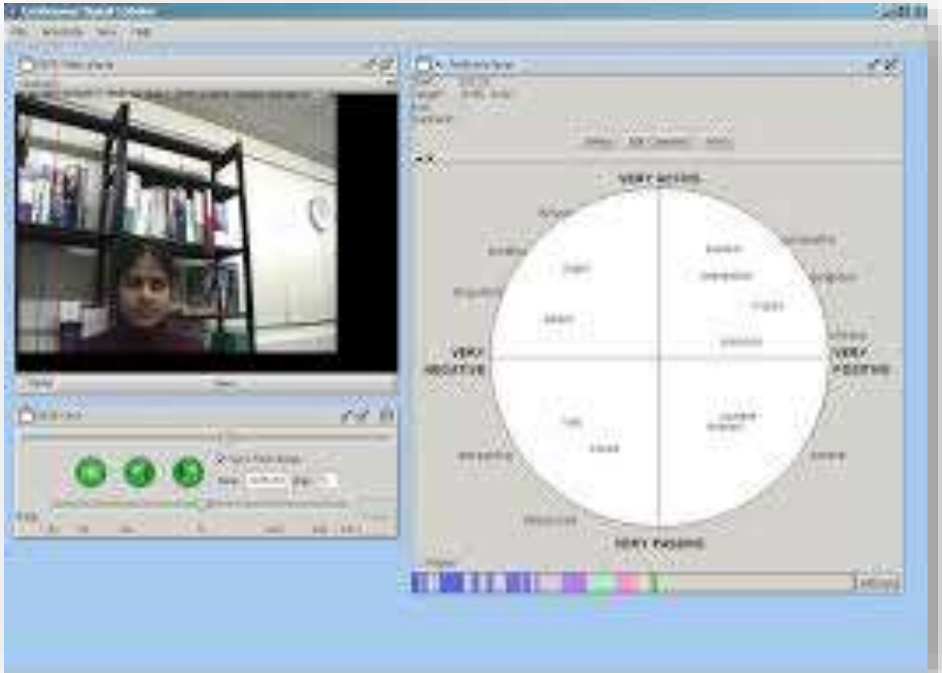




How Often to Annotate?

- Time-Discrete (e.g. self-assessment manikin)
- Time-Continuous (e.g. FeelTrace, *AffectRank*)





How Often to Annotate?

- Depends on
  - Application (speed of interaction: e.g. games vs. movies vs. e-learning apps)
  - Signal (e.g. physiology is slower than body movement and speech)
- No gold standard

When to Annotate?

- Before play (**Pre**-Experience)
- **During** (real-time) experience
- After play (**Post**-experience)

	Before	During	After
+	• Set the baseline of a player's state prior to playing a game • Information that enriches our models • Detecting the *relative* change from baseline	• Report on the spot! • Real experience (better ground truth?) • Limited memory effects	• Controlled • Non-intrusive
-	• No data about experience	• Highly intrusive • Distorts the experience (first person)	• Self-deception • Various reporting effects

A note about time and self-report!

- Self-reports are time-dependent
- Real experience vs. Post-experience
  - Few seconds → Real experience
  - Few minutes/hours → *Episodic* memory (context retrieval)
  - More → *Semantic* Memory (beliefs)

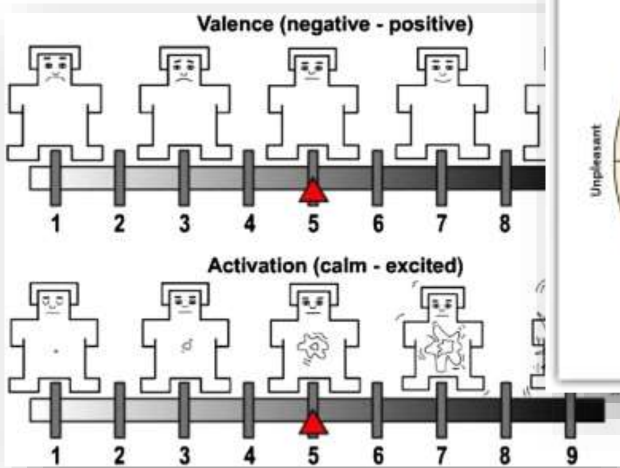
NB. The gap between our **memory of experience** and our **experience** is more prominent when we report unpleasant emotions such as anger, sadness and tension. Also: The experience felt near the end of a session (e.g. a game level or a game) affects our report – aka ***peak-end rule***.

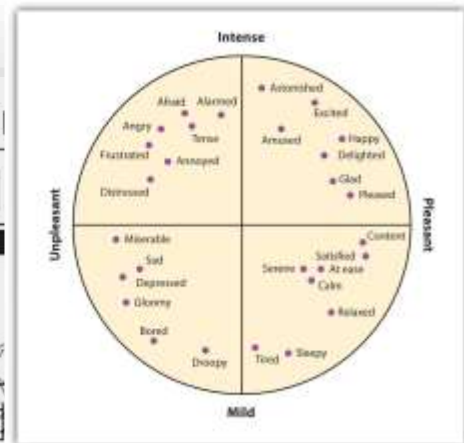
Which Annotation (Data) Type?

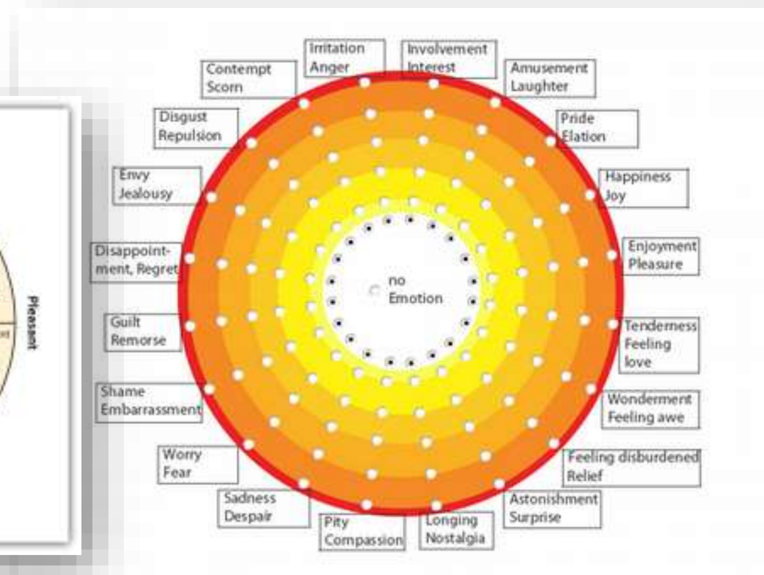
Which Annotation (Data) Type?

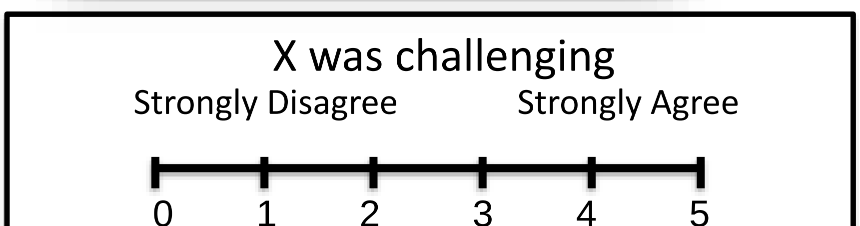
- Scalar (Likert scale, arousal/valence score, SAM) – **Rating**
- Binary value or a class – **Class**
- Preference between two or more options – **Rank**

Rating







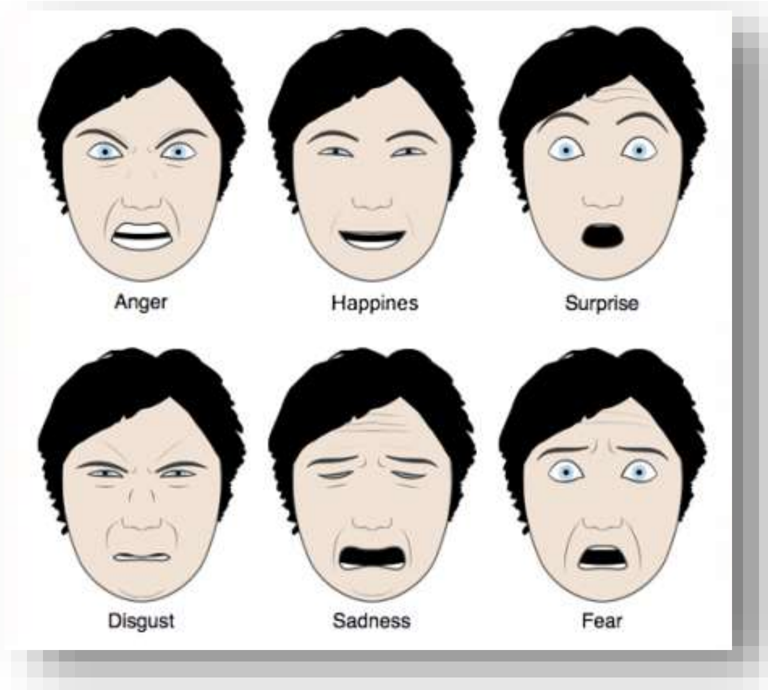


Examples: Geneva Wheel, SAM, Likert Scales, PAD values

Class

Examples:

- This facial expression is **happy**! (Eckman)
- Arousal values higher than 0.6 belong to class **aroused**
- This skin conductance peak denotes **stress**



Rank

X is **more/less** [challenging / frustrating / arousing / boring / fearful / …] than Y

challenging
frustrating
arousing
boring
fearful
…

- Requires at least two instances!
- N-Alternative Forced Choice (4-AFC is popular)

☐ X is **more/less** frustrating than Y
☐ Both are **equally** frustrating
☐ **Neither** is frustrating

Which Annotation (Data) Type? Summary

	Class	Scalar	Rank
+	Easy to analyse and process. Only one instance (to be annotated) is required. Part of subjectivity is eliminated.	Easy to analyse and process. Only one instance (to be annotated) is required.	Controls for reporting memory effects (increasing/decreasing the memory window). Eliminates subjectivity biases. It is natural to compare (key trend in marketing research). Evidence from neuroscience.
-	Assumptions made about classification. What is the "gold" threshold value that splits classes?	**Highly** subjective; Use of scale-bias; Culture-bias, Personality (temperament, interests)-bias; Increased between-participants effects; Logical errors: confused by ordinal scales. *Primacy* and *Recency* order effects. Memory effects. Key **fundamental issues** (see next slide).	*Primacy* and *Recency* order effects. Memory effects

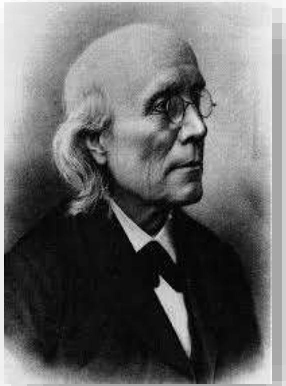
What is the Value of Player Experience?

A **thesis**: emotions are intrinsically **ordinal** (relative)

The thesis is supported by **theoretical arguments** and **empirical evidence** across disciplines

It **reframes** the way emotions are viewed, represented and analysed computationally

One of the first Challenges in Psychology



Mapping the intensities of responses to particular stimuli

That is basic to AC/HCI/UX/GUR…: we call it **labelling**

Two approaches have a long history

- The older (Fechner) was based on comparing stimuli, and finding '**just noticeable differences**'
- Much later, Stevens introduced '**magnitude estimation**' – asking people to give a number. Twenty years ago, psychologists tried a magnitude estimation approach to labelling. The data are in, and we know it doesn't work as straightforwardly as they hoped.

