Reinforcing motor re-training and rehabilitation through games: a machine-learning perspective

Maurizio Schmid[1,2]*

[1] *Department of Applied Electronics, Roma Tre University, Rome, Italy*
[2] *Harvard Medical School, Spaulding Rehabilitation Hospital, Cambridge, MA, USA*
* Correspondence: schmid@uniroma3.it

This opinion article wishes to highlight the importance of applying reinforcement learning in games designed for therapeutic or educational purposes and is directed in particular to the medical community. Games provide enjoyment: people learn and keep on playing just for the sake of it, because it is fun. In addition the intrinsic presence of competition and confrontation in any game might be critical in social behaviour. Playing also represents a powerful tool for individual growth and development. In fact games might fulfil different objectives, such as supporting maturation in child, or enhancing educational programs and ultimately facilitating recovery from diverse pathologies. For example, when focussing on motor activities, it is widely accepted that the repetition of sessions of movement therapy plays a key role in the modification of motor outcomes (Kwakkel et al., 2004). In this context, games, due to their ability to improve patients' compliance to treatments, are increasingly being exploited to strengthen and validate training programs directed to improve motor outcomes (Johnson, 2006; Mirelman et al., 2009).
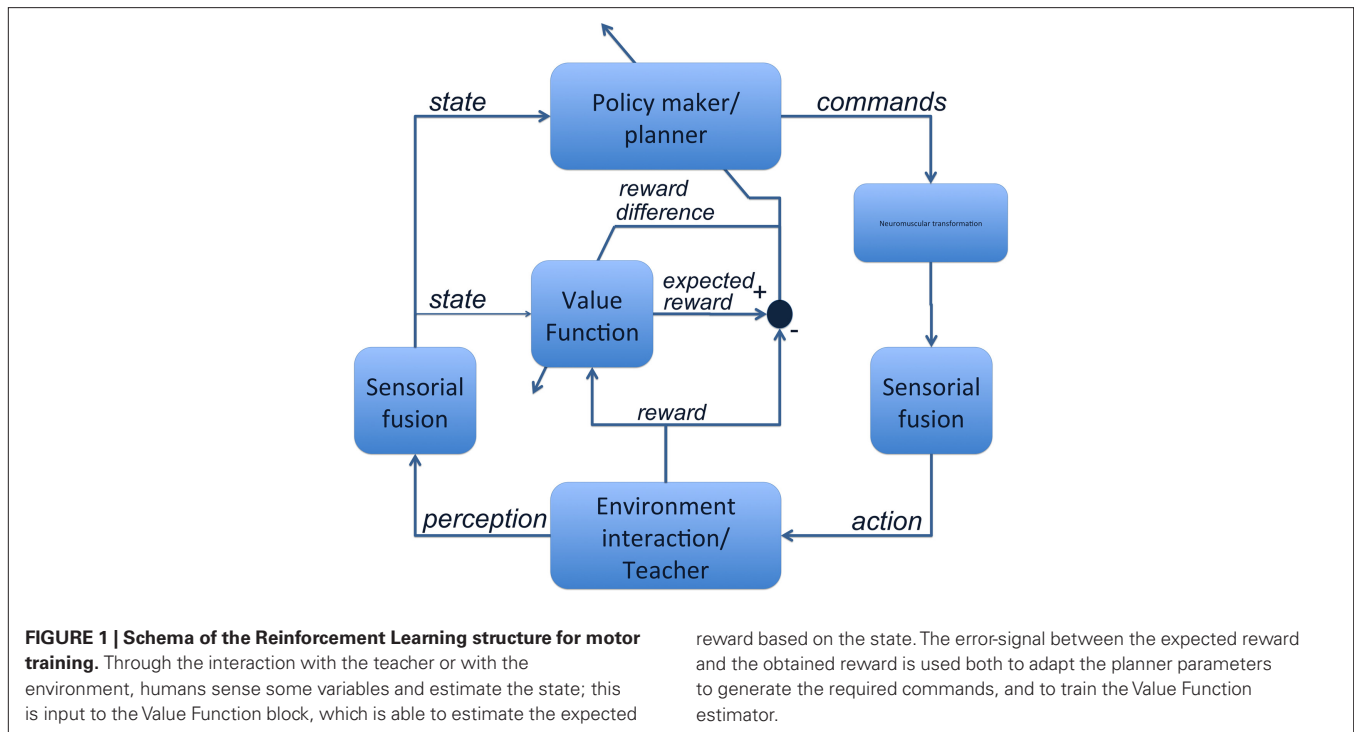
The role of games in improving motor outcomes during rehabilitation is quite obvious, however the rationale behind game clinical efficacy in rehabilitation programs which include games, when compared to the conventional ones (Broeren et al., 2008), has not been developed in a theoretical framework. Certainly, game-based rehabilitation is more effective due to its "more engaging" nature. In this framework, it may be helpful to take into consideration the theory behind machine learning and artificial intelligent systems, for which the outputs of motor re-training represent a valid replica: it is indeed demonstrated that repetitive task-oriented practice in an un-supervised scenario (which mimics a learning-by-doing paradigm) in motor re-training represents a viable and valid method for rehabilitation, and that another means to convey rehabili-

