Predictive Thinking in Virtual Worlds[1]

Launching a video game for the first time entails stepping into an unfamiliar virtual world. Whether it is a medieval fantasy setting like the Northern Realms of THE WITCHER 3: WILD HUNT, a sprawling modern metropolis such as Los Santos in GRAND THEFT AUTO V, or FALLOUT 4's (Bethesda Softworks 2015) post-apocalyptic rendering of New England, video games exhibit a diverse assortment of scenarios through which players have to learn to navigate by operating their virtual personas.

Not only do players need to learn their way around these settings, they also need to assimilate the mechanics that enable movement and interaction: The actions that the player character can perform and the responsiveness of the controls usually vary from game to game. Besides, the physical laws that govern each virtual environment might differ from the ones we are used to from the real world or other gameworlds.

The adjustment to these properties of the gameworld takes place through a trial-and-error process that Torben Grodal has dubbed the "aesthetic of repetition" (2003, 148):

> "When we arrive to a new city or a new building we slowly learn how to move around, and if we want to learn to drive or bike, we exercise those skills until we have acquired the necessary procedural skills. The video game experience is very much similar to such an everyday experience of learning and controlling by repetitive rehearsal" (ibid.).

1 An earlier version of this section was previously published in the anthology BILDVERSTEHEN. SPIELARTEN UND AUSPRÄGUNGEN DER VERARBEITUNG MULTIMODALER BILDMEDIEN edited by Lars C. Grabbe, Patrick Rupert-Kruse, and Norbert M. Schmitz (Alvarez Igarzábal 2017a).

Thus, players engage in a heuristic process through which they assimilate the design and mechanics of the virtual world until the control of the avatar becomes second nature. Furthermore, someone who is unfamiliar with the particular input device at hand (controller, joystick, mouse and keyboard) would need to learn how to operate it as well. Inexperienced players will typically take their eyes off the screen and look at the controller to locate the button they want to press, or they will lean to one side when they want the character to move in that direction but it is not responding as they expect. Even seasoned players used to a particular input device—mouse and keyboard, for example—might have trouble when switching to a new one—like the Xbox One controller.

The aesthetic of repetition presents itself therefore on two layers: (1) At the level of the physical interface and (2) at the level of the game mechanics. Naturally, the more confident a player is with the first layer, the faster they will be at assimilating the workings of the second.

To some extent, everyone is familiar with the aesthetic of repetition. Steve Baumgart was the winner of what the Rolling Stone magazine claims was the first video game tournament, held in 1972 at the Stanford Artificial Intelligence Lab. The game at the center of the competition was SPACEWAR! In an interview, Baumgart said: "Pretty soon, you don't think about the buttons (…) It's like speed typing – you just look at ships on the screen and make them move where you need them to go" (Baker 2016). But what takes players from needing to pay close attention to the actions they are performing to a state in which they can act without having to "think about the buttons"? A compelling answer to these questions comes from a theory that philosopher Andy Clark (2013) has dubbed *action-oriented predictive processing*, which asserts that the brain is a machine that applies Bayesian statistics to anticipate the state of its surroundings. The theory is principally based on research conducted by neuroscientist Karl Friston (2003, 2005, 2010, 2011, 2012; Friston and Kiebel 2009).

According to this paradigm, the brain creates models of the environment that it matches to incoming sensory information. Should there be an incongruity, the model in the brain is updated accordingly (Clark 2013, p. 182). If the model matches the upstream sensory signal, no update is necessary, so it remains unchanged. Thus, the more experience with a particular activity someone has, the more updated their model of said activity will be, allowing them at some point to operate on autopilot.

Evidence from numerous studies shows that this unifying framework can account for both perception and action. This theory goes beyond the layer of our direct experience into subconscious processes that lie beneath it. After all, a process that is second nature should be expected to be at least partly subconscious in

order to be performed without actively paying attention to it. In the words of Clark (2013, p. 197):

"The world, it might be said, does not look as if it is encoded as an intertwined set of probability density distributions! It looks unitary and, on a clear day, unambiguous. But this phenomenology again poses no real challenge. What is on offer, after all, is a story about the brain's way of encoding information about the world. It is not directly a story about how things seem to agents deploying that means of encoding information."

