The State Machine and the Present Moment

The perception of time is based on change. Therefore, the first thing a medium needs in order to be time-based is the capacity to change states. Each change of state is an event. The simple act of a pixel altering its color, or its location in space, allows us to perceive the difference between the earlier and the later state, and thus infer that time has passed.

Time is always passing in the real world, but in mediated worlds this is not necessarily the case. Different media represent time in different ways,[1] and these representations can vary from our everyday experience of it. Photographs, for instance, are still images that freeze a particular moment in time. Movement in this medium can be implied, but never actually shown. Thus, a speeding bullet piercing through an apple becomes an image of a bullet floating still next to an apple with two bursting orifices: one right behind the projectile and the other one on the opposite side of the fruit. From this image, we can infer that the bullet came in through one side of the apple and tunneled its way to the other at high speed.

Film offers the capacity to record and display a fragment of time by taking and then projecting sequences of static photographs at fast rates, causing us to perceive one continuous moving image. After the popularization of the moving image through film, television brought it into the living room and, more recently, digital video made it possible for moving images to be transmitted through the Internet into all sorts of devices, some of which can even be carried around in our pockets.

With moving image technologies, the capacity to slow down or speed up the passage of time is afforded, with slow motion and time-lapse techniques respec-

1 For a comparative analysis of the treatment of space and time in film and games (and other audiovisual media) see Wolf 2002b (pp. 77-80) and Freyermuth 2015 (pp. 131-139).

tively. By dilating time, we can see the heaving cheek of a wrestler as it is struck by its opponent's knuckles, or a corn seed slowly bursting into popcorn and rising from the pan into the air in a mist of floating oil drops. Time can also be compressed in order to witness events so slow that they would otherwise appear static. As a result, we can experience the long-drawn growth of a plant in just a few seconds.

Just like film and television, the video game is a member of this family of moving images, and it commonly uses the same display technology as the latter: a screen composed of a matrix of dots called a raster. These come in different forms, from the CRT screens used to play PONG (Atari 1972) in the '70s, to current OLED screens, and can display sequences of frames at high speeds that we perceive as a constant, moving image.

APPARENT MOTION

The moving image is a temporal illusion. It is what psychologists Ramachandran and Anstis call "apparent motion," that is, perceiving an "intermittently visible object as being in continuous motion" (Ramachandran and Anstis 1986, p. 102). Real motion, in contrast, is the perception of continuously moving stimuli in the visual field. A football match watched in a stadium constitutes real motion, whereas the same event watched on TV is in apparent motion. Since movement in video games is apparent, this section focuses exclusively on this perceptual category.

It was commonly believed that we are able to perceive movement from a series of still images thanks to a phenomenon named *persistence of vision*. According to this account, the light irradiating from the screen leaves an impression on our retina that lasts for a few fractions of a second after the source image disappears. The retina is therefore unable to register the changes in light that happen between the moment it gets stimulated by light and the end of this short time frame. This, so the theory goes, is the reason why we do not see the blackness in-between movie frames and the image projected onto the screen seems to be in motion. Persistence of vision approximates what perceptual psychologists call the *flicker fusion threshold*, but it merely solves the problem of why the image on the screen seems continuous instead of being perceived as a blistering slideshow (Anderson and Anderson, 1993, p. 4). The problem arises with the claim that the illusion of movement is a result of persistence of vision, a pervasive misunderstanding in some academic quarters in the previous century, espe-

cially in film studies (compare Anderson and Fisher 1978, and Anderson and Anderson 1993).

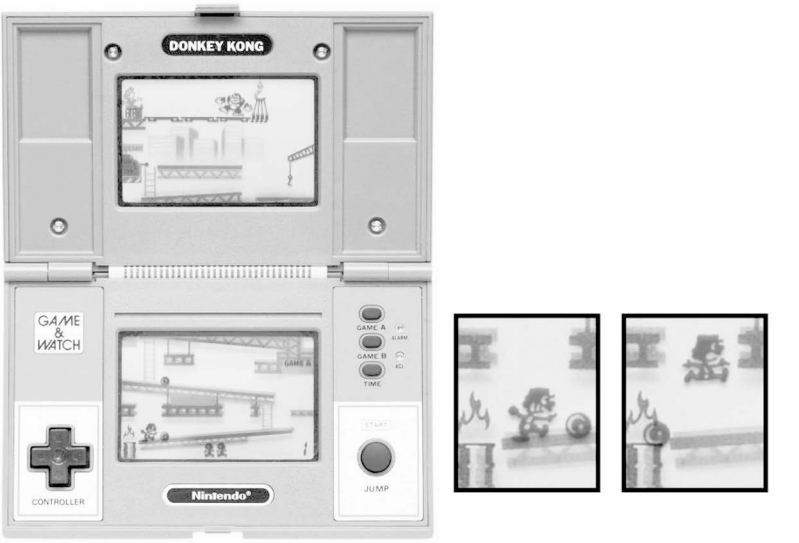
Back in 1978, Joseph and Barbara Anderson (née Fisher) wrote a paper entitled THE MYTH OF PERSISTENCE OF VISION, in which they rebuked the notion that this phenomenon can "account for the perception of the successive frames of a motion picture as a continuously moving image" (Anderson and Fisher 1978, p. 6). A decade and a half later, they revisited this argument in the aptly-titled paper THE MYTH OF PERSISTENCE OF VISION REVISITED (Anderson and Anderson 1993). The motivation to do so was not to make amendments to the original rationale, but to insist upon it, since the mistaken belief persevered despite their initial efforts to debunk it.