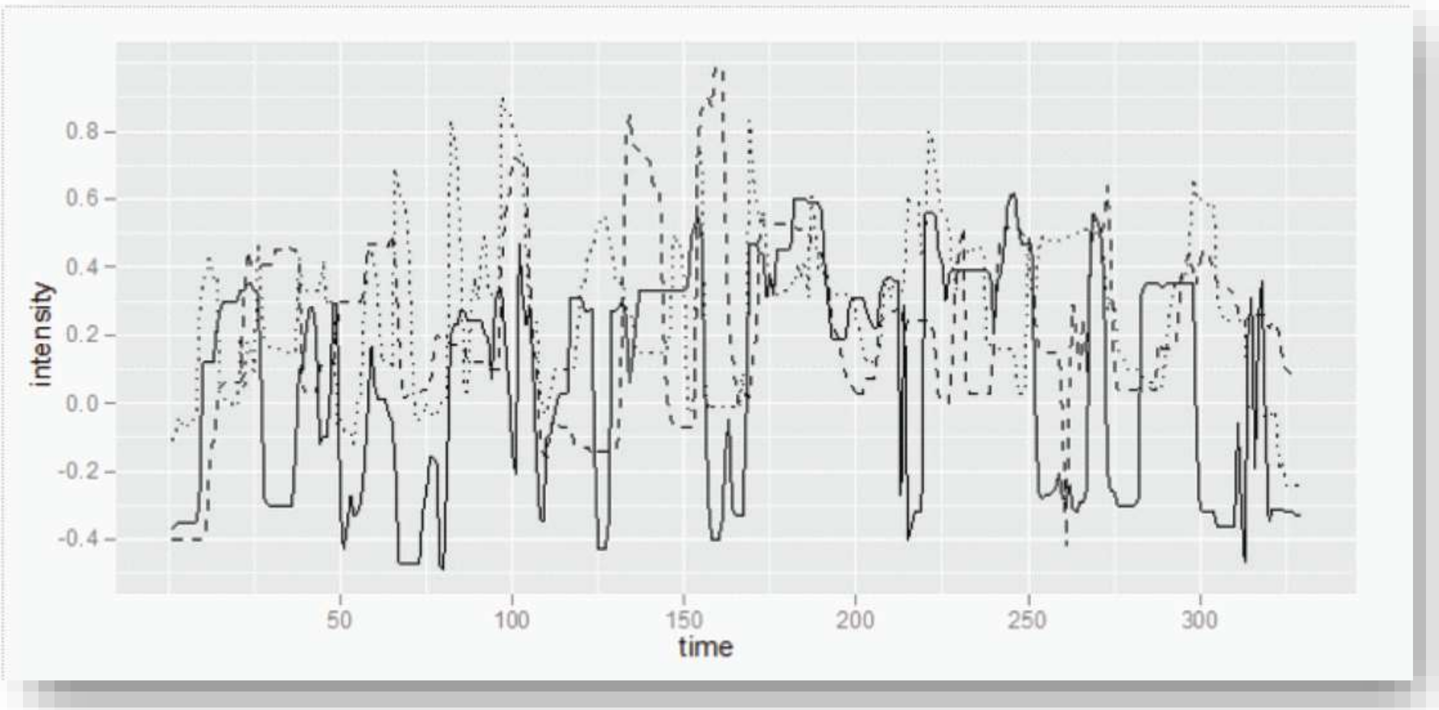
The core finding is simple…

When people are presented with a *piece of data* and asked to assign a *magnitude* describing an emotional response, they tend to **disagree** quite substantially.

Douglas-Cowie et al. "**Multimodal databases of everyday emotion: Facing up to complexity**," *Ninth European Conference on Speech Communication and Technology*. 2005.

Or else…

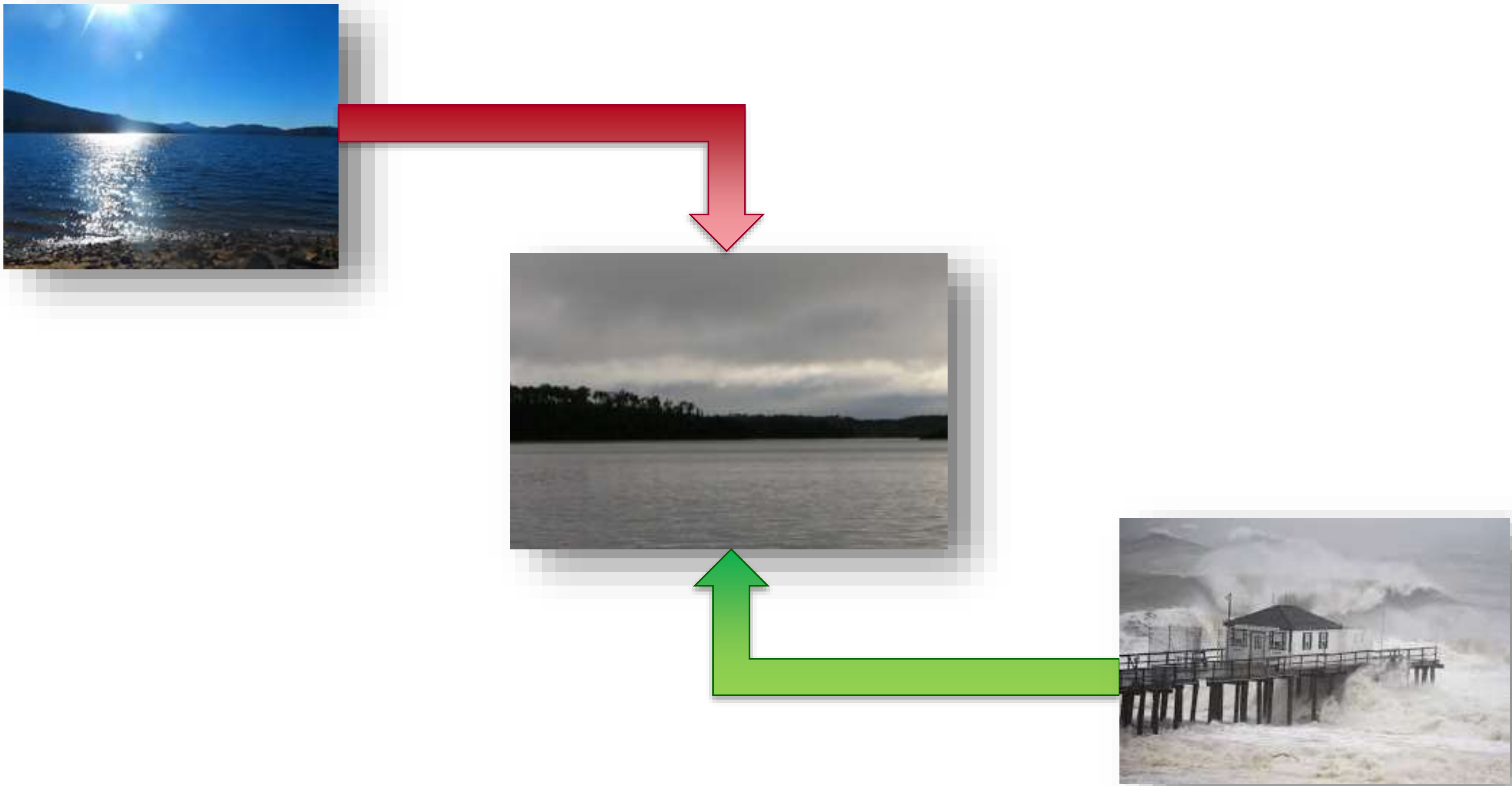
….there is **seriously** no point in this….



Douglas-Cowie et al. "**Multimodal databases of everyday emotion: Facing up to complexity**," *Ninth European Conference on Speech Communication and Technology*. 2005.

Why? Multivalued Emotion…

Why? Adaptation Level...



Are we living in an **ordinal** world?

Yannakakis, Cowie, Busso, **The Ordinal Nature of Emotions**, ACII, 2017 **[Best Paper Award]**

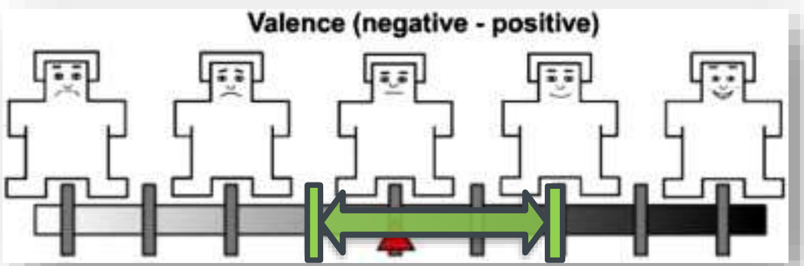
Magnitude is deeply context-dependent

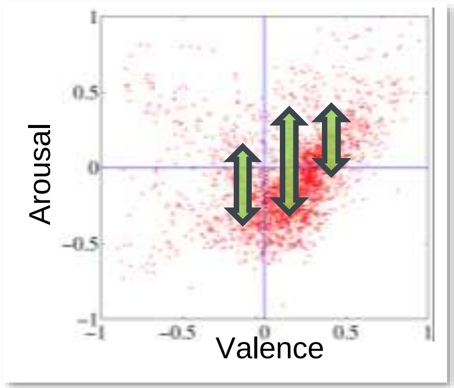
Pairwise comparison is robust! A reference point is forced

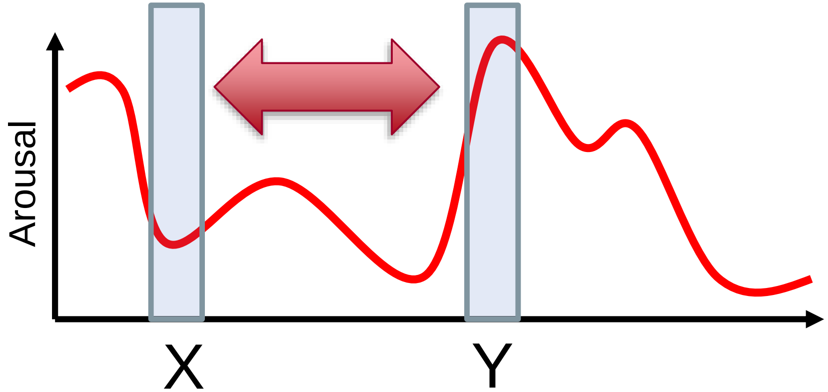
We encode values in a **relative** fashion

The **ordinal** (relative) approach

Yannakakis, Cowie, Busso, **The Ordinal Nature of Emotions: An Emerging Approach**, *IEEE Trans. on Affective Computing*, 2018