THE BAYESIAN BRAIN

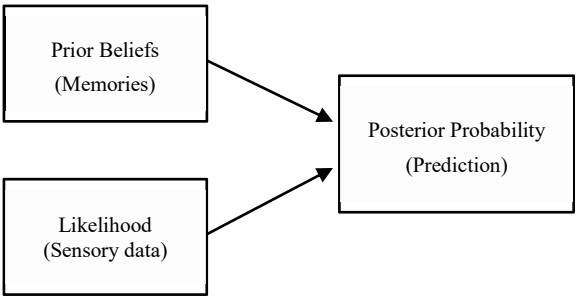
The Bayesian method is one of the two major theories of statistics—the other being classical or frequentist statistics (Romeijn 2016). The basic components of Bayesian statistics are: (1) the priors or prior beliefs—hypotheses based on previous experience; (2) the likelihood—gathered data in the present moment; and (3) the posterior probability—the most likely scenario determined by the information in the two first sets. That is, the priors and the likelihood are fed to a Bayesian estimator, which calculates how likely a particular event is to happen (figure 2.1).[2]

The brain is, within this framework, a Bayesian estimator that possesses models of the world obtained through previous experience or hardwired through evolution (the priors), collects information through the senses (the likelihood), and infers the most likely state of the environment from those two sets of data (the posterior probability). This process results in our experience of the world (compare Clark 2013, Friston 2011, Körding and Wolpert 2006). As Andy Clark remarks:

"[T]he task of the brain, when viewed from a certain distance, can seem impossible: it must discover information about the likely causes of impinging signals without any form of direct access to their source. Thus, consider a black box taking inputs from a complex external world. The box has input and output channels along which signals flow. But all that it "knows", in any direct sense, are the ways its own states (e.g., spike trains) flow and alter. In that (restricted) sense, all the system has direct access to is its own states […] The brain is one such black box" (Clark 2013, p. 183).

2 Central to this theory is Bayes' theorem, the rule with which the posterior probability can be estimated. Understanding the theorem is not necessary to grasp the logic behind Bayesian inference, so I have chosen to omit it for the sake of clarity.

Figure 2.1: The likelihood and prior beliefs are fed to a Bayesian estimator, which calculates the posterior probability.



From the perspective of the brain, even the body is a part of the external world (Friston 2011, p. 92) and, to complicate things further, the information obtained by the senses (the likelihood) is contaminated by noise (Körding and Wolpert 2006, p. 319). This means that the brain needs to estimate the state of the world and generate reactions to it in a constant state of uncertainty. Through movement, the brain probes the world and updates its priors in the light of the incoming stream of sensory information, generating a feedback loop that integrates perception and motor action into one model.

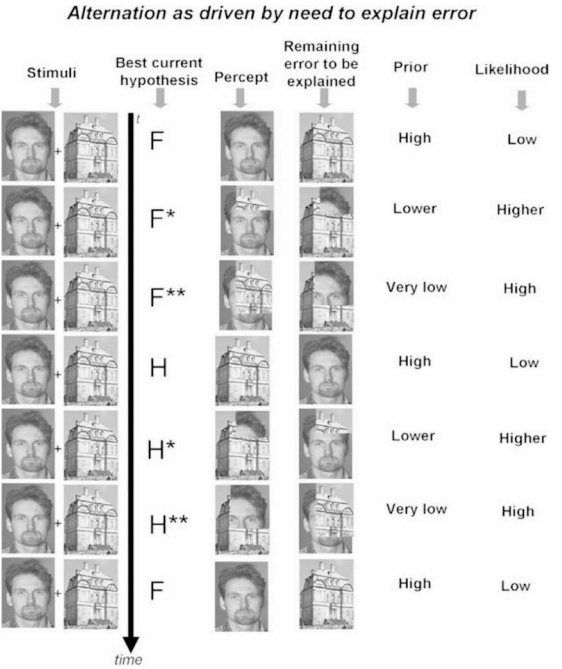
BAYESIAN INFERENCE IN VISUAL PERCEPTION

Cases of Bayesian inference in visual perception can help clarify the theory before moving to examples of motor control. Hohwy, Roepstorff, and Friston (2008) have argued that this Bayesian model is cohesive with diverse studies in binocular rivalry (see Alais and Blake 2005; Leopold and Logothetis 1999; Tong, et al. 2006). This phenomenon occurs when a person is presented with a bi-stable stimulus:

"If one stimulus is shown to one eye and another stimulus to the other, then subjective experience alternates between them. For example, when an image of a house is presented to one eye and an image of a face to the other, then subjective experience alternates between the house and the face" (Hohwy et al. 2008, p. 687).

