Automatic planning in cognitive training: application to multiple sclerosis

Published Version:
Mauro Gaspari, Federica Pinardi, Dario Signorello, Franca Stablum, Sara Zuppiroli (2024). Automatic planning in cognitive training: application to multiple sclerosis. HUMAN-COMPUTER INTERACTION, 39(1-2), 79-95 [10.1080/07370024.2023.2241452].

Availability:
This version is available at: https://hdl.handle.net/11585/938594 since: 2024-07-12

Published:
DOI: http://doi.org/10.1080/07370024.2023.2241452

(Article begins on next page)

DESIGN

Automatic Planning in Cognitive Training: Application to Multiple Sclerosis

Mauro Gaspari[a] and Federica Pinardi[d] and Dario Signorello[b] and Franca Stablum[bc] and Sara Zuppiroli[a]

[a]Department of Computer Science and Engineering, University of Bologna, Via Mura Anteo Zamboni, 7, Bologna, Italy; [b]Department of General Psychology, University of Padova, Italy;[c]Human Inspired Technology Research Centre, University of Padova, Padova, Italy; [d]IRCCS, Neuroscience, Bellaria Hospital, Bologna, Italy

ARTICLE HISTORY
Compiled July 12, 2023

ABSTRACT
Multiple Sclerosis is the second most common cause of neurological disability among young adults. Cognitive impairment, which typically worsen over time, is a major symptom of MS. Sign of cognitive impairment can be observed in many cognitive domains often including executive functions disorders. Planning is one of the main skills related to executive functions, it is fundamental for many cognitive and motor tasks. Brain games, initially available in a paper-and-pen format, have been designed to improve planning abilities. Current computerized cognitive training tools also include this kind of exercises; however, they have several limitations which can be addressed exploiting automated planning. This solution enables advanced forms of human-computer interaction, but poses several design challenges. In this paper, we tested the usability of two cognitive training exercises for executive functions based on automated planning which include different features and interaction mechanisms, then we present the results of a multidomain cognitive training addressed to individuals affected by MS including the exercise that performed better in the test. The aim of this study is to clarify design issues concerning executive functions exercises based on automated planning, showing that they can be actually used in a multidomain cognitive training with participants affected by MS.

KEYWORDS
Cognitive Training, Multiple Sclerosis, Executive functions, Planning.

1. Introduction

The incidence of neurodegenerative diseases like Multiple Sclerosis, Parkinson's disease, Alzheimer's disease (Gironi et al., 2016) is increasing due to the progressive ageing of the population (Hung et al., 2010; Hou et al., 2019). In this context Multiple Sclerosis (MS), a chronic, immune-mediated, and neurodegenerative disorder that is characterized by the inflammation and progressive demyelination of neurons, specifically those ones of central nervous system, stands out from other diseases due two main factors. On the one hand MS also affects younger subjects while most of the cases involve elderly in other disease. On the other hand, MS has a wider variety of potential

CONTACT M. Gaspari. Email: mauro.gaspari@unibo.it

symptoms which depend on the unpredictability of the underlying inflammatory processes also affecting cognitive functions. Multiple Sclerosis is the second most common cause of neurological disability among young adults (Compston & Coles, 2008). Cognitive impairment, which typically worsen over time, is a major symptom of MS. Its onset may be even at early stages of the disease, with a prevalence ranging from 43% to 70% of the whole affected population (Benedict & Zivadinov, 2011). The resulting cognitive dysfunctions in this cohort of patients include deficits in attention, working memory, processing speed, verbal memory, verbal fluency, and executive functioning (Chiaravalloti & De Luca, 2008) and may have a significant and negative impact on quality of life (Kalmar et al., 2008). A recent study (De Meo et al., 2021) involving 1212 MS patients identifies 5 different cognitive phenotypes. Most of the patients are characterize by a mild to severe multi-domain cognitive involvement, and only 19.4% showed preserved cognition. In summary, sign of cognitive impairment can be observed in many cognitive domains often including executive functions disorders, and cognitive rehabilitation for MS patients should possibly address all of them (Gaspari et al., 2023).