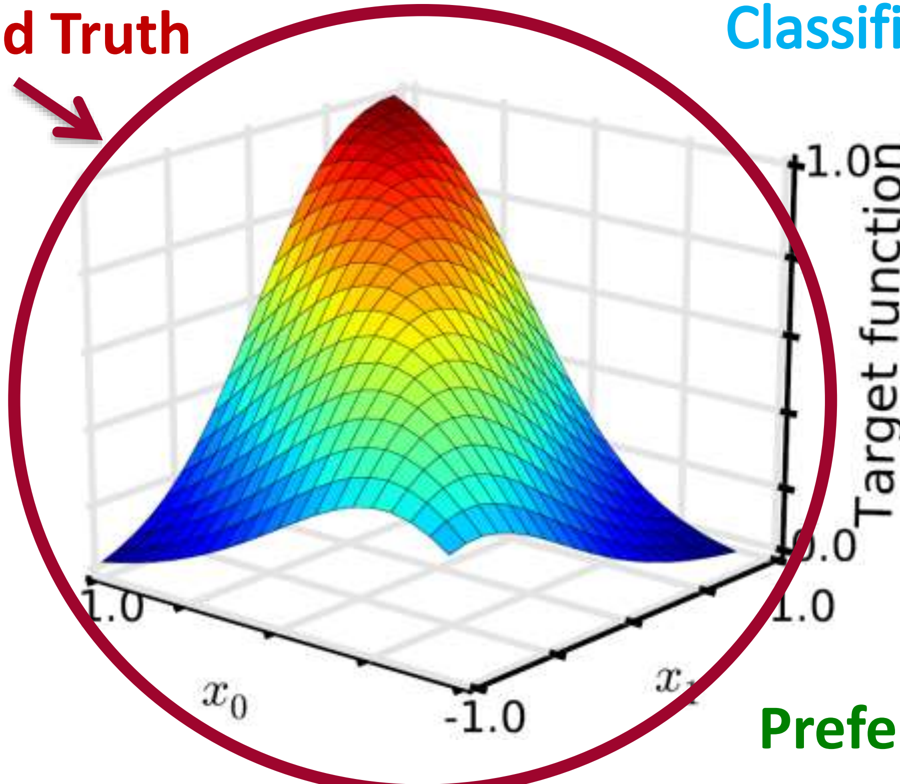


Some Evidence for the Ordinal Nature of Experience

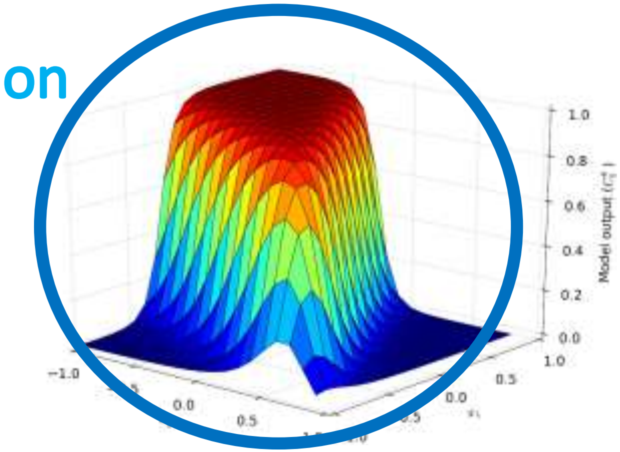
Classified Ratings vs Ranked Ratings

Martinez et al., **Don't classify ratings of affect; Rank them!** *IEEE Trans. on Affective Computing*, 2014

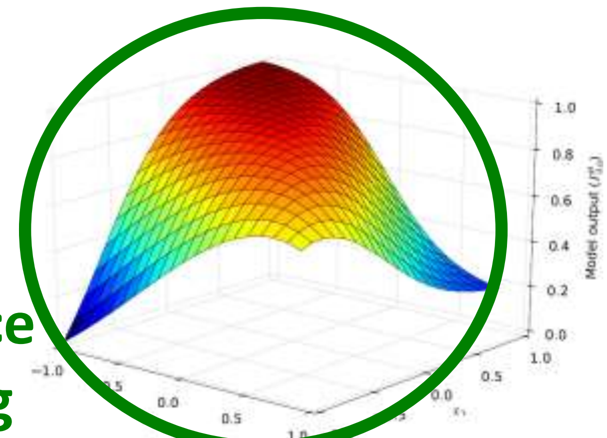
Ground Truth



Classification

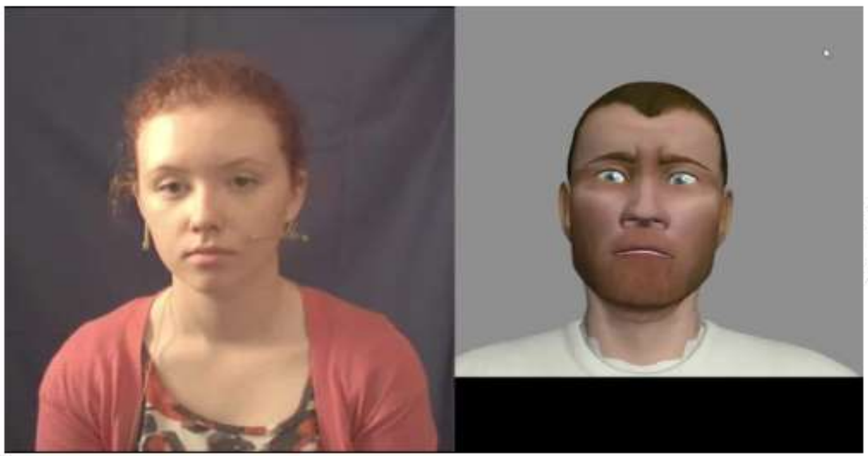


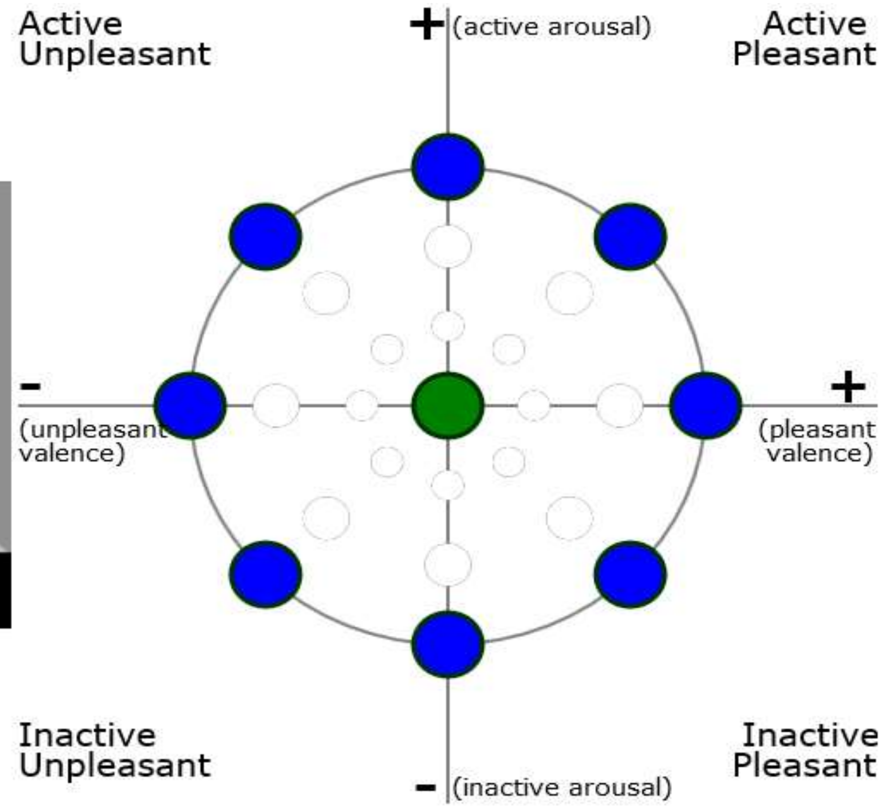
Preference Learning



AffectRank: Ordinal Emotion Annotation

Yannakakis and Martinez, **Grounding Truth via Ordinal Annotation,** *ACII*, 2015.





https://github.com/TAPeri/AffectRank

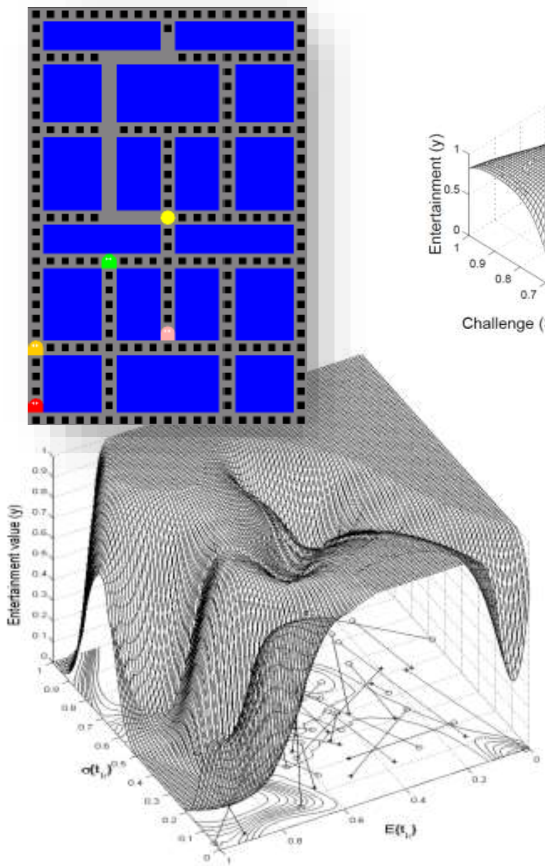
RankTrace: Relative Unbounded Annotation

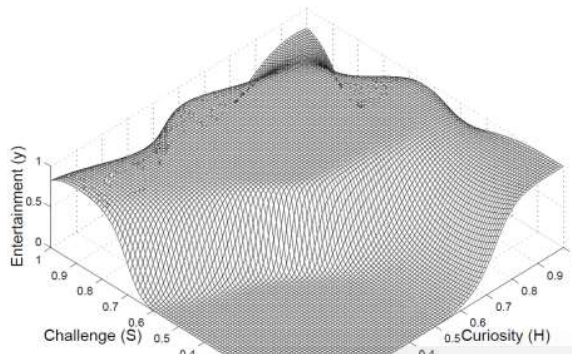
Lopes et al., **RankTrace: Relative and Unbounded Affect Annotation** *ACII*, 2017.