In this case, the subject not only sees either a house or a face, but the perception will shift from one to the other, with combinations of both in between. At first, the subject might only perceive the face. Then, seconds later, the perception will change to part face, part house, until finally only the house will remain visible. This phenomenon will then repeat back and forth indefinitely in intervals of around three seconds—that is, the duration of the experienced moment discussed in chapter one, section 1.1. Figure 2.2 shows a representation of the effect that said bi-stable stimuli have in perception.

Figure 2.2: Simplified Bayesian scheme for the alternation of stimuli in binocular rivalry.

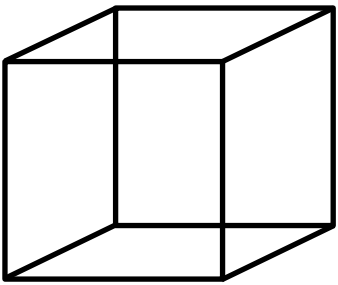


Source: Hohwy et al. 2008, p. 693.

The reason behind the phenomenon of binocular rivalry is that two different objects (the face and the house) appear to share the same spatiotemporal location and, thus, "[n]o single hypothesis accounts for all the data, so the system alternates between the two semi-stable states" (Clark 2013, p. 185). The incapability of two objects to be at the same time in the same place is a "systemic prior" (ibid.) or hyperprior: "…binocular vision, in primates, rests upon both eyes fo-

veating the same part of visual space. We have therefore learned that the explanation for binocular visual input is unitary" (Hohwy et al. 2008, p. 691). This "failure" of perception caused by an artificial state of affairs in the experimental environment gives us a glimpse behind the curtain that is consistent with the picture of the brain painted by the Bayesian framework. In normal circumstances, objects in the visual field do not share the same place at the same time. So, in the end, the brain settles for the strongest hypothesis, which will be the subject's experience of the world: "What ultimately determines the resulting conscious perception is the best hypothesis: the one that makes the best predictions and that, taking priors into consideration, is consequently assigned the highest posterior probability" (ibid., p. 690).

Figure 2.3: The Necker cube.