In these papers, Anderson and Anderson point out that the notion of persistence of vision originated from the misrepresentation of an article by the British physician Peter Mark Roget and the assumptions of 19th-century psychologists. Anderson and Anderson also stress that persistence of vision could not possibly account for why we perceive movement. Even without noticing the blanks in-between the frames of a film, we could just as well experience something reminiscent of Marcel Duchamp's "Nude Descending a Staircase," where images from consecutive moments in time would superimpose in overlapping layers (Anderson and Anderson 1993, p. 4). But while we are fully aware that, when watching film or television, we are looking at a series of images in very rapid succession,[2] the illusion overcomes our perception and we only see one continuous image in motion. When Forrest runs, we see him running. When Mario jumps, we see him jumping.

One clear indication that the perception of movement occurs at the level of the brain (and not the retina) is the rare condition known as *akinetopsia* or motion blindness, which appears as a result of a lesion to the visual cortex (Zeki 1991). Those suffering from this ailment in its most extreme form cannot perceive movement. They can see stationary entities, but these vanish from their perception when in motion. It is thus highly implausible that the perception of movement could arise solely as a retinal phenomenon. Movement occurs within the brain. This has been the understanding in psychology since the early 20th century, starting with psychologists like Max Wertheimer (1912), one of the founders of Gestalt psychology.

2 There are significant technical differences between how film projectors and television sets (and even between different types of television sets) display images, which are not relevant to the current argument. The important factor in the context of this study is that both technologies work by displaying several frames per second, which enables the perception of apparent motion.

Figure 1.1: DONKEY KONG GAME & WATCH (1982).



Source: http://pica-pic.com/ (accessed June 04, 2019).

Left: The avatar is on the ground, to the bottom-left of the lower screen.
Right: Detail of two frames of the avatar jumping.

Consider for instance electronic handheld games like the GAME & WATCH[3] series released by Nintendo in the 1980s. These are portable consoles with one single game. Their rudimentary LCD screens show a few interactive elements on top of a fixed background. These elements are static black shapes, like paper cut-outs, that represent the avatar and the enemies or obstacles, as seen in figure 1.1. When pressing the jump button, the image of the avatar disappears and one with the same design but in a different pose instantaneously pops up above it. From this, the player can intuitively infer that the avatar has moved, and not that it has disappeared and reappeared a few millimeters higher. A brief moment after the jump, the figure on top vanishes and the one below is displayed once again, implying that gravity pulled the avatar back down to the ground. There is nothing smooth or continuous to this animation, but no effort is needed to perceive

3 Digitalized versions of several electronic handheld games, including some of the GAME & WATCH series, can be found on http://pica-pic.com/ (accessed August 25, 2016).

movement; the illusion simply ensues. On the contrary, an extra effort would need to be exerted to avoid seeing movement.

That this simple technology achieves an illusion of movement already hints at a constitutive feature of the human mind: Our perception is not solely dictated by the bottom-up processing of stimuli, but it also involves the implementation of top-down assumptions about the world (I will explain this in more detail in chapter two, section 2.1). The fact that the GAME & WATCH figures at the top and the bottom possess the same characteristics allows the brain to connect the dots and infer that it is seeing the same character that has changed position. The information captured by the retinas is being processed and interpreted as movement in the brain's visual cortex. If the two figures were a red square and a green triangle, the brain would have more difficulty interpreting the signals as the product of movement.

According to Ramachandran and Anstis (1986, p. 1), the first feature needed to perceive apparent motion is *correspondence*. Players can see the figure jump in DONKEY KONG because it retains its characteristics with only a slight change of pose. The consistency of features allows our visual perceptual system to detect an object and follow it in space.

Correspondence alone, however, is not enough to fully explain the perception of motion. A series of ingenious experiments by Ramachandran and coworkers (Ramachandran and Anstis 1986) have shown that the brain also makes assumptions about how objects in the world behave:

1. "Objects in motion tend to continue their motion along a straight path. The visual system perceives linear motion in preference to perceiving changes in direction" (ibid., p. 105).
2. "Objects are assumed to be rigid; that is, all points on a moving object are assumed to move in synchrony" (ibid., pp. 105-106).
3. "A moving object will progressively cover and uncover portions of a background" (ibid., p. 107).

If all of these conditions are in place, the illusion of movement ensues effortlessly through the mere act of watching. A way to challenge players is to violate these principles, making it harder for them to follow objects in motion.

This capacity of the video game medium to elicit the perception of apparent motion does not differ from what film and television can accomplish. But video games do diverge in one substantial respect from other moving images: They are *interactive*.

THE STATE MACHINE

Game scholar Jesper Juul argues that games are *state machines*, a term he borrows from computer science: "a state machine is a machine that has an *initial state*, accepts a specific amount of *input events*, changes state in response to inputs using a *state transition function* (i.e., rules), and produces outputs using an *output function*" (Juul 2005, pp. 59-61).[4] This notion applies both to analog and digital games. The state machine is thus defined by the rules, which allow for input from players and produce an output that informs them of the current state of the game.