Tools @ emotion-research.net

Ratings (Likert) vs Preferences (Ranks)

Yannakakis and Hallam, **Rating vs. Preference: A comparative study of self-reporting**, ACII, 2011
Yannakakis and Martinez, **Ratings are Overrated!** *Frontiers in Human-Media Interaction*, 2015





X is engaging

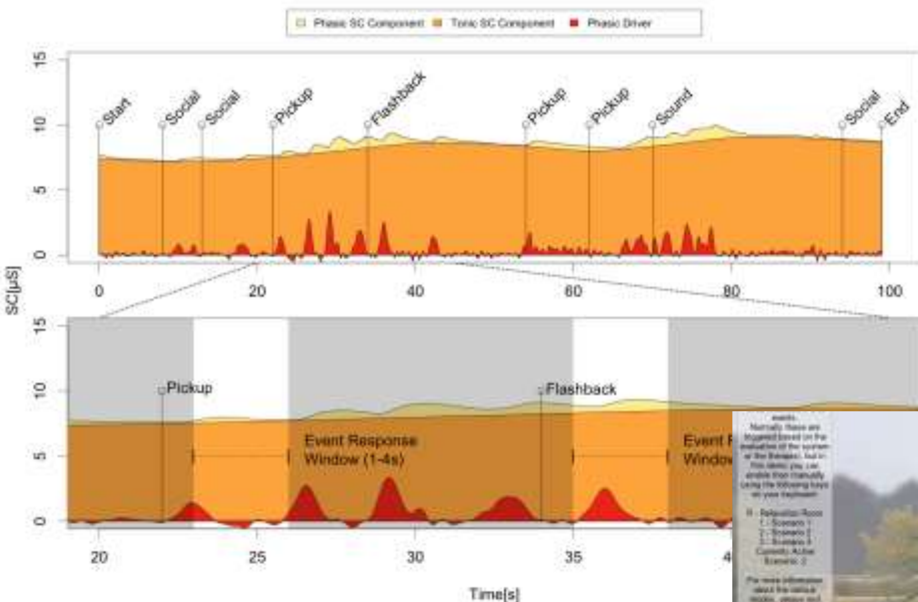
Disagree — Agree

-2 -1 0 1 2

☐ X is **more/less** engaging than Y
☐ Both are **equally** engaging
☐ **Neither** is engaging

Stress Annotation: Classes vs Preferences

Holmgård, et al **"To rank or to classify? Annotating stress for reliable PTSD profiling"**, ACII, 2015



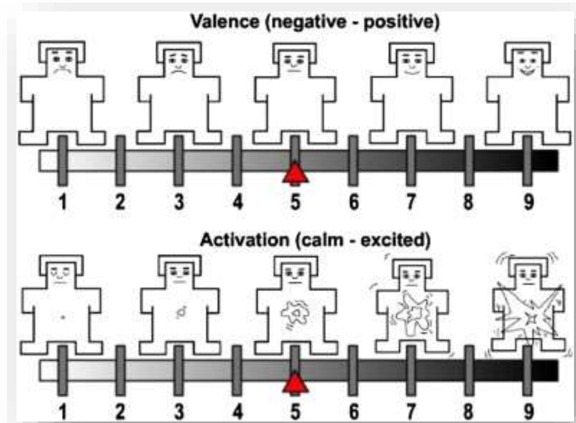
GAMES FOR HEALTH

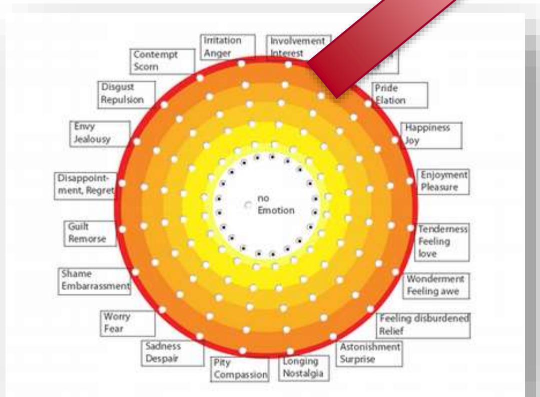
Ratings (and Classes) vs. Ranks

Martinez, Yannakakis and Hallam, **Don't classify ratings of affect; Rank them,** *IEEE Transactions on Affective Computing*, 2014

Treat ratings as **ordinal data : rank them**!

X is **more/less** challenging / frustrating / arousing / boring / fearful / … than **Y**





X was challenging

Strongly Disagree — Strongly Agree

0 1 2 3 4 5

To sum it up: **don't** do this…

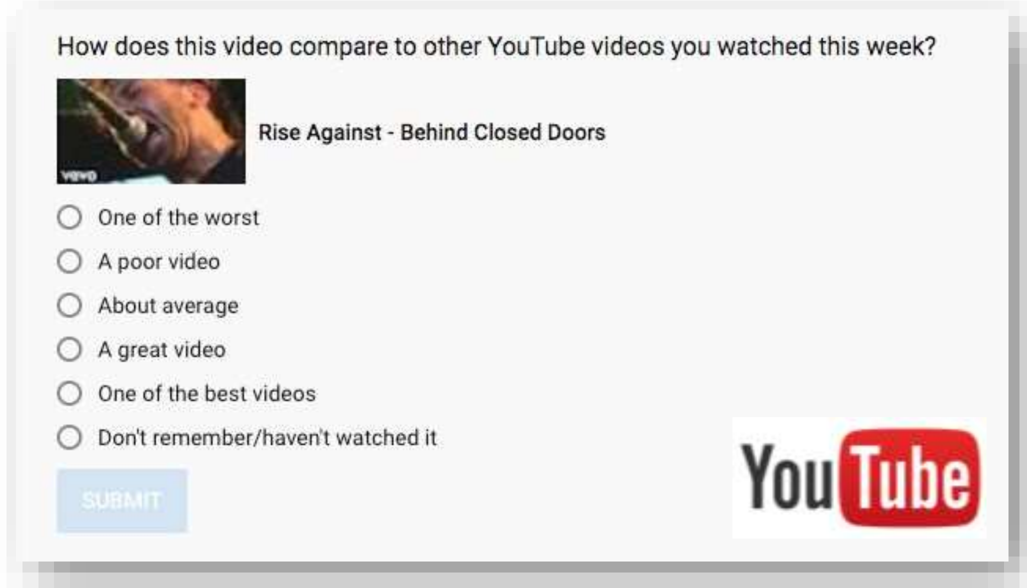
- Wasteful Info due to
  - Scale-bias
  - Personal-bias
  - Labels are **NOT** numbers
  - High inconsistency (randomness)
  - …

Do this instead…

- ☐ I like **Julian's** class more/less than **Georgios'** class
- ☐ I like them both equally
- ☐ I like neither

You gain on:

- Reliability
- Validity
- Generality


How does this video compare to other YouTube videos you watched this week?

Rise Against - Behind Closed Doors

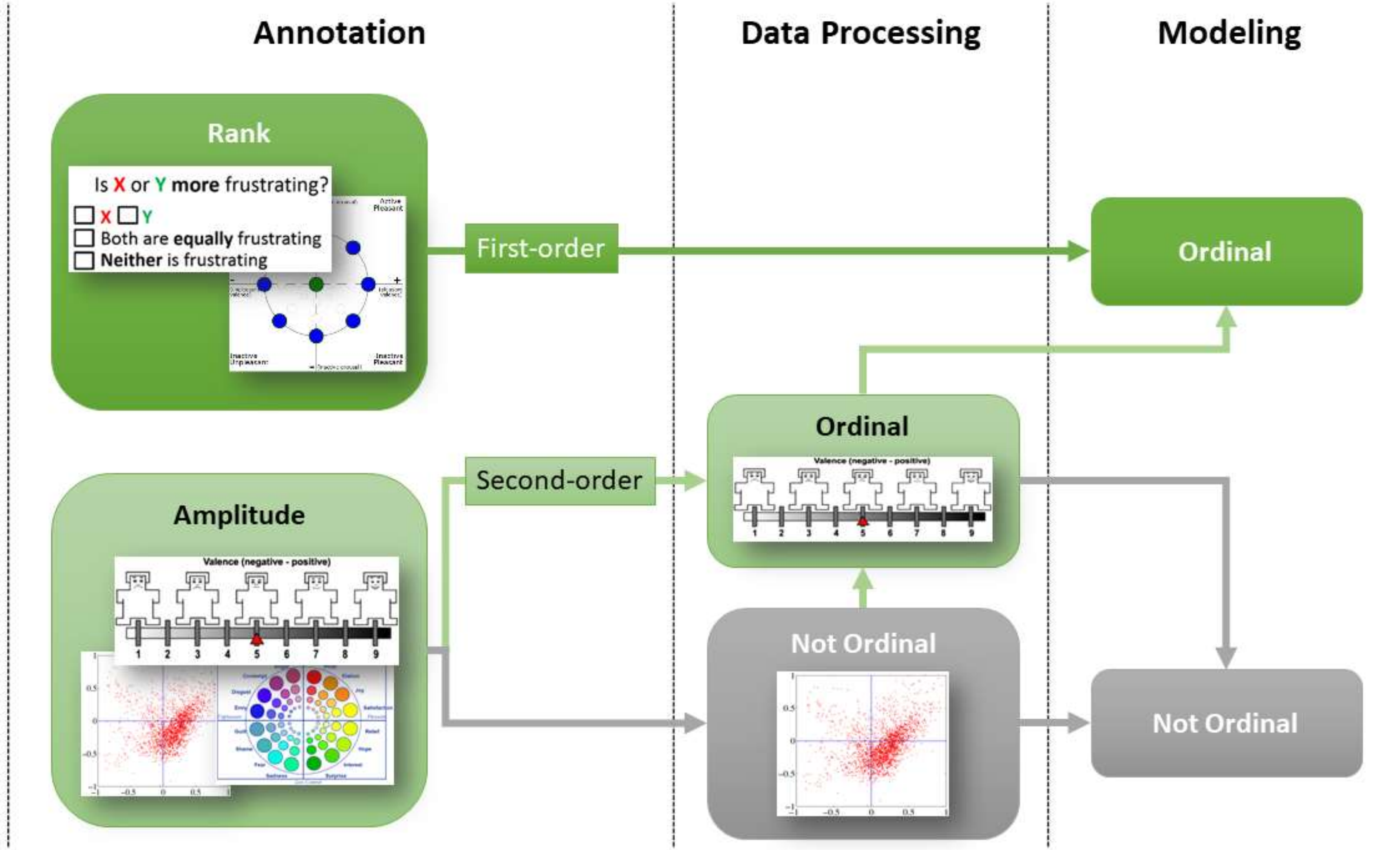
- One of the worst
- A poor video
- About average
- A great video
- One of the best videos
- Don't remember/haven't watched it

SUBMIT

YouTube

An Ordinal Perspective

Yannakakis, Cowie, Busso, **The Ordinal Nature of Emotions: An Emerging Approach**, *IEEE Trans. on Affective Computing*, 2018