tation is usually obtained by either a real teacher (e.g. the physical therapist), or a virtual teacher (such as in virtual reality-based rehabilitation), thus triggering an instance of the learning-by-imitation scheme. If the rehabilitation exercises are administered in a game fashion, which means asking the player to choose a solution in terms of movement execution out of a number of discrete possible alternatives, we can hypothesize that this is an example of a reinforcement-learning scheme (Sutton and Barto, 1998).

The internal model hypothesis, i.e. the presence of an internal representation of the dynamics and kinematics of movement at the neural level, is acknowledged as one of the leading theories of motor control and learning in neuroscience (Jordan and Rumelhart, 1992), and many authors proposed different mathematical proxies of this empirical paradigm (see e.g. Tin and Poon, 2005). In the framework of rehabilitation, it is assumed that, in the brain (the controller), this internal representation is re-trained by changing its functional structure over time through interaction with modifications either in the environment or in the affordances of the body. It is acknowledged that most of the computational models that can simulate this ability towards change are based on the presence of rules (needed to adjust the parameters of the controller), where the interaction between the controller and the controlled plant (the body) allows some kind of modification. Among them, reinforcement learning is one of those machine-learning approaches where an entity decides a strategy that will maximize some reward function, as a result of the actions (Minsky, 1961). The controller, which is called the *agent* in machine learning definitions, is not actually told which actions need to be taken, but learns to adapt the policy, based on the obtained reward. Algorithms able to analytically provide this kind of training are now at hand (Williams, 1992), and the mechanism itself has been

fully studied in behavioural analyses of games (see e.g. Zaghloul et al., 2009), where simple reinforcement learning schemes have been demonstrated as efficient in respecting both the law of effect, for which choices that carry positive effects in the short term tend to be repeated in the future, and the power law of practice, for which the performance of a natural system when facing unseen conditions tends to increase with a decreasing rate over time (Erev and Roth, 1998).

In the framework of motor control, this approach has been hypothesized as the only one that grants consistent improvement in general learning-by-doing schemes. At the same time, the implementation of reinforcement learning is generally based on the presence of a discrete number of different choices (a number of primitives), among which the agent is able to choose. The availability of (non-strictly) determined games that involve decision making and execution of a discrete and finite number of motor tasks is fostering new scenarios for enhancing re-training based on reinforcement. It is indeed acknowledged that the prefrontal motor cortex serves as the command generator that sends input to the descending pathways. The way this system chooses the pattern of commands based on the desired action can be represented as the function of a planner, whereas a controller at a lower hierarchical level transforms these neural commands into actual movers. If no mechanism intended to modify the parameters of the planner is present, the controller will only be able to adapt to modifications based on the nature of the planner, in a short-term paradigm based on the difference between the intended action and the executed one. If, instead, some measure of long-term reward is used to let the planner modify its behaviour, the latter will adapt its parameters based on how much the reward differs from the reward it was expecting. Once this difference is negligible, the planner will be trained to drive the

FIGURE 1 | Schema of the Reinforcement Learning structure for motor training. Through the interaction with the teacher or with the environment, humans sense some variables and estimate the state; this is input to the Value Function block, which is able to estimate the expected reward based on the state. The error-signal between the expected reward and the obtained reward is used both to adapt the planner parameters to generate the required commands, and to train the Value Function estimator.

lower hierarchy blocks in such a way that the actions are the ones needed to correctly perform the task, and obtain the reward which was expected.