In board or card games, the elements used to play the game keep track of the game state. In chess, these are the board and pieces; in poker, the cards on the table and in each player's hands, and the chips used to place bets. In the case of chess, the whole game state is accessible to both players. In the case of poker, the game state is partly hidden from players, who only have access to the shared information on the table and their own hand (other player's hands and the cards in the deck are hidden parts of the game state).[5] The state machine includes not only the positions of the pieces on the board or the cards and chips, but also the rules that determine what those elements can do at any time in the game. Pawns in chess can only move forward, one square at a time, but capture other pieces diagonally; to get a flush in a game of poker you need five cards of the same suit that do not form a sequence.

In single-player video games, the interaction occurs between a computer and a player. The computer runs a program (the game) that keeps track of its own states. The program informs the player of these states primarily through image and sound. What the player perceives, then, is the state of the mediated gamespace. The player processes this information, chooses a course of action, and provides the inputs required to set the strategy in motion. These inputs are

4 Programmers might find this analogy disorienting, since finite state machines are one of the programming patterns that can be used to code games (see Nystrom 2014). But the use of "state machine" in this passage shouldn't be taken too literally. Despite this potential confusion, I believe that referring to video games as state machines can be a productive analogy.

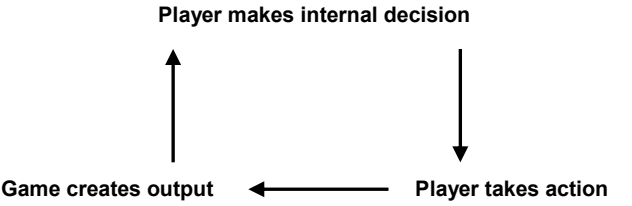
5 Mark J. Thompson distinguished between games of *perfect* and *imperfect* information. Chess does not conceal any information from players and it does not include random mechanics like dice. Therefore, it is what Thompson calls a game of *perfect* information. In poker, on the other hand, players hide their cards from their opponents, which makes it a game of *imperfect* information (Thompson 2000).

commonly made with a controller or mouse and keyboard. The program then updates its state to reflect the incoming information.

While playing a game of TETRIS, the game controls the speed at which the tetrominoes descend and the order in which different pieces appear. The program also keeps score, and draws the image and reproduces the sounds that constitute the game's output. The player can accelerate the pieces' descent, rotate them left and right in 90-degree angles, and can determine their position on the horizontal axis. In more complex games, the variables controlled by both the player and the program itself can be much higher in number.

Game designers Katie Salen and Eric Zimmerman summarize the process of interacting with a game (whether analog or digital) in a concise three-part diagram:

Figure 1.2: Gameplay loop by Salen and Zimmerman.



Source: Salen and Zimmerman 2004, p. 316.

Complex states can arise from this gameplay loop even in games in which players have a minimal range of action—like TETRIS. This is the concept of *emergence* (Schell 2008, pp. 140-143; Salen and Zimmerman 2004, 163-165). With just a simple set of moves, players can participate in the formation of intricate patterns of action and reaction from which new game states emerge. The combination of simple elements can shape the state machine in unexpected ways. Salen and Zimmerman note that "[a] successfully emergent game system will continue to offer new experiences, as players explore the permutations of the system's behavior" (2004, p. 165).

While the computer runs the game, processes player input, and sends output signals with its state-changing peripherals, the player's mind is continually constructing a moving picture of the state of the game. This picture is our conscious experience, which takes place in the window of time we call the present.

THREE MOMENTS IN TIME

The *now* in which we live our lives and play our games is a laborious mental construction. It is not a slice in the constant flow of time, but rather a window that integrates both the proximate future and the immediate past. Husserl called the traces of the past *retention* and the anticipation of future events *protention*. Philosopher Dan Lloyd exemplifies this with the Beatles' song "Hey Jude":

"[Experiences] occur and leave their traces in 'retention,' like a comet's tail. As Paul McCartney lands on 'Jude,' the 'Hey' is retained though no longer sensed. Likewise, as 'Jude' sounds, we anticipate something to follow ('Don't make it bad,' if one knows the song, or something less definite) – this is 'protention'" (Lloyd 2012, p. 696).

Building on a dual taxonomy first proposed by Ernst Pöppel (1997), psychologist Marc Wittmann speaks of three *moments in time* (table 1.1) that constitute our experience of the present: The *functional moment*, the *experienced moment*, and *mental presence* (Wittmann 2011; 2015, pp. 54-62). These perceptual units are nested like a Russian doll of temporal awareness, with their boundaries demarcated by different time frames that span from mere milliseconds to a few minutes.

Table 1.1: The three moments of present experience.

Moment	**Duration**
Functional Moment	Milliseconds 30 ms – 300 ms
Experienced Moment	Few seconds 300 ms up to 3 seconds
Mental Presence	Several seconds up to a few minutes

Source: Wittmann 2015, p. 58.

The Functional Moment

The *functional moment* is the basic building block of our temporal consciousness. It is the level at which the brain discerns if events happen simultaneously or non-simultaneously. At this stage, the brain operates in spans of milliseconds.