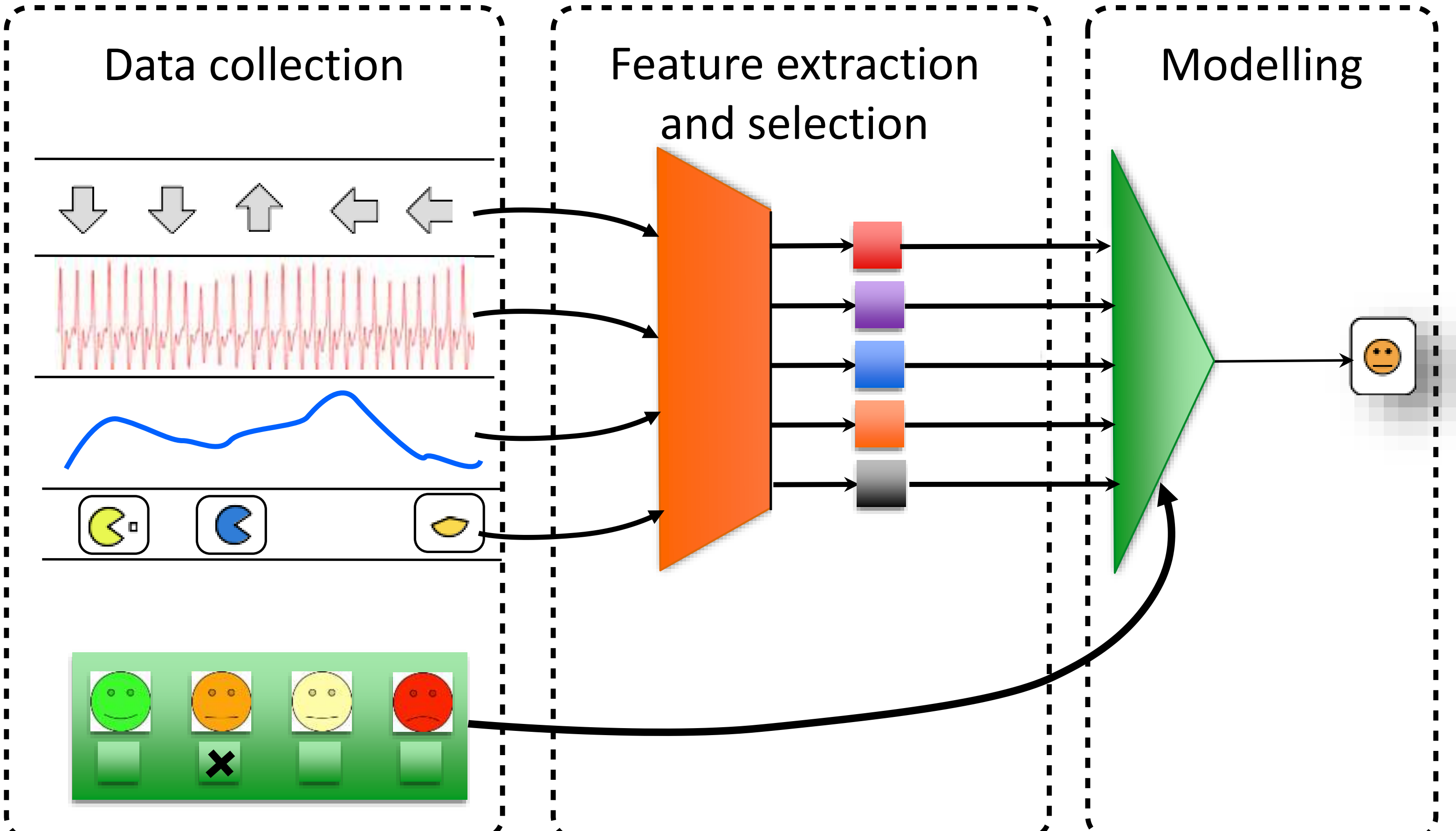


Annotation – Take away messages

- 1st vs. 3rd person: depends on the application
- Try to get reports as close to the *true experience* as possible (time-wise)
- No report is ideal (they suffer from biases)
- Annotate experience as **ranks** whenever possible
- If ratings are available
  - Regression of ratings is **fundamentally wrong**
  - Do not convert them to classes – it will cost you on model performance
  - **Convert them to ranks** (treat them as ordinal scales)!

How Can we Model Players?

Supervised Learning



Supervised Learning

- The output of the model is the *estimated experience*
- The **ground truth** is given by annotated experience given as
  - Nominal values (e.g. sample A is frustrated)
  - Numerical values (e.g. sample A is 0.86 frustrated)
  - Ordinal values
    - Ranks (e.g. sample A is more frustrating than sample B)
    - Ratings (e.g. sample A is ‘extremely frustrating’ and sample B is ‘fairly frustrating’

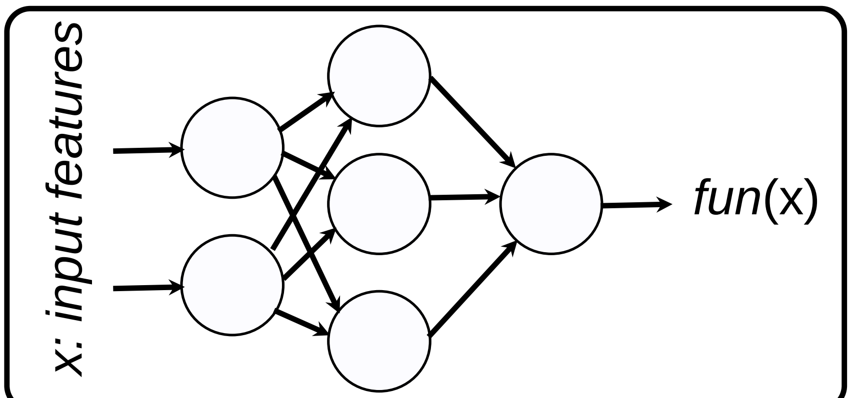
Which Training Method?

Preference learning | Classification | Regression

Example: modeling *fun* ratings

How much *fun* was that game?

1	2	3	4	5
not at all	slightly	moderately	fairly	extremely



Reminder: Backpropagation

(1) Initialise to random weights

(2) For each training pattern p:

  (a) Present input pattern $\vec{x}^{(p)}$

  (b) Compute output(s) using forward mode

  (c) Compute output error $E^{(p)}$

  (d) Compute error derivatives $\frac{\partial E}{\partial w_{jk}}$

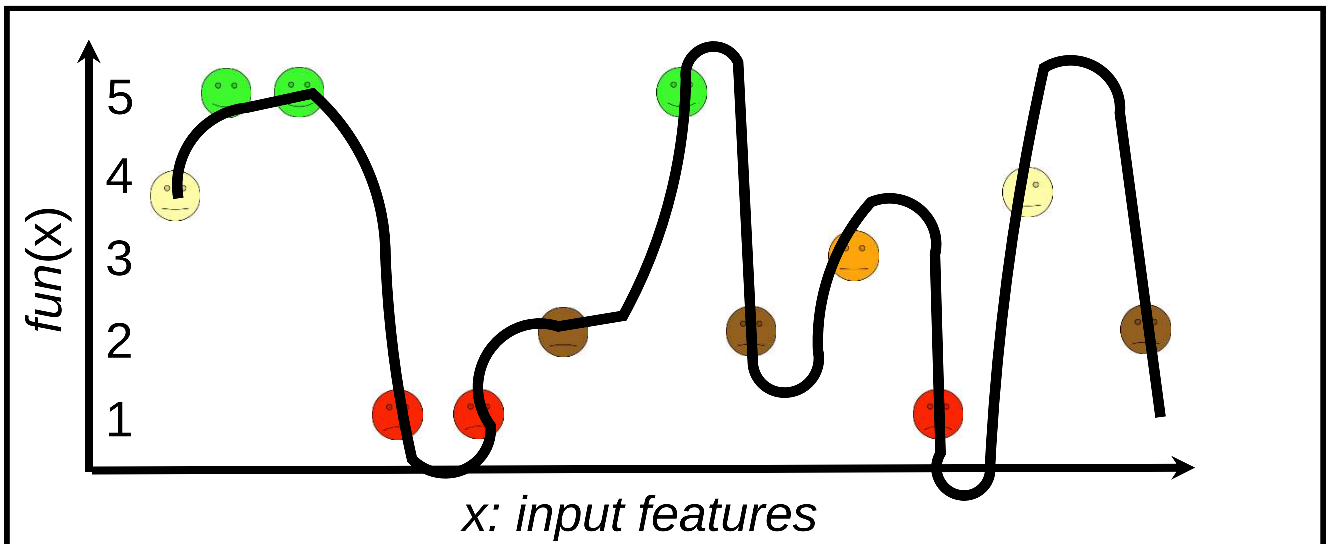
  (e) Update weights by $-\eta \frac{\partial \mathrm{E}}{\partial w_{ij}}$

(3) Is error small?

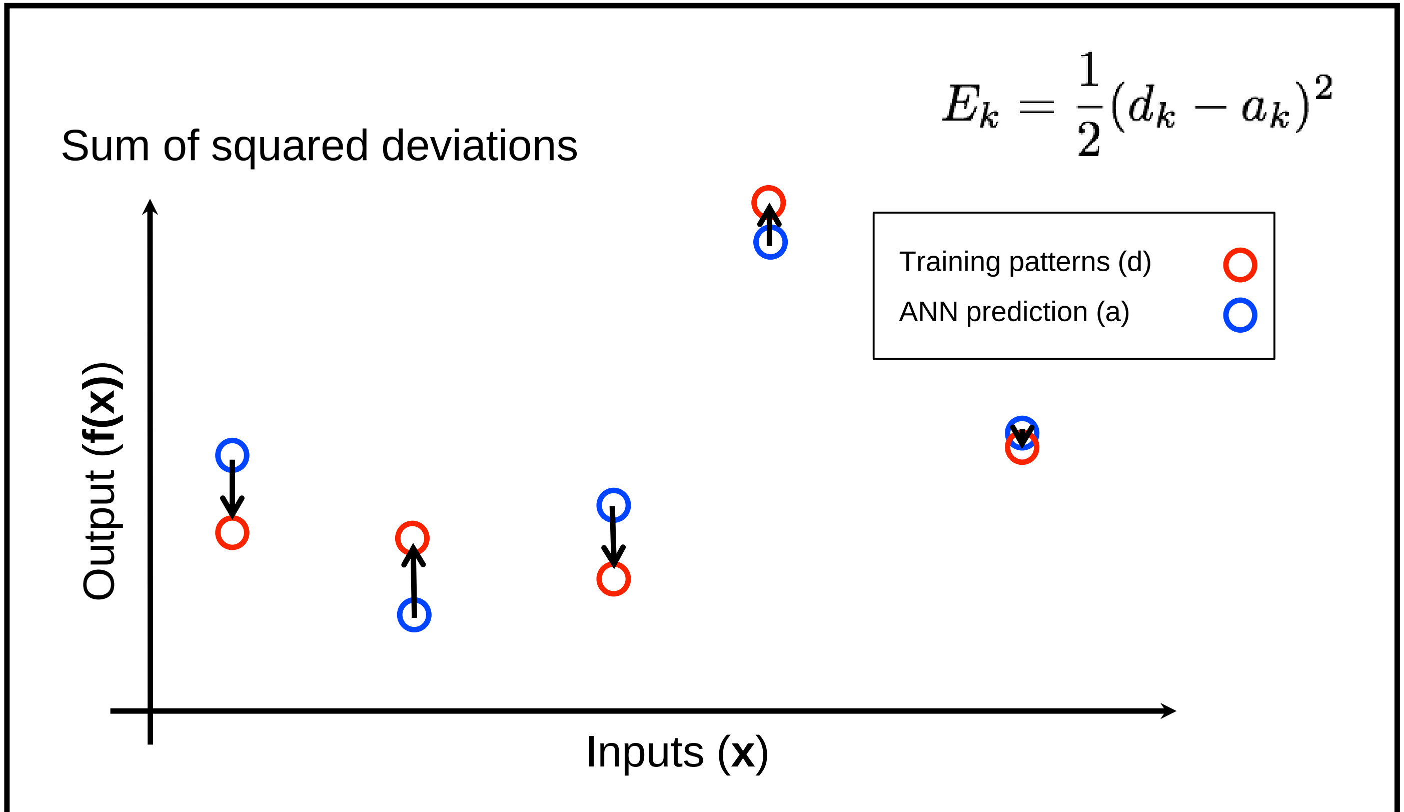
- Yes: then STOP
- No: loop to step (2)

The bad: Regression

- Remember: ratings are **NOT** numbers!
  - Not everyone uses scales in the same way
  - Items in the scale are not equidistant