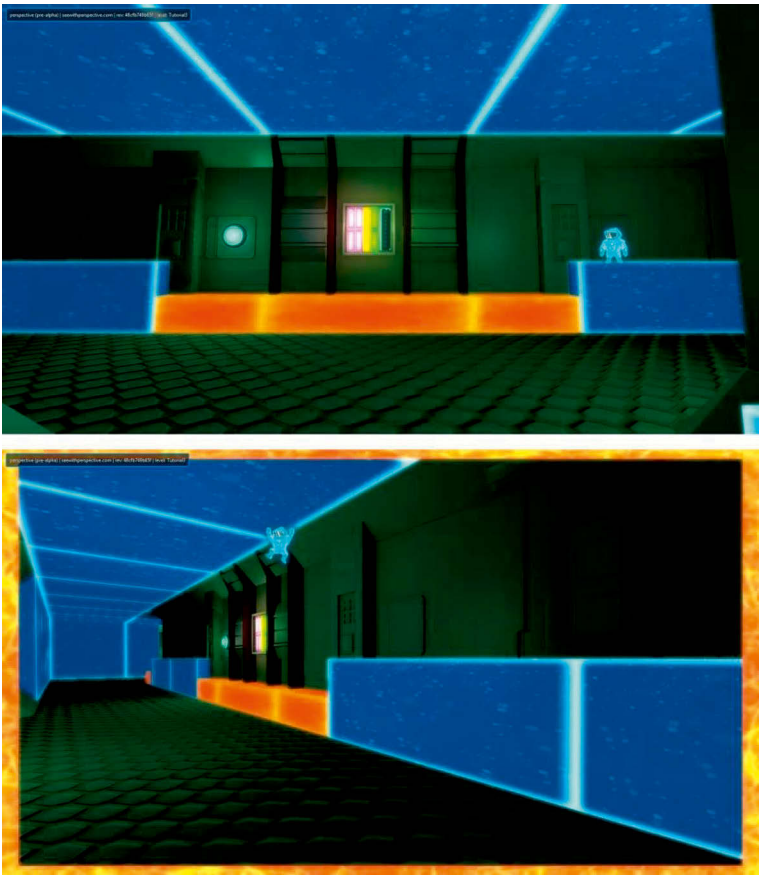


Hohwy et al. also assert that this explanation applies to bistable stimuli like the Necker cube (ibid., p. 699). In section 1.1, I discussed this type of ambiguous imagery with the example of the Rubin Vase. The Necker cube (figure 2.3) is a two-dimensional figure made up of straight lines arranged in such a way that the brain interprets them as a cube. This cube, however, can be seen from two different perspectives: either from the top, with the front face of the figure leaning to the left, or from the bottom, with the front face to the right. Since actual three-dimensional objects cannot be seen simultaneously from two perspectives, the brain tests both hypotheses by alternating between them (circa every three seconds). Once again, there is no solution to this conundrum, so the brain can only carry on shifting perspectives.

While visual stimuli in video games tend to be congruent—and nothing like the extreme example of the face and the house—, some make use of ambiguous imagery evocative of Escher's famed works "Belvedere" or "Waterfall," or impossible figures like the Penrose triangle (Penrose and Penrose 1958) in order to

obfuscate the player's interpretation of the gameworld and, thus, complicate navigation (see Hensel 2015). ECHOCHROME (Sony Computer Entertainment Japan Studio 2008) and PERSPECTIVE (DigiPen 2012) are two examples of this.

Figure 2.4: PERSPECTIVE.



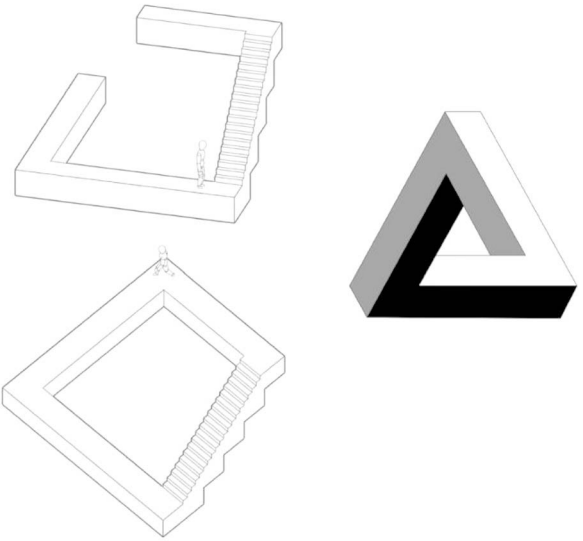
Source: https://www.youtube.com/watch?v=XSS6QBMtfqI (1:06, 1:17).

Top: the blue platforms are too far apart for the player character to jump over the deadly orange platform. Bottom: Moving the camera, and thus changing perspective, brings the blue platforms closer together in the two-dimensional plane, enabling the character to jump across.

Both games combine the logic of a two-dimensional image with a three-dimensional environment, and the player needs to find the most suitable point of view with the camera for the avatar to be able to move from surface to surface. These aesthetic and mechanic elements toy with our systemic priors and make it demanding to determine the avatar's position in relation to platforms and other objects in the world.

In PERSPECTIVE, the player needs to alternate between two discrete play modes: camera movement and platforming. For example, if two platforms are too far away to jump across the gap between them from a side view (as seen in figure 2.4), the camera can be moved to change the angle of the platforms and place them closer together in the two-dimensional plane. This action enables the player-character to make the jump.

Figure 2.5: ECHOCHROME (left) and the Penrose triangle (right).



Source: Left: https://www.youtube.com/watch?v=Pm-4gfJshA8 (accessed June 6, 2019). Right: https://commons.wikimedia.org/wiki/File:Penrose-dreieck.svg (accessed February 9, 2018).

Left: A sequence from ECHOCHROME's tutorial showing the mechanics used to connect platforms through changes in perspective. Right: The Penrose triangle for comparison.

ECHOCHROME (figure 2.5) implements an aesthetic more reminiscent of Penrose's impossible imagery and, in a very similar way to PERSPECTIVE, the player needs to move the camera around to find a suitable arrangement of the platforms that will allow the character to traverse the gamespace. The character, however, is not player controlled, but walks automatically. The challenge for the player is to swiftly move the camera and adjust the perspective according to the needs of the moment so that the character can reach the end of the stage.