Given that the brain needs time to process stimuli, our experience of the present can lag behind the actual occurrence events for hundreds of milliseconds. In addition, different senses operate at different temporal resolutions: Auditory signals function at the highest temporal resolutions, meaning that they are transduced quickly, whereas visual signals require the longest time, and thus operate at the lowest temporal resolution. Acoustic stimuli that are merely two or three milliseconds apart can be perceived as non-simultaneous. For visual signals this gap extends to over ten milliseconds (Pöppel 1988, p. 16; Wittmann 2011, p. 2). Within the "window of simultaneity" (Pöppel 1988, p. 12) stimuli always appear to be concurrent, even if there is an objectively measurable time difference between them. The temporal limit of this window of simultaneity for sound signals is around four-thousands of a second (Pöppel 1988, pp. 18-19). Two clicks, each one millisecond long, played one or two milliseconds apart, would sound simultaneous. In fact, they would seem like one single click—a phenomenon called *click fusion* (Pöppel 1988, p. 11).

As perplexing as it may sound, detecting the non-simultaneity of two events does not necessarily imply that these are perceived in succession. Within particular time frames, it is possible to recognize the non-simultaneity of two events without being able to determine which happened first and which second. If you hear two consecutive clicks, they would need to be around 20 to 40 milliseconds apart for you to be able to tell in which order they were played (Pöppel 1988, p. 19; Pastore and Farrington 1997; Wittmann 2011, p. 2). You would hear that they are non-simultaneous if they were, for instance, ten milliseconds apart. The window of simultaneity varies from sense to sense, but this sequencing threshold of 20 to 40 milliseconds is the same for at least touch, sound, and vision (Pöppel 1988, p. 19). Under the sequencing threshold and over the window of simultaneity, Pöppel speaks of "incomplete subjective simultaneity." Within the window of simultaneity below four milliseconds, we experience "complete subjective simultaneity" (Pöppel 1988, p. 19).

Following Pöppel, the pulse of our temporal consciousness is around 30 milliseconds (on average). In other words, our perception of time is divided into discrete building blocks of a few tens of milliseconds. Below this threshold, we are unable to perceive sequence or duration. To assess the length of any independent stimulus, we would need to clearly determine the point A in time when

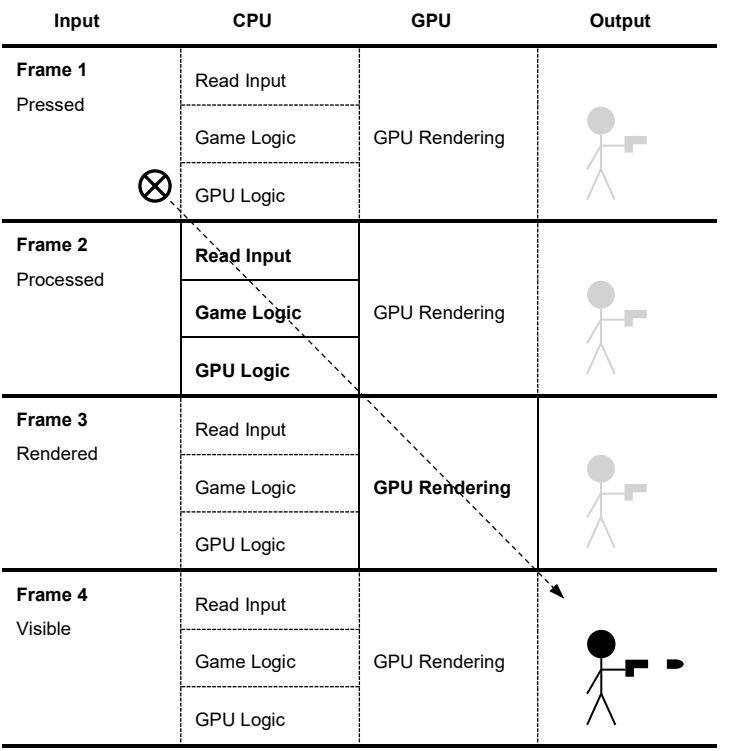
it started and the point B when it ended—a sequence of events. That is, the stimulus would need to be longer than the window of simultaneity (Wittmann 2011, p. 3). All of this also implies that the seamless flow of our experience of time is a fabrication after the fact.

To ensure that a game's controls feel responsive, it is crucial to provide feedback to player input as close as possible to the window of simultaneity. Jesse Schell maintains that "[g]enerally, it is a good rule of thumb that if your interface does not respond to player input within a tenth of a second, the player is going to feel like something is wrong" (2008, p. 231). Several factors influence the responsiveness of a game, some of which are not under the control of the developer (such as the refresh rate of the screen the player is using). One crucial factor is the frame rate at which a game runs. The reason why gamers tend to obsess with this variable is not so much the smoothness of movement on the screen, but the responsiveness of the game. If a game runs at 30 frames per second, movement will be displayed in increments of 33.3 milliseconds. This is slightly above the average threshold to perceive simultaneity. Even if a game displayed the response on the very next frame after a button press, this frame rate is still dangerously close to the limit where we stop perceiving it as simultaneous. At 60 frames per second, that value descends to 16.6 milliseconds, enabling a response to player input below the threshold where we start to detect sequence.