This mechanism is the basis of training in games, where the Value Function inner block in **Figure 1** replicates the mechanism of "acting to maximize the reward". If no games are present, the training is performed based on the difference between the intended action and the obtained one, with no long-term goal. As a result, if this were the case of rehabilitation programs for humans, the absence of reward mechanisms can lead to a "less adaptive" behaviour where no changes in the parameters of the planner appear due to long-term rewards. On the contrary, a system able to excite the reward mechanism (such as the case of games in a training session) will facilitate modifications of the planner, improve retention of the actions that grant a reward, and possibly facilitate generalization ability. It is here speculated that, also in real life activities, the same mechanism appears, where the human brain is able to extract some sort of reward function based on the performance of the executed task, in terms of accomplishment of the desired objective (see e.g. Cohen, 2008). Keeping this in mind when planning training and

rehabilitation programs, it is no surprise that the increase in efficacy in terms of motor outcomes when games are present is a clear-cut consequence of a more efficient training strategy.

In conclusion, games might represent an effective way of strengthening rehabilitation programs, not only as the result of an increased motivational effect, but also because they entail and trigger a more efficient learning paradigm, based on reinforcement learning. Experimental data targeted to dissect motivational aspects from learning factors might help in supporting/confuting this hypothesis.

REFERENCES

Broeren, J., Claesson, L., Goude, D., Rydmark, M., and Sunnerhagen, K. S. (2008). Virtual rehabilitation in an activity centre for community-dwelling persons with stroke: the possibilities of 3-dimensional computer games. *Cerebrovasc. Dis.* 26, 289–296.

Cohen, M. X. (2008). Neurocomputational mechanisms of reinforcement-guided learning in humans: a review. *Cogn. Affect. Behav. Neurosci.* 8, 113–125.

Erev, I., and Roth, A. E. (1998). Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* 88, 848–881.

Johnson, M. J. (2006). Recent trends in robot-assisted therapy environments to improve real-life functional performance after stroke. *J. Neuroeng. Rehabil.* 3, 29.

Jordan, M. I., and Rumelhart, D. E. (1992). Forward models: supervised learning with a distal teacher. *Cogn. Sci.* 16, 307–354.

Kwakkel, G., van Peppen, R., Wagenaar, R. C., Wood Dauphinee, S., Richards, C., Ashburn, A., Miller, K., Lincoln, N., Partridge, C., Wellwood, I., and Langhorne, P. (2004). Effects of augmented exercise therapy time after stroke: a meta-analysis. *Stroke* 35, 2529–2539.

Minsky, M. L. (1961). Steps toward artificial intelligence. *Proc. IRE* 49, 8–30.

Mirelman, A., Bonato, P., and Deutsch, J. E. (2009). Effects of training with a robot-virtual reality system compared with a robot alone on the gait of individuals after stroke. *Stroke* 40, 169–174.

Sutton, R. S., and Bart o, A. G. (1998). Reinforcement Learning: An Introduction. Cambridge, MA, MIT Press.

Tin, C., and Poon, C. S. (2005). Internal models in sensorimotor integration: perspectives from adaptive control theory. *J. Neural Eng.* 2, S147–S163.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. Learn.* 8, 229–256.

Zaghloul, K. A., Blanco, J. A., Weidemann, C. T., McGill, K., Jaggi, J. L., Baltuch, G. H., and Kahana, M. J. (2009). Human substantia nigra neurons encode unexpected financial rewards. *Science* 323, 1496–1499.