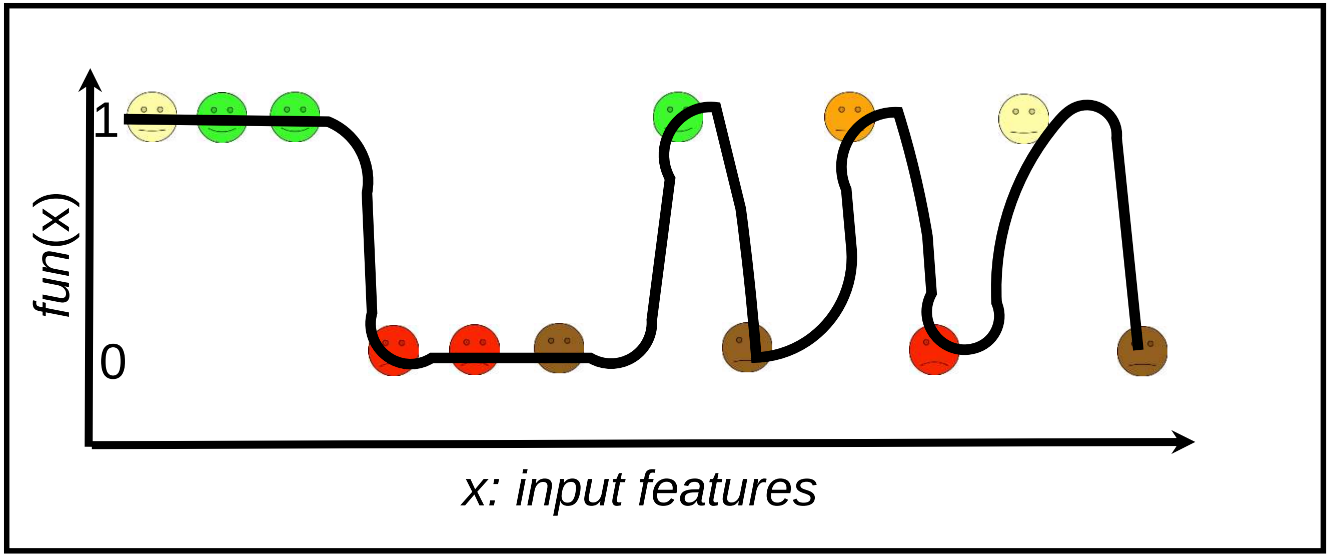


Regression with Backpropagation



The **ugly**: Classification

- Converting ratings into classes eliminates a lot of information and it can introduce biases

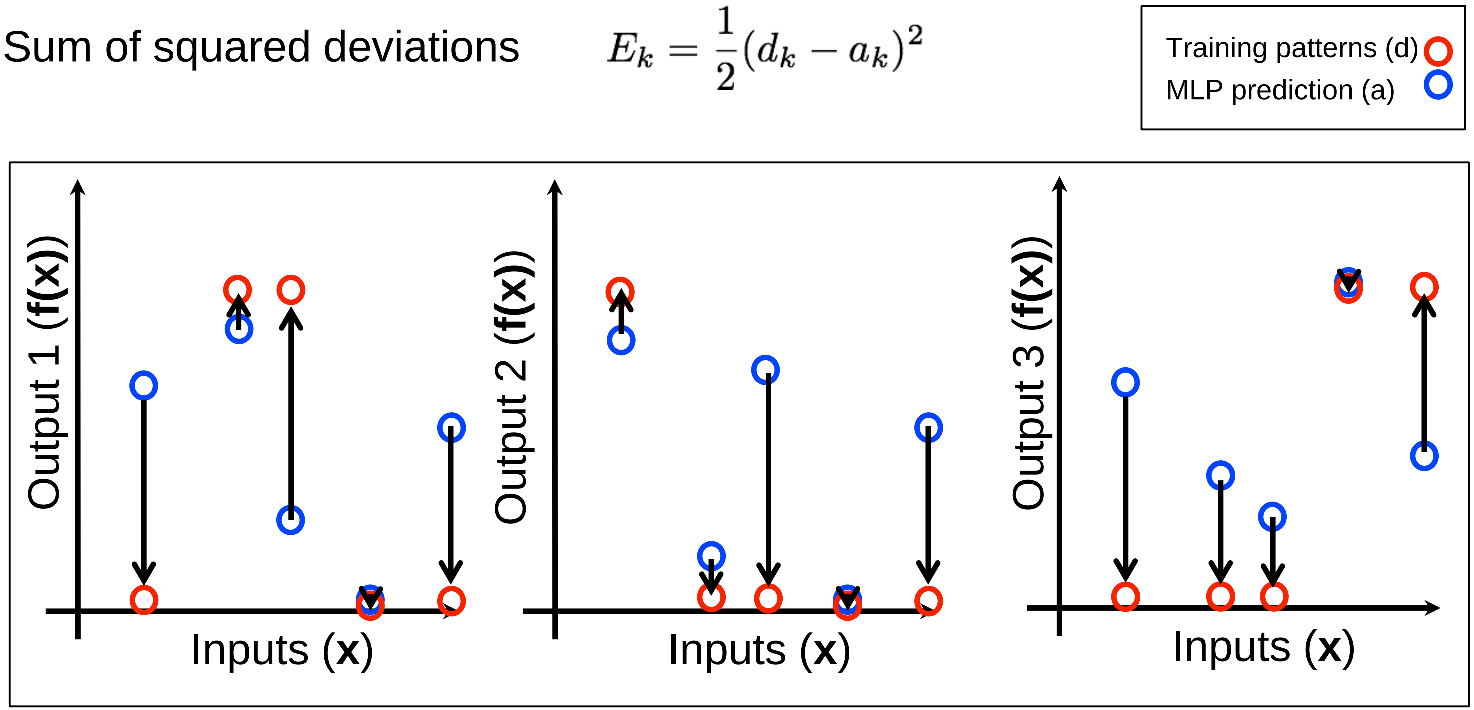


H. P. Martinez, G. N. Yannakakis and J. Hallam, "**Don't Classify Ratings of Affect; Rank them!**," *IEEE Transactions on Affective Computing,* 2014

Classification with Backpropagation

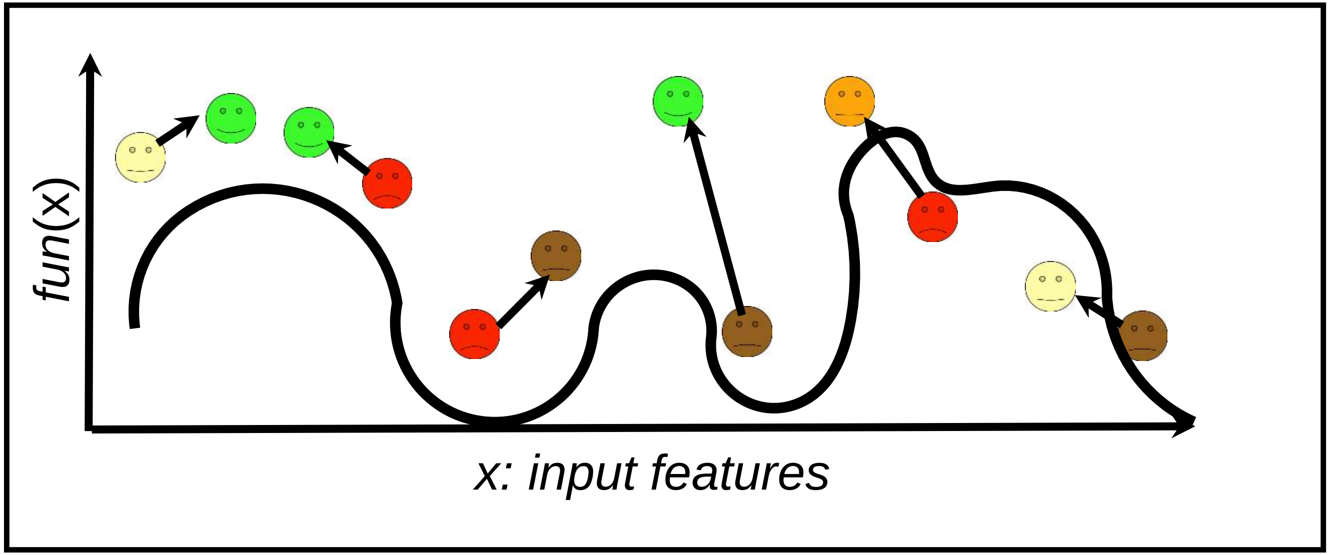
- Same as regression but with one output per class

Sum of squared deviations $E_k = \frac{1}{2}(d_k - a_k)^2$



The good: Preference Learning

- Learn only the ordinal relations
- Valid whenever the annotator is consistent on her use of the scale

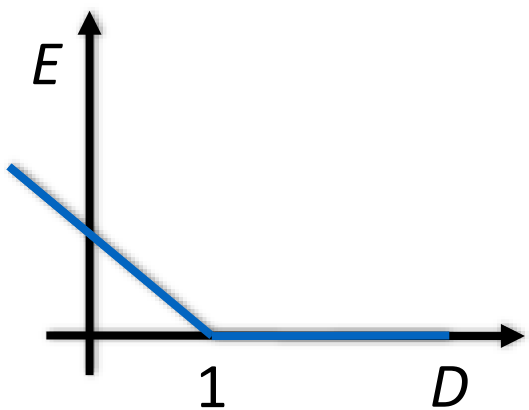


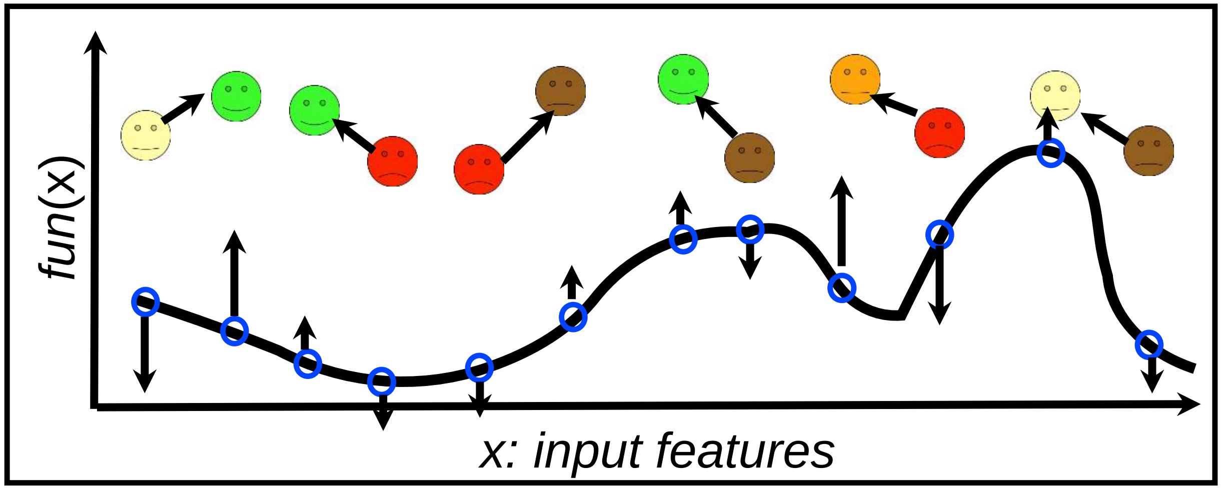
(Deep) Preference Learning

- Error function maximizes the distance between the output for the preferred sample (d^A) and the output for the non preferred sample (d^B)

$$E = max(0, 1 - (d^A - d^B))$$

$$\frac{\partial E}{\partial w_{ij}} = \begin{cases} -\frac{\partial d^A}{\partial w_{ij}} + \frac{\partial d^B}{\partial w_{ij}} & \text{, if } d^A - d^B < 1 \\ 0 & \text{, otherwise} \end{cases}$$