Executive functions are the set of higher-order processes widely accepted as fundamental components of human cognition. Executive dysfunctions may lead to cognitive, behavioural and social interaction problems and to difficulties in performing routine household tasks. Exploiting neuroplasticity and cognitive reserve, cognitive training can be effective in mitigating executive dysfunction. Planning is a crucial executive function, and is required in many cognitive and motor tasks. Brain games, initially available in a paper-and-pen format, have been designed to improve planning abilities. Often they are designed as ecological exercises which resemble everyday life activities, like Plan-A-Day proposed by Funke & Kruger, (1995) (see Holt et al., (2011) for a computerized version), where subjects have to schedule a list of tasks to complete during a day, while considering various constraints, or the Zoo Map, developed in 1996, which is part of the Behavioral Assessment of the Dysexecutive Syndrome (BADS) battery (Oosterman et al., 2013; Wilson et al., 1996). Exercises with a high ecological validity showed positive impact on the participants' quality of life (Zelinski, 2009).

State of the art computerized cognitive training tools also include similar type of exercises (Naeeni Davarani et al., 2022; Kaller et al., 2011; Rios et al., 2020; Messinis et al., 2017; Nguyen et al., 2019; Holt et al., 2011), however they have several limitations, for example the set of problems to solve is limited to a few cases and the iteration with the user is primitive. To overcome these limitations, we have explored an approach based on automatic planning (Nau, 2007), where the training task is defined as a planning problem (Baschieri et al., 2018) and the planner interacts with the user during the training activity. This approach enables advanced forms of human-computer interaction, but poses several challenges (Gaspari & Donnici, 2019). For example, a planner can be used to check the feasibility of dynamically generated exercises, or to validate the actions of patients at runtime, provided that the underlying planning problem is carefully designed.

In this paper, we tested the usability of two cognitive training exercises for executive functions based on automated planning which include different features and interaction mechanisms, then we present the results of a multidomain cognitive training addressed to individuals affected by MS including the exercise that performed better in the test.

The usability test involved 30 healthy adults (age ranging from 19 to 56 years) for a first evaluation of two different design approaches:

- A *synchronous approach* where all the actions of the participants are immediately checked by the planner, which ends the exercise as soon as it becomes unsolvable. In this way participants have an immediate feedback on their actions. These features are implemented in Day of Commitments, a novel computer-based version of the Plan-A-Day (Baschieri et al., 2018).

- An *asynchronous approach* where the planner does not interact with the participants, and is only used to evaluate the provided solution. A new version of the Zoo Map, named the ZooSafari Visit, exploits this approach (Gaspari et al., 2023).

At the end of this study, that was conducted during the lockdown period, questionnaires were administered to evaluate usability and the type of strategies used by the participants to plan the activities. The results of this test led us to add the ZooSafari Visit, which performed better, in the set of exercises to be administered to individuals affected by MS.

Successively we have conducted a multidomain cognitive training with 13 individuals affected by MS with preserved cognition or mild cognitive dysfunctions, according to the phenotypes presented in the De Meo et al., (2021) study.

The aim of this research is to give an answer to the following questions concerning the design of executive functions exercises based on automated planning: Which are the more effective interaction mechanisms to embody in planning exercises? Should exercises at different difficulty levels be previously generated and checked, or is it possible to generate them automatically during training? Can we observe an impact on users due to the two different implementation approaches? It is feasible to include exercises based on automatic planning in multidomain cognitive rehabilitation sessions with patients affected by MS? Can MS patients effectively use these exercises? Which is the impact of the whole multidomain training on MS patients?

2. Design approaches for Executive Functions Exercises

To enhance and rehabilitate the planning skills of people with cognitive disabilities two elements are crucial:

(1) the development of systems that allow automatic generation of a sufficient number of new exercises;
(2) the development of systems that allow the user to implement different solution strategies.

These issues, which are difficult to tackle with traditional paper-and-pen techniques, can be addressed in computerized systems exploiting automated planning in the design of serious games for training executive functions (Baschieri et al., 2018). For example, new exercises can be generated automatically including different stimuli, constraints and goals, and the planner can be used to check their feasibility, and the actual level of difficulty. Indeed, the precise level of difficulty of these kinds of exercises mainly depends on the number of steps needed to reach a solution rather than the number of stimuli involved. Moreover, different solution strategies can be assessed by comparing them with the optimal solution computed by the planner, to determine the performance of the patient. A further improvement with respect to paper-and-pen approaches concerns user interaction, for example the planner can be called after every action of the patient to check if the problem can still be resolved. Nevertheless, this solution poses serious efficiency problems because the planner needs to answer in a