Mick West, the co-founder of Neversoft Entertainment (TONY HAWK'S PRO SKATER (1999), GUITAR HERO III: LEGENDS OF ROCK (2007)), succinctly explains the challenge of programming a responsive game in a Gamasutra feature (West 2008a). As seen in table 1.2, West demonstrates that at least three frames should elapse from the moment the player presses a button until visual feedback is displayed on-screen. When playing a third-person shooter, for example, if the player presses the shoot button, one frame is missed on the press. In frame two the input is read, the logic state is updated, and the CPU (Central Processing Unit) performs its part of the rendering. In frame three, the GPU (Graphics Processing Unit) renders the state. In frame 4, the new game state is finally displayed, that is, the character shoots.

The game loop is what makes the clock in virtual worlds tick. Each loop will produce a frame, and the faster the loop runs, the more frames per second the game will have. A game running at 30 frames per second will then take $3/30^{ths}$ (that is, one-tenth) of a second to display feedback. This equals 100 milliseconds, which is considerably over the sequencing threshold of 30 milliseconds. Running at 60 frames per second, the time the game will take to display feedback is of $3/60^{th}$ (or one-twentieth) of a second, which equals to 50 milliseconds. While this is a substantial improvement, it is still slightly above the threshold.

Table 1.2: The Game Loop.



Source: West 2008a.

After the player's input, it takes a game at least three frames to create visual feedback.

However, games are perfectly playable at 60 or 30 frames per second.[6] This is because the window of integration for multimodal stimuli (that is, pertaining to more than one sense) is larger than for unimodal stimuli. These integration periods can vary considerably, depending on the length of the stimuli, their frequen-

6 The gaming news website IGN offers a table that compares the resolutions and frame rates of several video games on Xbox One and PlayStation 4 (IGN Xbox One Wiki Guide 2019). Though the data are not necessarily reliable in every case, the table shows that the standard frame rates for both mainstream consoles are 30 and 60 frames per second.

cy, and the senses involved (Wittmann 2011, p. 3). Pressing a button and receiving mediated feedback requires the integration of tactile, visual, and auditory stimuli. The duration of these integration tasks is in the order of hundreds of milliseconds, varying from 100 to 250 depending on the complexity of the stimulus. At best, this should be tested in every individual case. Still, it remains true that, the closer the feedback is to the window of simultaneity, the more responsive controls will feel.

In a follow-up piece, West (2008b) describes how he measured the responsiveness of GRAND THEFT AUTO IV (Rockstar North 2008). The technique he uses is clever and simple: he uses a video camera to record the screen and his hand pressing the button at the same time and then measures the response of the game by counting the frames in the video. That is, if in the video West presses the trigger to fire the handgun at frame one, the frames it takes for the game to display the feedback on the screen will show the delay. In West's measurement, GTA IV exhibits a delay of ten frames or 166 milliseconds.[7] This responsiveness dangerously approaches the upper limits of what our perception can synchronize. We do not perceive this time difference directly, but sense that something is odd instead. West describes this feeling as the game being "laggy" or "sluggish."

The speed at which sound is processed speaks to the importance of auditory feedback in video games. Visual feedback is crucial but, when it comes to quick reactions (as in games of skill), it is imperative to pay attention to acoustic signals. For this very reason, the start of a race is signaled with a gunshot instead of a flash. Even though light reaches the runners sooner than sound, a bang is processed faster than a flash, allowing for a swift, almost involuntary reaction. There is however a limit to how quickly we can react to stimuli. The fastest we can respond to an auditory stimulus is 120 milliseconds (the average is 150 to 200 milliseconds) (Anson 1982). From there, reaction times to auditory stimuli can only go up. For visual stimuli, the reaction limit is of approximately 150 milliseconds (with the average at 250 to 300 milliseconds) (Najenson et al. 1989; Jaskowsky et al. 1990). This difference of 0.05 seconds in reaction speed does not seem like much, but it can have an impact in video games where quick reaction times are key for success, such as fighting games like STREET FIGHTER V (Capcom 2016) or competitive first-person shooters like OVERWATCH (Blizzard Entertainment 2016) (compare Pöppel 1988, pp. 23-32).

7 The refresh rate of the television used can produce a greater delay. West used an LCD TV, which actually added lag, bringing the responsiveness of the game up to 200 ms.

The Experienced Moment

At the next level of our temporal consciousness, the window expands from the 0.03 to 0.3 seconds of the *functional moment* to around three full seconds. This three-second span is what Wittmann calls the *experienced moment*, in which our feeling of *nowness* unfolds. This fact has led Pöppel to express his famed law that "we take life three seconds at a time" (Pöppel 2004).

This three-second time window is the longest possible duration of the now, "the temporal limit of consciousness" (Pöppel 1988, p. 49). As author Claudia Hammond puts it: "It is as though every few seconds the brain asks what's new" (Hammond 2012, p. 76). If the context so requires it, the "now" can compress into shorter integration periods. We can even do this actively when focusing on a particular stimulus. Pöppel illustrates this with the example of a metronome: By setting the metronome to 120, one can hear two beats per second. The interval is constant, and each beat sounds exactly the same, but by directing our attention, we can actively bundle the beats into differently sized units—as if we heard "one, two; one, two" or "one, two, three; one, two, three." Here lies one of the main differences between the functional and the experienced moment: in the former, we are passive recipients of stimuli and, in the latter, we acquire some agency as to how the stimuli are perceived (Pöppel 1988, 66; Wittmann 2015, pp. 58-59).