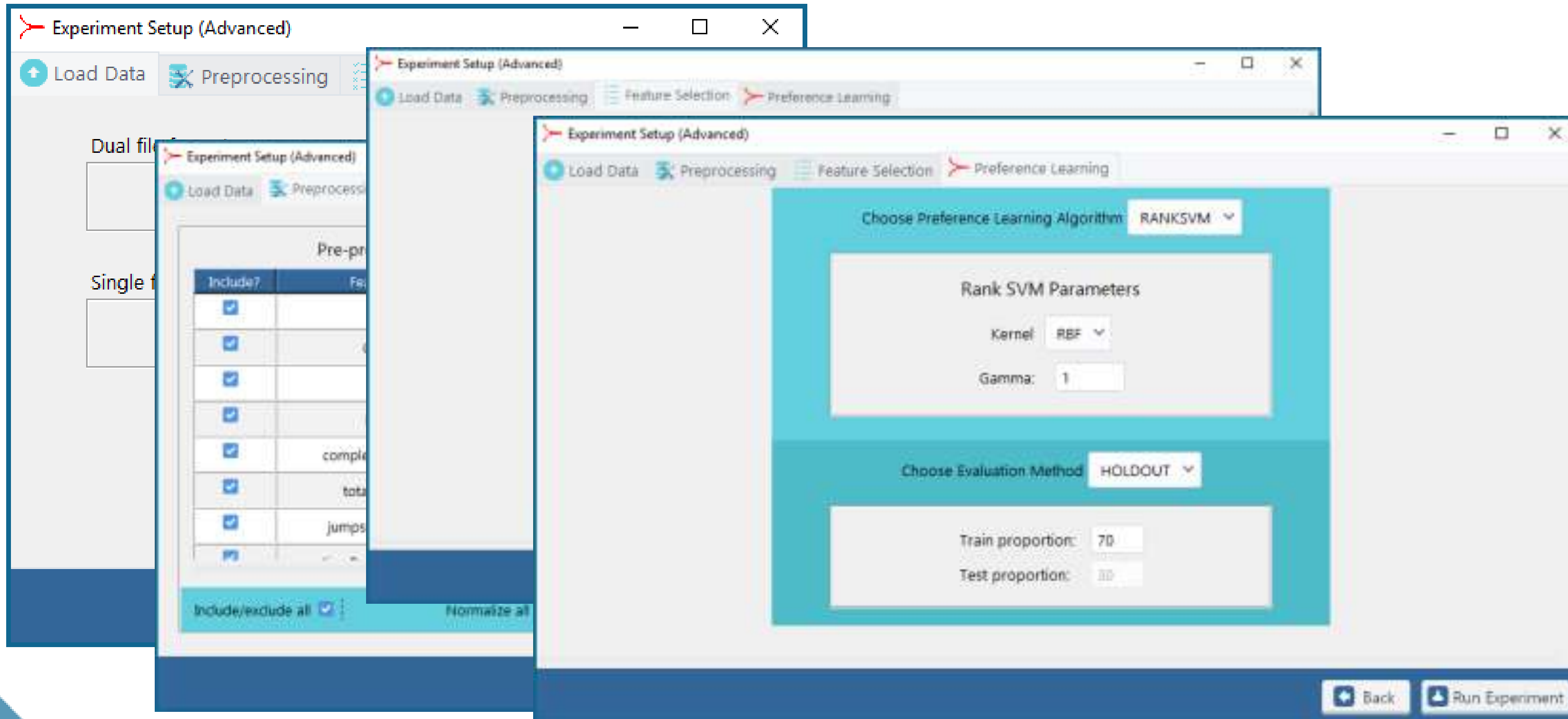


(Deep) Preference Learning Beyond BP

- The concept of **learning from pairs of preferences** can be implemented in most supervised learning methods by adapting the error/fitness function
  - Neuroevolution
    - Fitness that rewards match of pairs
  - Rank-based ANN (RankNet)
  - SVMs (RankSVM)
  - Decision Trees
  - ...

Preference Learning Toolbox

Farrugia, Martinez and Yannakakis, **The Preference Learning Toolbox**, arXiv preprint, 2015



http://plt.institutedigitalgames.com/

Preference Learning Examples

Player Experience Modeling in Super Mario

- 327 subjects (1308 games)
- Input: Playing Behavior and Content Features
- Output: Engagement, Frustration, Challenge self-reported ranks (pairwise) of short games
- ANN trained via **Neuroevolutionary Preference Learning**
- Player experience model accuracy: **73-92**%

Super Mario Bros Example:
The Annotated Experience (ANN output)

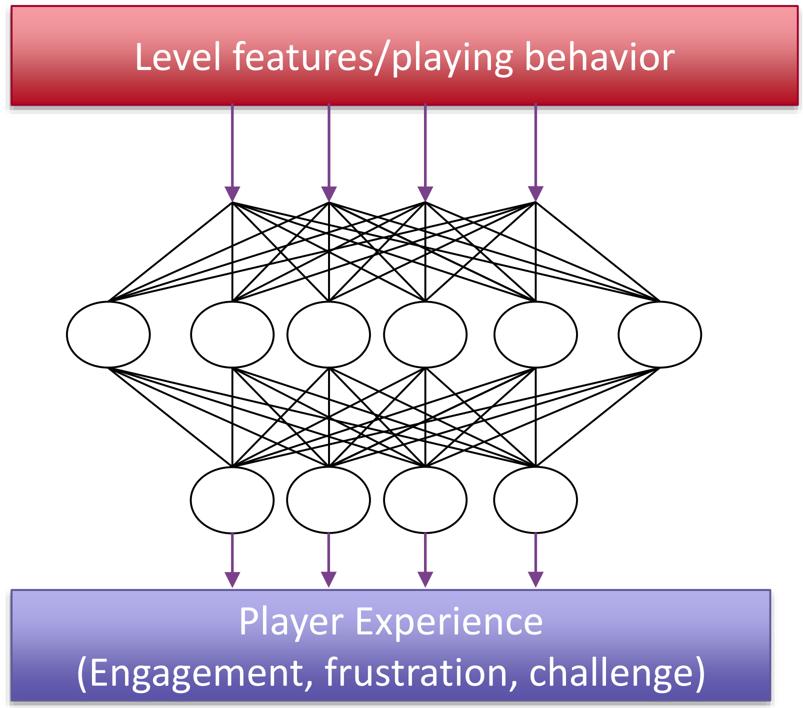
Shaker, Asteriadis, Yannakakis and Karpouzis, **Fusing Visual and Behavioral Cues for Modelling User Experience in Games,** *IEEE Trans. on Systems, Man and Cybernetics* (B), 2013

- Three Player experience states modelled:
  - *Engagement*, *Frustration*, *Challenge*
- Player Experience is self-reported (post-experience) via a 4-alternative forced choice questionnaire:

☐ **Game A** is more/less **engaging** than **Game B**
☐ Both are equally **engaging**
☐ Neither is **engaging**

Super Mario Bros Example: The Modeling Approach

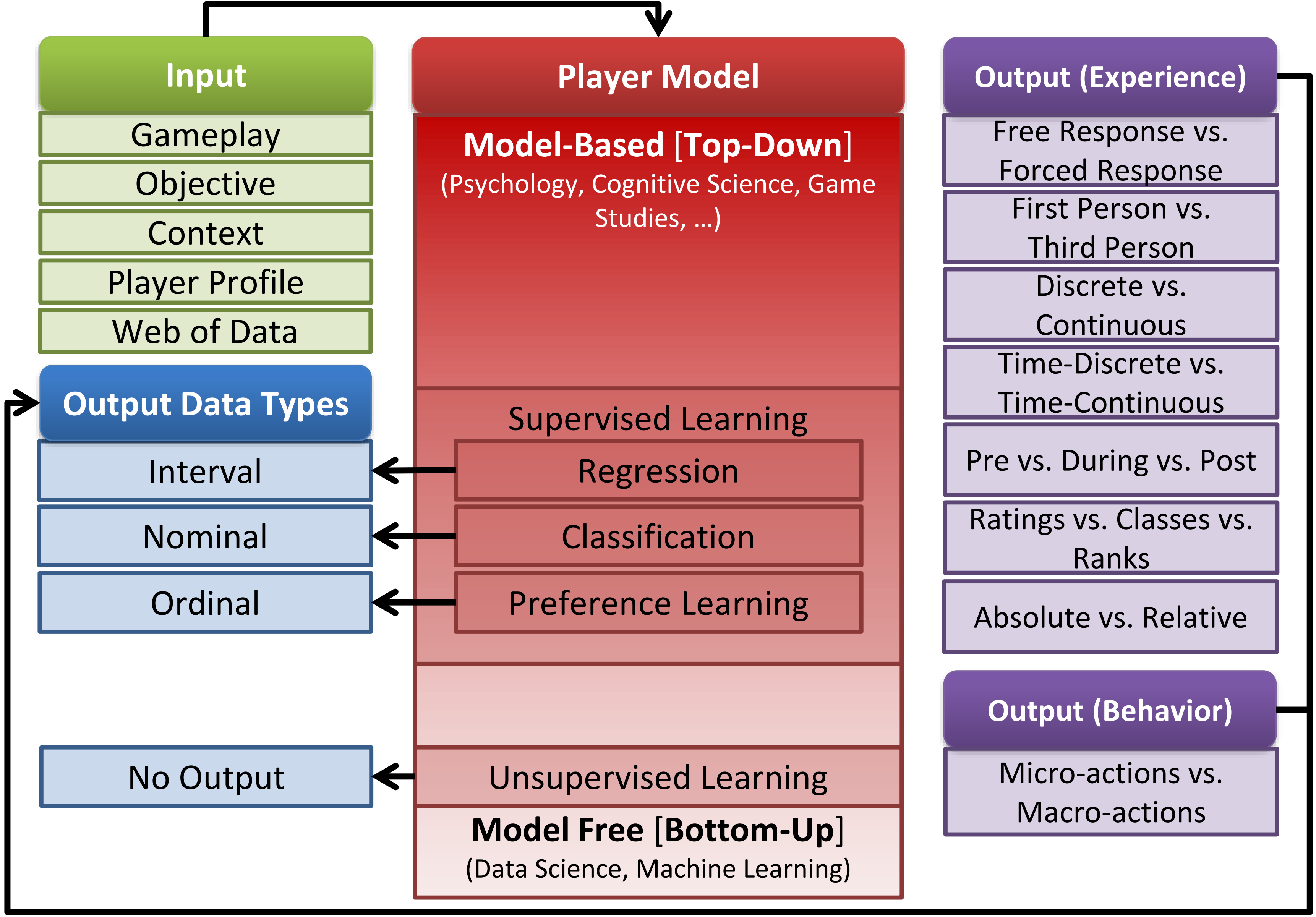
Shaker, Asteriadis, Yannakakis and Karpouzis, **Fusing Visual and Behavioral Cues for Modelling User Experience in Games,** *IEEE Trans. on Systems, Man and Cybernetics* (B), 2013