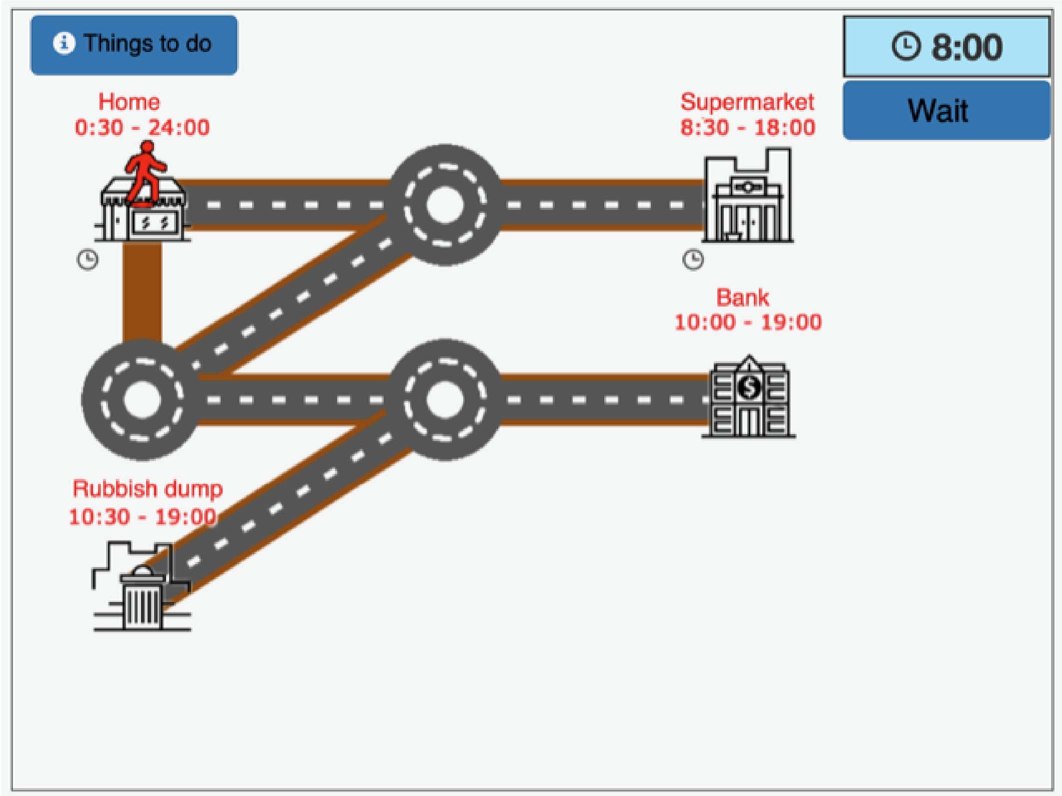

Figure 1. An automatically generated map of the Day of Commitments exercise.

short time, no more than a few seconds for each step. As a consequence, a careful design of the planning problem is needed and the complexity of the proposed tasks may be limited. An alternative approach, is to avoid control at every step, and provide asynchronous interaction mechanisms that subjects can use during the training, for example:

- A "Verify" button which explicitly calls the planner to check whether the problem is still solvable or not. In the latter case, it will inform the patient of which goal is insolvable, giving them the options to either continue in order to satisfy the other goals, or end the exercise.
- A "Hint" button to give the subject a suggestion on which action to perform next.

The introduction of these additional features may have an impact on the difficulty of the exercises and on the performance of the users.

Both synchronous and asynchronous design approaches were used in the past for implementing serious games based on automated planning. For example, the execution control architecture for simulating the behaviour of virtual infantry soldiers (Menif et al., 2017) requires a kind of synchronous re-planning activity when some situations arise; while the blocks world based game for rearranging anagrams (Do & Tran, 2013) only exploits the planner asynchronously for ranking different problems and evaluating user performance.

Considering cognitive training exercises, as far as we know, our approach is the first that exploits automated planning. The exercises we have found are all based on ad-hoc solutions and do not embody an automated planner. For example, the implementation of Plan-A-Day presented in Holt et al., (2011) only provides eight fixed problems with increasing difficulty levels, and the performance is based on calculating the solution time rather than the goodness of the plan. The user can only interact using a "back" button to click when he/she finds a mistake in the plan.