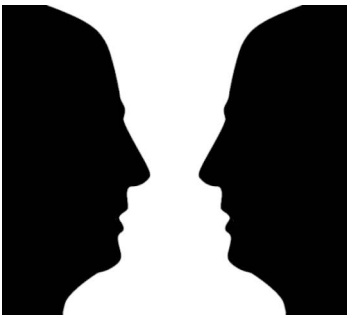
As long as the interval between beats remains around one second, they can be effortlessly integrated into patterns. If the intervals are stretched, there comes a point where it is impossible to keep binding the beats together. At intervals of 2.5 to 3 seconds, the subjective integration of beats into units becomes impossible for most people (Pöppel 1988, pp. 53-54). Beyond this threshold, the interval becomes too long, and it replaces the stimulus (the beat) as the focus of attention (Wittmann 2011, p. 5).

The effects of the three-second present have also been observed in speech and poetry. Spontaneous speech is divided into individual utterances. After each unit of utterance, speakers pause to plan the next one. These units do not surpass the circa three-second limit. The effect is not always observable when reading aloud, because, in that case, the speaker does not need to pause to plan the next utterance. In spontaneous speech, however, the three-second boundary was observed in languages like English, German, Chinese, and Japanese. Children of all ages also speak in units of three seconds. As Pöppel remarks, this is especially notable in children under ten years of age, who speak more slowly than adults but still in three-second units (Pöppel 1988, pp. 71-73). Studies of poetry in several languages (including English, French, Japanese, Latin, and ancient Greek)

conducted by Pöppel and poet Frederick Turner have also shown that the majority of poems are structured in three-second verses. Pöppel also notes that the same holds for poetry with longer verses, such as hexameter and pentameter. In these cases, lines are subdivided into three-second utterances when recited. Deviations tend to manifest below three seconds of duration, with the exception of some postmodern poetry, which does go above the limit of the experienced moment (Pöppel 1988, pp. 75-81).[8]

In vision, bi-stable stimuli like the Rubin vase (figure 1.3) or the Necker cube (discussed in section 2.1) provide further evidence of the experienced moment (see Wittmann 2012, pp. 58-59). The Rubin vase (named after psychologist Edgar Rubin) is an ambiguous figure with two possible interpretations: it can be seen either as a white vase on a black background or as two black profiles facing each other on a white background. These two interpretations switch every two to three seconds unless we direct our attention to one of them earlier—in which case the switch can happen faster. However, even when we purposely focus our attention on say, the vase, it is not possible to maintain this interpretation for long, as the brain will change it by itself when the three-second window is over.

Figure 1.3: The Rubin vase.



8 It should be noted that postmodern art is characterized (among other things) by the abandonment of the value of beauty. While beauty standards can and often are culturally defined, many aspects of beauty are related to our cognitive dispositions. Considering the evidence supporting Pöppel's claim that poetry structured in three-second intervals tends to be preferred, it should in fact be expected that postmodern poetry would reject this structure. Thus, postmodern poetry, with verses that surpass the duration of the experienced moment, is still in line with Pöppel's assertions.

Studies have also found evidence for the duration of the experienced moment in tasks involving sensory-motor control. In one of them, participants were presented a sequence of tones that played in intervals from 300 milliseconds up to 4800 milliseconds. Participants then needed to synchronize taps to the rhythm of these sounds. At intervals longer than 2400 milliseconds participants could not follow the timing and synchronization started to break down. The margin of error increased in proportion to the distance between intervals (Mates et al. 1994). A further study analyzed behavioral data collected from different cultures (Europeans, Trobriand Islanders, the Yanomami people, and Kalahari Bushmen). It showed that human short-term, goal-directed behavior is universally segmented in three-second units of movement (Schleidt et al. 1987).

How the experienced moment manifests in video games remains to be examined. It could be hypothesized that player character animations (attacks, jumps, weapon reloads, movement cycles) tend to stay within the three-second time window. While this seems likely, only systematic observation could confirm if it is true. The performance of combos in fighting games could also display these segmentations. For developers, it is essential to keep the duration of the experienced moment in mind when animating and designing game mechanics. While the responsiveness of a game is detected at the range of milliseconds, activities that require synchronization or the integration of events into bundles should remain within this time window for ease of interaction.