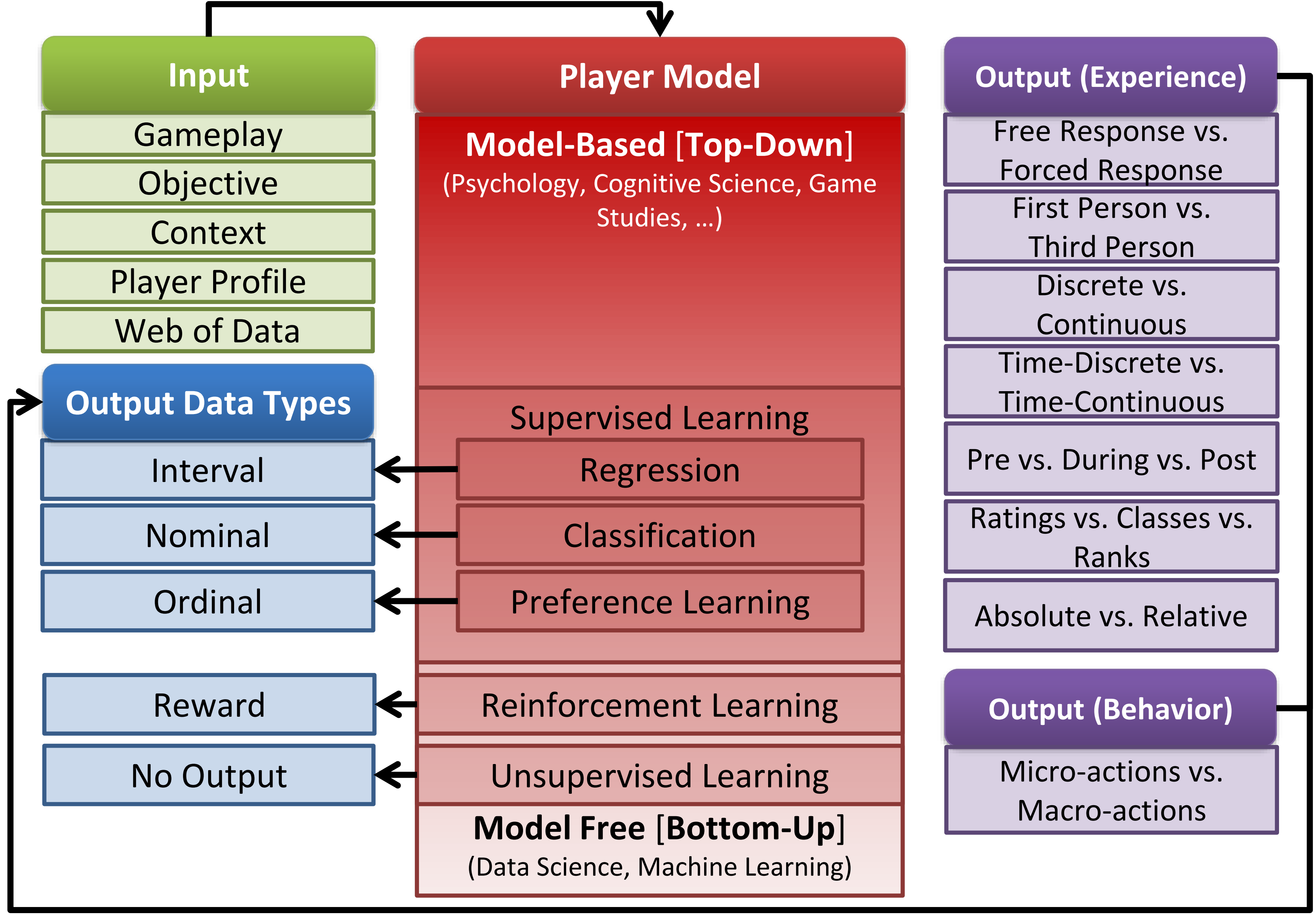


Platformer Experience Dataset

K. Karpouzis, G. Yannakakis, N Shaker, S. Asteriadis. **The Platformer Experience Dataset**, Sixth Affective Computing and Intelligent Interaction (ACII) Conference, 2015.

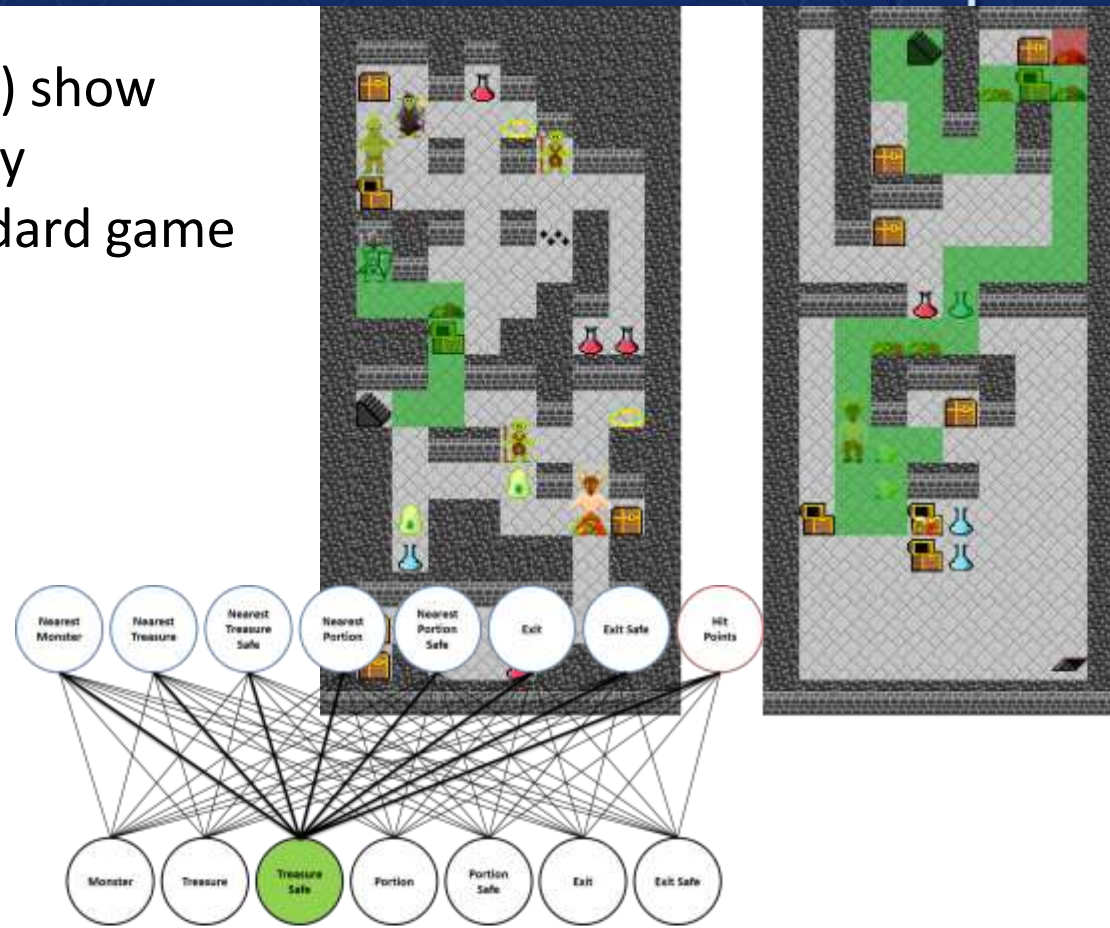
http://ped.institutedigitalgames.com/

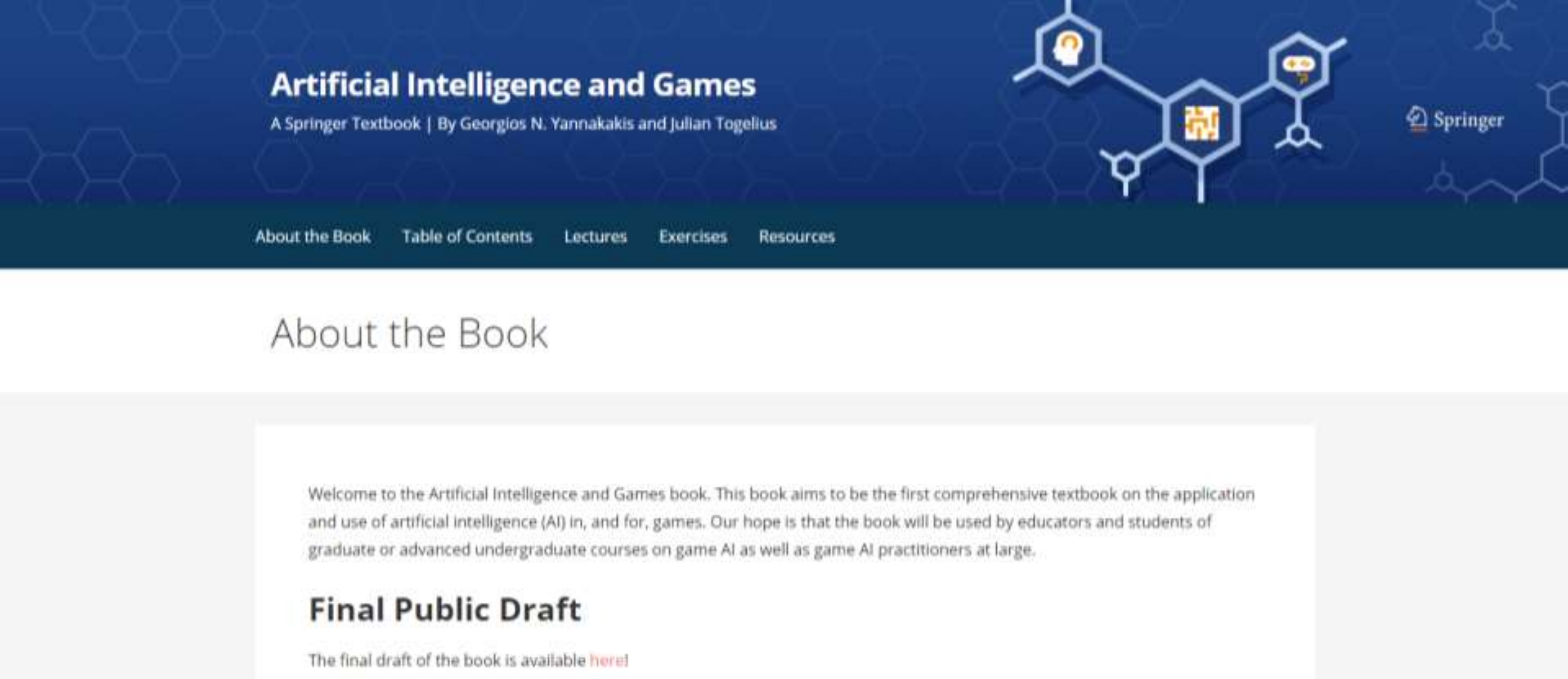

Input
Gameplay
Objective
Context
Player Profile
Web of Data
Output Data Types
Interval
Nominal
Ordinal
No Output
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, ...)
Supervised Learning
Regression
Classification
Preference Learning
Unsupervised Learning
Model Free [Bottom-Up]
(Data Science, Machine Learning)
Output (Experience)
Free Response vs. Forced Response
First Person vs. Third Person
Discrete vs. Continuous
Time-Discrete vs. Time-Continuous
Pre vs. During vs. Post
Ratings vs. Classes vs. Ranks
Absolute vs. Relative
Output (Behavior)
Micro-actions vs. Macro-actions


Input
Gameplay
Objective
Context
Player Profile
Web of Data
Output Data Types
Interval
Nominal
Ordinal
Reward
No Output
Player Model
Model-Based [Top-Down]
(Psychology, Cognitive Science, Game Studies, ...)
Supervised Learning
Regression
Classification
Preference Learning
Reinforcement Learning
Unsupervised Learning
Model Free [Bottom-Up]
(Data Science, Machine Learning)
Output (Experience)
Free Response vs. Forced Response
First Person vs. Third Person
Discrete vs. Continuous
Time-Discrete vs. Time-Continuous
Pre vs. During vs. Post
Ratings vs. Classes vs. Ranks
Absolute vs. Relative
Output (Behavior)
Micro-actions vs. Macro-actions

Procedural Personas

- Given utilities (rewards) show me believable gameplay
- Useful for human-standard game testing
- RL
  - MCTS
  - NeuroEvolution
  - …
- Inverse RL





Thank you!

gameaibook.org