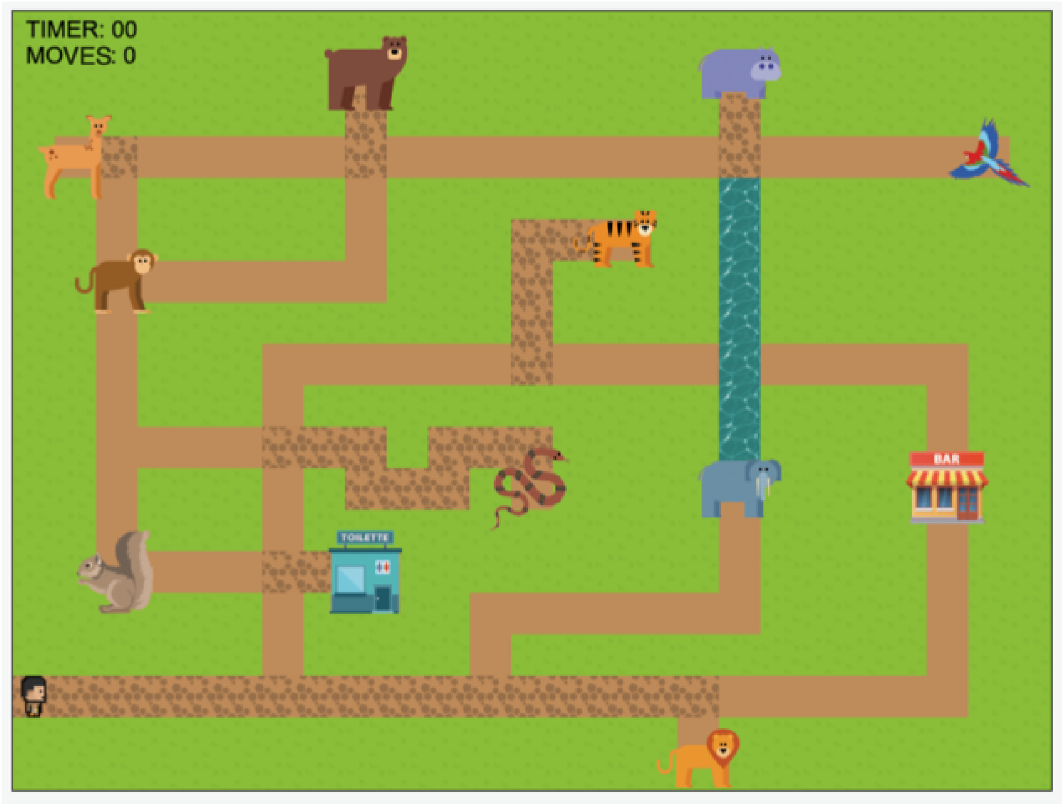

Figure 2. ZooSafari Visit.

2.1. Automated Planning Based Exercises

The planning exercises used in this study are both based on the Planning Domain Definition Language (PDDL) (Haslum et al., 2019) a well-known automatic planner, exploiting PDDL4J (Planning Domain Description Library for Java) (Pellier & Fiorino, 2018), an open source Java toolkit that provides state-of-the-art planners based on PDDL.

The exercises are part of the MS-rehab software (Gaspari et al., 2023) and only a web browser is needed to use them. MS-rehab provides an interface for administrators (rehabilitation staff) that allows them to assign exercises to subjects involved in the training and monitor their activity, and an interface for patients. Caregivers are not considered as category of users, and thus they can only access the system using the account of patients.

The first exercise is Day of Commitments, inspired by the Plan-A-Day (Funke & Kruger, 1995). The patient is given a series of unordered tasks to be carried out in a day (e.g., going to the bank to withdraw money or taking their daughter to the swimming pool), a map of the city where the tasks have to be executed, and a set of time constraints for the places on the map (e.g. the bank opening hours, or the time when the daughter starts her swimming course). The patients have to find a plan that allow them to execute all the tasks, respecting the constraints, and optimizing the number of actions. An example of an automatically generated map is presented in Figure 1; the user explores this two dimensional map using the mouse.

The second exercise is ZooSafari Visit. This exercise was inspired by the Zoo Map test of the BADS battery (Oosterman et al., 2013). It can briefly be described as an exercise where the patient must plan his/her route by referring to a map of the ZooSafari and visiting a selection of animals and locations (see Figure 2). Arrows on the keyboard are used for moving in the map. Initially, there are a number of different routes which the patient can choose, but there is a river that can be crossed only once. Therefore, the number of choices become more limited.