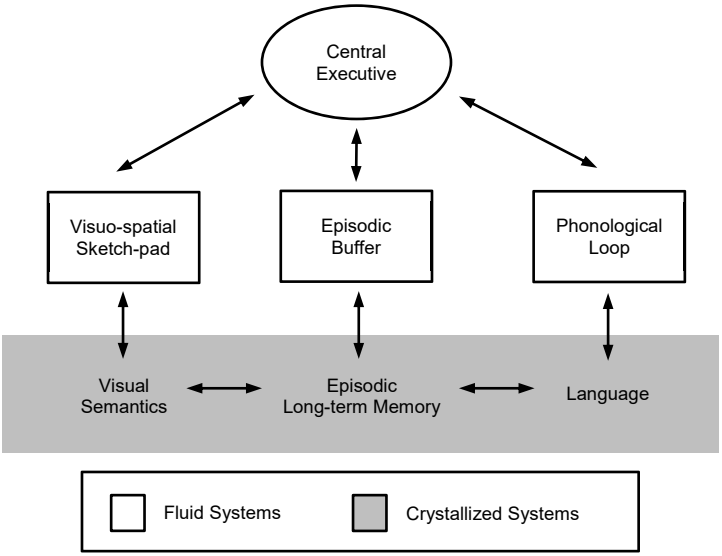
Mental Presence

Beyond the limits of the experienced moment lies what Wittmann calls *mental presence*, the third level of our temporal awareness (Wittmann 2011, p. 5). The mechanism operating at this level is working memory, which allows us to retain visuospatial, phonologic, and episodic representations for periods that can exceed the three seconds of the *now*: "An *experienced moment* happens *now*, for a short but extended moment. *Mental presence* encloses a sequence of such moments for the representation of a unified experience of presence" (Wittmann 2011, p. 5). The temporal boundary of mental presence is much fuzzier than those of functional and experienced moments. Our mental representations progressively lose resolution until they are erased from working memory. Some information might be stored in long-term memory for later recollection, but most of the representations vanish eventually. The information kept in working memory can be obtained from sensory data or retrieved from long-term memory. But working memory is not simply a storage space—the term "short-term

memory" is used to refer to this particular aspect of working memory (Baddeley 2012, p. 4). It is also a system that can manipulate information.

Starting in the 1960s, psychologist Alan Baddeley has conducted groundbreaking research that led to what is now perhaps the most widely accepted model of working memory (see Baddeley 2003, 2012). This model proposes several components of the system and describes their interactions. In figure 1.4, the top four elements labeled as *fluid systems* represent working memory. These are the *central executive*, the *visuo-spatial sketch-pad*, the *episodic buffer*, and the *phonological loop.*

Figure 1.4: Baddeley's working memory model.



Source: Baddeley 2012, p. 16.

The central executive is our spotlight of attention. It can divide attention between two stimuli, switch tasks, and retrieve information from long-term memory. The visuo-spatial sketch-pad specializes in the storage and processing of visual information (color, shape, movement), and the phonological loop on auditory information (pertaining also to language). The episodic buffer works as storage for information from diverse sources, integrating stimuli into coherent perceptual phenomena. Information like touch, sound, or image is processed in different areas of the brain at different speeds. The episodic buffer creates a unitary repre-

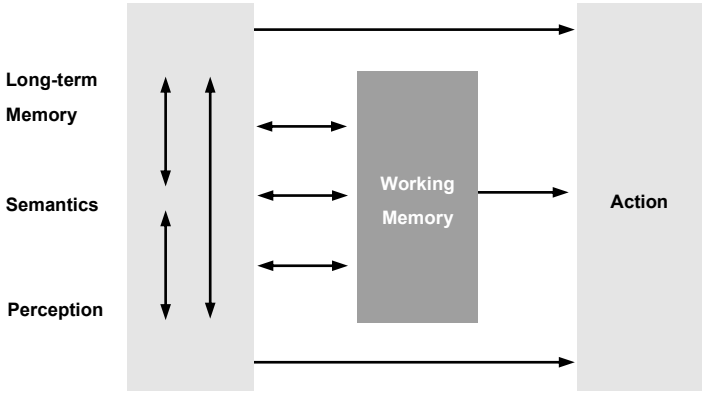
sentation of events in which all the different stimuli are bound together. The episodic buffer is also capable of binding information creatively, which allows us to fantasize about winged horses and talking snakes. Moreover, it allows us to organize different events in narrative sequences. The combination of the creative and narrative capacities of the episodic buffer enables us to anticipate future events (prediction will be further discussed in chapter 2.1). Baddeley also hypothesizes modules of working memory specialized in smell and touch, but these remain to be studied (see Baddeley 2012).

The grey area in figure 1.4 features elements of long-term memory (the crystallized systems) that interface with the different aspects of working memory. Working memory depends on systems in long-term memory in order to function. The different sets of information stored in these crystallized systems are loaded into working memory and bound together in the episodic buffer. To process language, for example, we save the vocabulary and grammar of the language we speak in long-term memory. This information is then loaded to working memory when conversing or reading. Episodic long-term memory, also referred to as our "narrative self" (Wittmann 2011, p. 5), is the information of past events that provides us with a general story of our lives.

Long-term memory is commonly divided into two main parts according to type of information encoded: *declarative* and *procedural* memory (Cohen and Squire 1980). Declarative knowledge is data-based and can be expressed—names, telephone numbers, email addresses, or a conversation we had with someone. Procedural knowledge is rule-based and related to actions—driving, riding a bicycle, playing the piano, or playing SUPER MARIO BROS.