The video game examples above show the interrelation between sensory information (visual in this case) and motor action: Placing the camera at a particular angle in PERSPECTIVE might lead a player to believe that the character will be able to jump from one particular platform to another. While performing the action, however, they might realize that the gap between both surfaces is too wide as they see the player character fall through it. This event will bring about an update of the player's prior beliefs, who will subsequently adjust the camera angle. With each challenge, the player will become better at estimating the appropriate distance between two platforms and whether the character is capable of making the jump or not.

BAYESIAN INFERENCE IN MOVEMENT

Imagine an everyday scenario in which you go to the kitchen to get a glass of water. The pitcher is opaque, so you cannot see exactly how much water is in it (the information is incomplete). Since you filled it earlier, you assume it is still full and apply the necessary force to lift a pitcher containing approximately two liters of water. However, your roommate has drunk most of the water without you noticing and did not refill the container. The pitcher will thus offer less resistance than expected, rising surprisingly fast. However, in an instant, you can readjust the applied force to avoid hurling the pitcher into the air.

The curious aspect of this scenario is that you do not need to consciously think about the contents of the pitcher to assume that it is full. The estimation can be, and often is, performed tacitly. If there had been no discordance between your belief and the feedback, you probably would not have noticed the assumption you were making about the weight of the pitcher. But, when the expectations about the environment do not match its actual state, your belief is updated as soon as new information is received and you become aware of your presuppositions.

To put it in slightly more Bayesian terms: You approach the pitcher with a hypothesis about its state that guides your motor actions. When your hand grasps

the container and your arm applies force to lift it, a feedback signal moves up the sensory stream. Since this information does not match your model of the world, the feedback is understood as an error signal. This mismatch between hypothesis and incoming sensory information is called *surprisal*—different from *surprise*, which relates to the conscious experience of an unexpected event (Clark 2013, p. 3)—and it causes the model of the environment to be corrected. That is, bottom-up information obtained by the senses is compared to the top-down model of the world and, given that there is a disparity between prediction and sensory information, the model is updated.

These predictions—such as the one your brain made about the weight of the pitcher—are essential to interact with the world. In the words of Daniel Dennett: "[t]he brain's task is to guide the body it controls through a world of shifting conditions and sudden surprises, so it must gather information from that world and use it swiftly to 'produce future'—to extract anticipations in order to stay one step ahead of disaster" (Dennett 1991, p. 144).

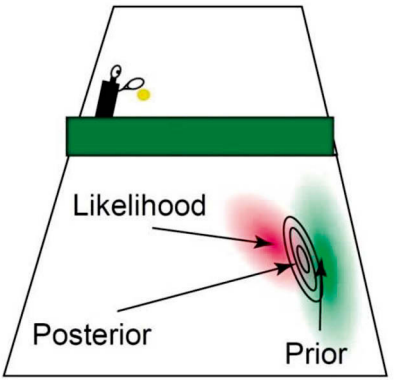
The following example (figure 2.6) by Körding and Wolpert (2006, p. 319) illustrates this notion quite eloquently:

> "[W]hen playing tennis we may want to estimate where the ball will bounce. Because vision does not provide perfect information about the ball's velocity there is uncertainty as to the bounce location. However, if we know about the noise in our sensory system then the sensory input can be used to compute the likelihood […] We can combine this with information that is available over repeated experience of tennis: the position where the ball hits the ground is not uniformly distributed over the court. For example the bounce locations are likely to be concentrated within the confines of the court and the distribution might be highly peaked near the boundary lines where it is most difficult to return the ball."

Figure 2.6 shows three probability distributions: the red gradient indicates the likelihood, the green the prior distribution, and the black ellipses mark where the ball is more likely to bounce, or the posterior probability, as computed by a Bayesian estimator (in this case, the player's brain). I have previously discussed Ernst Pöppel's distinction between simple and decision reactions (section 1.1). In simple reactions, there is one automatic response to one stimulus—I hear a bang, so I start running. These responses can be trained through practice to be faster. Decision reactions are slower but can vary in complexity. In the tennis example described above, the player is met with a decision reaction—determine the speed and direction of the ball, run to a position in the court where the ball can be intercepted, and swing the racket in time to hit the ball in the preferred di-

rection and with the desired strength. The more prior information the player possesses about the game, the faster and more accurate this decision reaction will be.