The performance in both the exercises takes into account how far the plan found by the patient deviates from the optimal one (Baschieri et al., 2018).

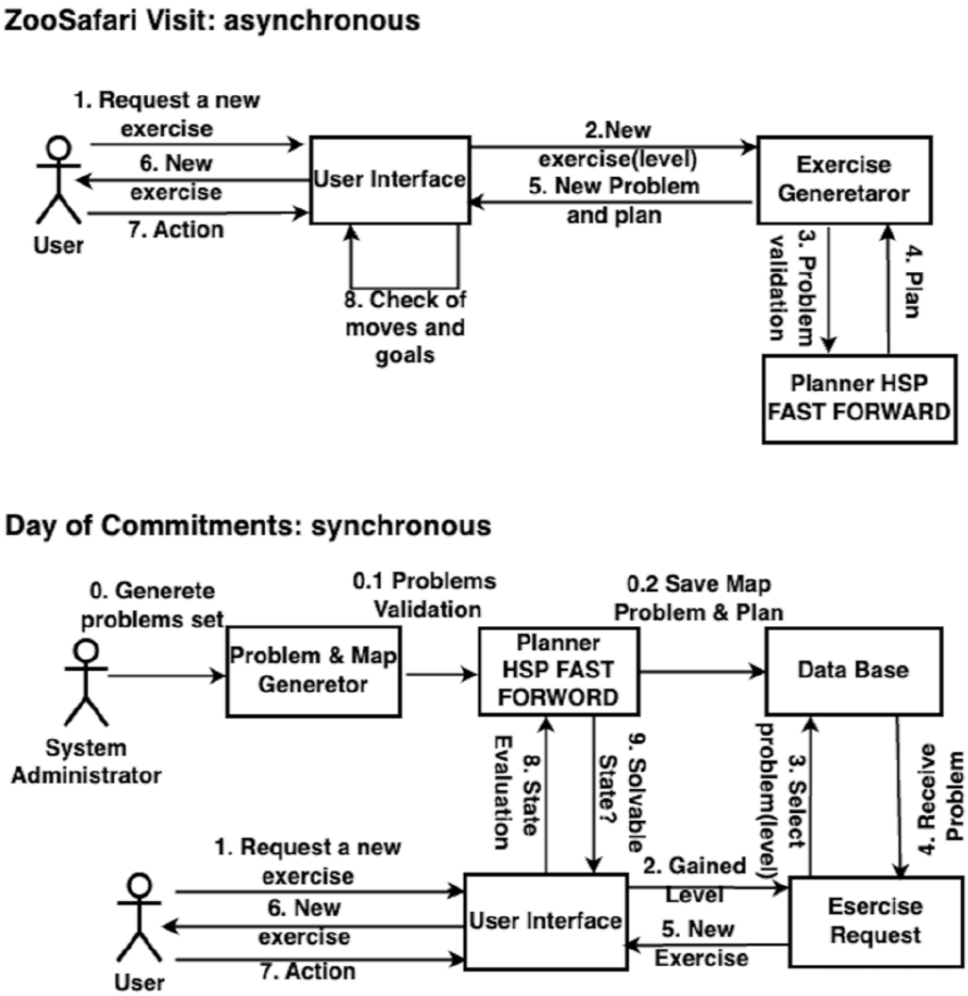

Figure 3. Synchronous vs asynchronous solution. Boxes indicate the components of the system and numbered arrows the sequence of communications among them. The user starts the process with a request for a new exercise, and then he/she performs a sequence of actions. In Day of Commitments we also model the initial phase where a system administrator generates the available exercises (these arrows starts with 0).

Figure 3 illustrates how automated planning is used in these exercises considering both the synchronous and the asynchronous solutions. We use a notation like communication diagrams (Fowler, 2004) which highlights interactions of system's components using numbered arrows. The user only interacts with the user interface component; all the other arrows indicate operations of the system.

In ZooSafari Visit problems are generated dynamically when a request for a new exercise at a given level is issued. The Heuristic Search Planner (HSP) (Pellier & Fiorino, 2018) using the FAST FORWARD heuristic checks if a solution for the generated plan exists. Actions of the user are not checked by the planner, the test indicated with the arrow (8. Check of moves and goals) only verifies if there are still possible moves and goals to satisfy.