Our behavior can be influenced by long-term memory either directly or indirectly through working memory (figure 1.5). In the case of runners hearing the starting gun, for example, the bang directly sets them in motion. This reaction can be trained to become instantaneous, almost like a reflex (but not an innate reflex like moving the hand away from a hot stove). In this example, the procedural information in long-term memory on how to perform the start (acquired through training) is associated with the bang to cause action directly. Working memory influences action when making more complex decisions that require deliberation and when an instant reaction is not necessary. Ernst Pöppel talks of *simple reactions* in the first case and *decision reactions* in the second (Pöppel 1988, pp. 23-32). In simple reactions, one stimulus is connected to one response. Decision reactions can vary in complexity and are much more malleable.

Figure 1.5: Baddeley's view of the links between long-term memory and working memory.



Source: Baddeley 2012, p. 18.

A critical aspect of working memory is that it is limited. Imagine playing the fighting game MORTAL KOMBAT X (Netherrealm Studios 2015) for the first time. You select the character Sub-Zero and proceed to play against a friend. In order to use the character to its full potential, you would need to know all of Sub-Zero's basic moves, special moves, combos, and finishing moves. Additionally, each character has three variations, which have different attacks, weapons, or special moves. Sub-Zero has the *Cryomancer*, *Unbreakable*, and *Grandmaster* variations.

Table 1.3 shows the list of button combinations needed to execute Sub-Zero's special moves and *kombo* attacks with the Cryomancer variation. These lists can be accessed from the pause menu during a fight. It would be absurd to try to memorize all the available combinations in your first fight using the character. The more sensible strategy would be to repeat one or two combinations in your head that you could then test during combat against a friend. The only way to learn all of Sub-Zero's moves and combos is through sustained practice.

By repeating commands and probing and exploring the gamespace, players store information in long-term memory, which enables them to produce faster reactions—both simple and decision reactions—than new players. The mastery of a game is thus dependent on the information stored in long-term memory, which enables players to react more rapidly in goal-directed ways.

Table 1.3: List of moves that Sub-Zero can perform in MORTAL KOMBAT X.

Kombo Attacks		**Special Moves**	
Frosty	□, □	*Ice Burst*	⇩⇨ □
Hailstone	□, □, △	*Frost Bomb*	⇩⇨ □ + R2
Permafrost	□, □, ○	*Frost Hammer*	⇩⇨ △
Black Ice	□, △	*Crushing Hammer*	⇩⇨ △ + R2
Straight Slash	⇦ □, △	*Air Frost Hammer*	(Air) ⇩⇨ △
Throat Slice	⇦ □, △, △	*Ice Ball*	⇩⇦ △
Tundra	□, △, ×	*Ice Blast*	⇩⇦ △ + R2
Snow Fall	⇨ □, △	*Slide*	⇨⇦ ○
Quick Slice	□, □, □	*Icy Slide*	⇨⇦ ○ + R2
Cold Punish	△, ○		
Ice Pain	△, ○, △		
Ices Up	⇨ ×, ×		

This table lists Sub-Zero's Cryomancer mode kombos (left) and Special Moves (right). The button combinations are based on the PlayStation 4 version of the game.

The limits of working memory depend on many conditions, which cannot be described within the scope of this work. A useful heuristic is to think of this limit as seven, plus or minus two (Miller 1956). That is the approximate number of pieces of information that a person can maintain in short-term memory. If we try to remember a series of letters, we would experience difficulties with more than seven. To store more pieces of information in working memory, we can rely on a process called *chunking* (ibid.), which allows us to cluster information into groups. Through chunking we can, for instance, reproduce lists of up to seven words, which include many letters each, allowing us to remember more letters than if we were just repeating them individually.

However, the retention of information can be quickly interrupted when other stimuli capture our attention. Video game designers know this quite well, since video games commonly aid players with menus (such as the ones in MORTAL KOMBAT X) and other elements (for example, button prompts) that function as memory aids. Role-playing games, for example, typically have menus where players can review the quests that are open or still pending. In this way, players

do not need to rely on memory to recall the objectives or any other information relevant to ongoing or pending quests.

The phenomenon of *immersion* discussed in game studies literature is explained as a consequence of the structure and limited capacity of working memory. Being immersed is, according to Janet Murray, "the sensation of being surrounded by a completely other reality […] that takes over all of our attention, our whole perceptual apparatus" (Murray 1997, p. 98).[9] When playing a video game, working memory is mostly filled with information about the state of the gameworld, and only little information about our body or the environment. Furthermore, the spotlight of attention (the central executive) is directed to the stimuli relevant to the task at hand (playing the game), moving awareness away from irrelevant stimuli. Therefore, we lose awareness of our surroundings and our body, and the feeling of being immersed in the gameworld ensues (section 3.1 will look further into this phenomenon in relation to the state of flow).

The details of our perception of the present are highly complex, and the above lines cannot do them justice. With this overview, I mean to show how games are shaped by the temporal architecture of our minds in relation to the present moment. Understanding these subjacent structures can help scholars better understand video games and developers gain more control over the experiences they wish to create. There is still much more to time perception, as the rest of this study will show. But first, it is important to pay closer attention to the medium. The following section dissects the video game into the basic components that make up its temporality. These elements are arranged into a typology of temporal structures.

9 For an overview of the discussion on immersion in the game studies field (in German language) see Neitzel 2012 (pp. 76-80).