Figure 2.6: Illustration of tennis example by Körding and Wolpert.



Source: Körding and Wolpert 2006, p. 320.

It is easy to see how this example could translate to a video game like PONG, which is a simplified, virtual version of table tennis. The motor actions that the players would have to execute are different in each case: The tennis player would run towards the alleged landing location of the ball and swing their arm holding the racket accordingly, while the PONG player would move the virtual paddle by means of whatever interface they are using at the moment—such as pressing a key on the keyboard or rotating a knob, as in the case of the original PONG machine. However, both players would estimate the trajectory of the ball with the same sets of data: the current visual information of the ball and their previous experience with the game. But any video game that involves the development of skills rests on the principle of learning through repetition, which relies on the mechanism of action-oriented predictive processing.

COPING WITH UNCERTAINTY

Thomas Malaby defines games as "a semibounded and socially legitimate domain of contrived contingency that generates interpretable outcomes" (Malaby 2007, p. 96). Applied to the specific realm of video games, the domain of contrived contingency is typically a gamespace with entities that behave in different

ways and influence each other. The role of the player is to set different variables into motion in pursuit of a particular outcome, usually dictated by the game's objectives. The result of the player's actions is indeterminate, and the challenge of video game design is to strike a satisfactory balance between control and uncertainty within this contingency.

Roger Caillois noted that "[a]n outcome known in advance, with no possibility of error or surprise, clearly leading to an inescapable result, is incompatible with the nature of play" (Caillois 2001, p. 7). As stated in the previous pages, uncertainty is an inescapable fact of life that is not exclusive to play or games. We deal with incomplete and inaccurate information on a daily basis. But, while other systems are designed to reduce uncertainty, games emphasize it. Play theorist Brian Sutton-Smith argues that

> "[a]ll creatures, animal and human, live with some degree of existential angst, and most of them spend some portion of their time attempting to secure themselves from this angst by controlling their circumstances […] We constantly seek to manage the variable contingencies of our lives for success over failure, for life over death. Play itself may be a model of just this everyday existentialism" (Sutton-Smith 1997, p. 228).

Malaby's concept of "contrived contingency" proves fitting in this context. It recognizes games as artifacts in which a scenario with fluctuating variables is orchestrated by a designer for a player to interact with according to certain rules while pursuing particular goals. They emulate the uncertainty of everyday life in a more constrained system and give players the promise of control over this artificial environment.

At first, it is to be expected that the interaction with the gameworld is informed by bottom-up sensory information to a greater degree, since many of the brain's predictions will fail to anticipate the state of the novel virtual scenario. Through interaction with the virtual world, players can update their priors—that is, improve their models of said world—and become better at predicting its states and future events. With time, actions will rely increasingly on top-down models of the environment and less on the incoming sensory stream of data. Given that bottom-up sensory information requires more time to be processed, the better players become at predicting the states of the world, the swifter and more precise their reactions will be. Such is the central mechanism behind the aesthetic of repetition described by Grodal.

In this context, playing video games can be understood as a process of uncertainty reduction through the accumulation of prior knowledge. The accrual of priors leads to increasingly accurate mental models of the virtual environment

and, thus, to greater control over it. Therefore, in order to master a game's mechanics, players will perform the same actions repeatedly. Mastering the jump in SUPER MARIO BROS., for instance, entails pressing the jump button again and again in order to assess different variables—for instance, how far or high Mario can jump, or to what extent he can change direction mid-air. Additionally, the player can test how these values are affected by the momentum acquired through running. Most of these repetitions are performed in safe conditions: If a player fails to leap over one of the game's warp pipes because they did not jump high enough, the avatar will often just hit the pipe's side and drop to the ground, losing nothing but a couple of seconds in the process. Players may also simply jump around aimlessly without being motivated by the environment, either as an intentional form of practice or just because they can. The majority of interactions in video games are of this nature. They tend to be less salient than actions that could damage or kill the player character, but they are greater in number and are part of the prior updating process.

Often, however, players need to jump over bottomless pits, spikes, or other hazards that might threaten the life of the player character or diminish valuable resources (e.g., health). In this context, there is one further characteristic of video games that must be taken into account: the capacity of games to reset time.