On the contrary, in Day of Commitments different problems and maps were generated by the system administrator in a batch mode and stored in a database including the solution plan that the HSP planner has found for them (see Section 2.2 for details). During training these exercises are retrieved from the database when a request for selecting a problem at a given level is issued, see arrow (3. select problem(level)). In this synchronous approach after each action of the user in the map, the HSP planner is called (arrows 8 and 9) to verify if the problem can still be solved from the reached state. The FAST FORWARD heuristic is also used for this test.

In the following, we briefly show how PDDL is used in the domain specification of ZooSafari Visit, for more details on the encoding of Day of Commitments, see Baschieri et al., (2018). In this exercise the map is fixed, whereas places and connections are part of the domain definition. The available actions are also defined in the domain, in terms of their preconditions and effects using a set of predicates and functions. For example, the following PDDL rule represents an action that allows the visitor to move from a location to another.

```
(:action Walk parameters (?visitor ?from ?to)
  :precondition
      (and (visitor ?visitor) (at ?visitor ?from) (road ?from ?to)
          (not (= ?from ?to)) (able-to-walk ?from ?to))
  :effect
      (and (at ?visitor ?to) (not (to-discover ?to)) (not (at ?visitor ?from))
          (not (able-to-walk ?to ?from)) (not (able-to-walk ?from ?to))))
```

Names starting with a question mark represent variables. This action can be executed if all the predicates in the precondition match with facts in the current state, here negation as failure is used. The positive facts in the effect part are added to the state, and the negative ones removed. The following predicates are used in this rule:

- (at ?thing ?place): an object ?thing is at position ?place.
- (visitor ?v): ?v is a visitor of the ZooSafari.
- (road ?from ?to): a road that connect two places.
- (to-discover ?place): a place that need still to be visited.
- (able-to-walk ?from ?to): it is possible to walk from two places.

When defining a specific problem, all involved objects are defined and initialized, and the planning goals are set. Using the ZooSafari Visit domain several problems can be generated, which might differ by the number of initial facts and goals.

2.2. Automatic generation of planning problems

Traditional pen-and-paper exercises used for cognitive rehabilitation usually have a small number of scenarios. To overcome this limitation and provide a good number of different situations to solve, we developed a flexible scenario generator in both the exercises, but the approaches used are different.

In Day of Commitments the scenario generator consists of two modules:

- a Map Generator creates a realistic environment with buildings, roundabouts, and items;
- a Problem Generator, given a map, generates solvable problems for that map.

The generators work on two predefined lists. The first includes about ten places (e.g., bank, home, station etc.) that are potentially part of a scenario; for each place, we specified a set of attributes, such as the opening hours, the items that can be taken or left, and the actions the patient can execute. The second list includes about twenty-five items; for each item we specified the actions the patient can perform. The map generator initially selects a random number of places, connects them with main roads (to assure a fully connected graph), and adds further random connections. Then, the generator adds to each place a random number of items (among those associated with it). This approach ensures randomness in the structure of the game without renouncing consistency.

Subsequently, the problem generator creates a set of PDDL goals, i.e., the tasks to be executed by the patient. Goals are related to staying (e.g., be-at, stand) or to obtaining and releasing items (e.g., got-at, drop-at). Each goal is created by randomly selecting a PDDL predicate and populating its fields coherently with the previously generated map. The created goals are added to the problem one by one, until the problem has reached the required difficulty (Baschieri et al., 2018).

In ZooSafari the scenario Generator is able to generate n! different situations, where n is the number of possible objectives, considering only solvable ones. Differently from Day of Commitments the map does not change, and only the order of the elements showed and the number of one-way paths is affected.

2.3. Usability Test

The usability test of the above planning exercises was conducted choosing a heterogeneous group of 30 healthy individuals (12 males and 18 females) ranging from 19 to 56 years with an average age of 30 (SD: 11) and an average schooling of 15 years (SD: 2.03). At the beginning of the test an informed consent was signed by all the participants and a Cognitive Failures Questionnaire (CFQ) (Broadbent et al., 1982) was administered to all of them. The analysis of CFQ results shows on average good executive function abilities in the selected sample (M: 36; SD: 11.5).

The test was conducted at home autonomously during the lock down period, exploiting the MS-rehab web based interface, providing a link to the online version of the MS-rehab user manual, and with the suggestion of contacting the operators in case of problems. The assigned task was to perform at least three rehabilitation sessions a week that they were free to organize in accordance with their commitments. Each session should have lasted 40 minutes to be divided between the two planning exercises. The sessions were monitored exploiting the remote control facilities of MS-rehab (Gaspari et al., 2023).

At the end of the training two questionnaires were administered: a satisfaction survey to evaluate software usability, and a metacognition questionnaire to investigate the planning strategies adopted in the training. Moreover, a meeting was scheduled at the end of the training for all the subjects that were free to participate or not.

The estimation of internal coherence of these questionnaires using the Cronbach Alpha coefficient (Brown, 2002) (acceptable values > 0.65) gave the following values: 0.903 for the satisfaction survey and 0.71 for the metacognition questionnaire.

The qualitative analysis of the satisfaction questionnaire, 14 questions based on a 5 items Likert scale expressing goodness, shows that the majority of participant was satisfied by the proposed training (M: 3.30, SD: 0.75), and would take part again to a similar training (M: 3.67, SD: 1.09), similar grades were obtained for all the questions. A point that is worth to highlight is that the answer to the question "Do you think that the fact that the training was carried out online had a negative influence on the obtained results?" obtained a low score (M: 2.33, SD: 1.27), indicating the effectiveness of the online training.

From a detailed analysis of the opinion of the single participants in the final meeting emerged that, some of the low scores were influenced by reported problem in the user-friendliness of the Day of Commitments exercise, which require more complex technical solution due to the synchronous interaction with the planner. Indeed, the results shows that only 11 subjects completed the training in the Day of Commitments exercise (on average 6 attempts were necessary), while almost all the participants (29) were able to reach the last level in the ZooSafari Visit in four or five attempts, thanks to an improved accessibility of the technical solution adopted.

With regards to the metacognition questionnaire, 9 question based on a 5 items Likert scale (ranging from 1 to 5) expressing accordance, the results show that the majority of participants reported that the training was useful to improve their planning abilities (M: 3.50, SD: 0.82), despite the fact that most of them (M: 4, SD: 0.87) declared to have good planning abilities.

Concerning the scenario generator testing we found that:

- The scenario generator of ZooSafari Visit is more efficient than that of Day of Commitments because it has less possible actions, pre and post conditions to consider. Thus it returns a solvable problem in reasonable time, enabling automatic generation of new situation during training.
- Given that in the ZooSafari Visit approach the difficulty level is not influenced by a changing map, but only by the number of targets and one way paths, the learning process was faster reaching the maximum difficulty lever in fewer steps.

The first aspect also follows from well-known computational complexity studies on planning problem generators (Bäckström & Nebel, 1995), while the second was confirmed by the results reported above.

The comments reported by users involved in the training activity and the errors they found were crucial for identifying problems and solving them: In some situations, there was a discrepancy between the actual difficulty of exercises and the level to which they were proposed. Some of the instructions of the tested exercises were ambiguous and/or difficult to understand. The provided feedback sometimes was not correct, and the users reported that the feedback would be crucial to improve their strategies and increase their motivations.

The gathered information was precious for improving several aspects of the two exercises: the level of some exercises was changed; ambiguous instructions were made clearer; the feedback was improved.

For the Day of Commitments exercise, the following changes have been made: the initial instructions have been improved; the opening hours of places have been made visible on the map; a specific button have been added to view the remaining goals during the training (this action is not considered in the plan). For the ZooSafari Visit exercise the following changes have been made: the initial instructions have been improved; specific instructions have been provided on how to use the arrows and pause buttons to move around the map; a maximum time in which exercises should be finished has been established.

The results of this usability study led us to choose the ZooSafari Visit as a planning exercise for cognitive rehabilitation with MS patients, because, featuring a better usability and reliability, and it appeared to be more well suited to fit with the cognitive difficulties that MS patients normally shown. Also the training with MS patients will be conducted at home in full autonomy.

3. Application to Multiple Sclerosis

A multidisciplinary team including a Neurologist, a Neuropsychologist, a Computer Scientist and a member of AISM Bologna was set up to conduct the training with individuals affected by MS. The role of AISM (Italian Association for Multiple Sclerosis) was to recruit individuals affected by MS and to coordinate the activities. The cognitive training project was structured in 4 main phases, following the approach presented in Gaspari et al., (2023), without including a personalized rehabilitation project for each participant. We designed a playful project constituted by two sets of predefined exercises, the same for all the participants.

3.1. Initial screening and multidimensional assessment

In the first phase we identified 13 individuals affected by MS with preserved cognition or mild cognitive dysfunctions. They were contacted by AISM. After an initial meeting with the multidisciplinary team that hold at the AISM premises, where the training project was presented, participants were asked to sign an informed consent. Given that activities were conducted on a voluntary basis at home, they did not necessarily require an ethical committee approval.

Thirteen participants began our project but only 10 completed the training and did the final assessment, 3 participants decided to stop their training for personal reasons while they were doing it. Thus, in the following we only include data of the 10 participants that completed the test, 6 of them having a Relapsing-Remitting Multiple Sclerosis (RRMS) and 4 a Primary Progressive Multiple Sclerosis (PPMS), with an MS diagnosis done 5 years ago or more. Before the training participants have undergone a neuropsychological assessment at the AIMS premises in order to identify their specific deficits. The RAO battery (Amato et al., 2006) was administered at the AISM premises by the neuropsychologist for participants who were without an assessment (assessments administered three months before the training have considered valid). More precisely, the administered subtests were the following: Selective Reminding Tests (Long Term Storage SRT-LTS, Consistent Long Term Retrieval SRT-CLTR, and Delayed SRT-D), Spatial Recall Test (SPART), and on Symbol Digit Modality Test (SDMT), Paced Auditory Serial Addition Test 2 sec. and 3 sec. (PASAT 2 and 3).

Using the results of this assessment it was possible to classify the participants in the different phenotypes illustrated in the De Meo et al., (2021) study.

Table 1 shows how many participants were classified in the proposed phenotypes after the initial and final assessments.

Table 1. How participants involved in the training are classified in the phenotypes presented in the De Meo et al., (2021) study according to the initial (before training) and final (after training) assessments.

Phenotype	Initial	final
preserved cognition	3	5
mild–verbal memory/semantic fluency comprised	4	3
mild-multidomain comprised	3	2
severe–executive/attention comprised	0	0
severe–multidomain comprised	0	0

3.2. Definition of the Multidomain Rehabilitation Program

In the second phase the multidisciplinary team defined two sets of exercises. The first set, simpler than the second, included four exercises to stimulate attention and three to stimulate memory. The second set contained four exercises to stimulate attention, three to stimulate late memory and one to stimulate executive functions (the ZooSafari Visit exercise). This second set of exercises was constituted by more complex tasks requiring a higher level of concentration and was designed to be carried out only when the first training had terminated. The cognitive training activity was carried out at home, thanks to the MS-rehab web interface and remote monitoring facilities. Each patient had to connect to the system for 30 minutes twice a week, in order to complete the exercises that presented an increasing degree of difficulty. All the cognitive training exercises were done in total autonomy, this was possible because all the exercises presented in our system are auto-adaptive, therefore, they automatically increase in difficult, in relation to personal abilities, without the need of operator intervention.

3.3. Cycle of Cognitive Training

Before the training, meetings were scheduled to illustrate the MS-rehab software to participants. The training process was the third phase, it began for one group of participants in February and a second group at the end of March. A technical support was provided during all the training. Each set of exercises lasted thirty minutes and comprised both memory and attention exercises. Participants who achieved at least 70% in the work assigned could move onto the second set of exercises.

When a group of participants finished the first set of exercises, a meeting with the multidisciplinary team was organized to present the new exercises. It should be noted that not all the participants were able to complete the first set of exercises. Thus, some participants have not started the second set of exercises, they spent all the time (four months) on the first group of exercises.

3.4. Final assessment

After the completion of the second level, or in any case after four months from the beginning of the cognitive training, the participants were called for a neuropsychological evaluation repeating the RAO battery (this was the fourth phase). A first group of participants started their training in February and they received the second set of exercises in April, stopping the rehabilitation in June. This group was tested in June. A second group started at the end of March they received the exercises second set in June and they stopped their training in September/October. This second group was tested in November. For this second group we did not take in consideration the months of July and August as rehabilitation time.

Table 2. Exercises administered in the two sets of the training. For each exercise it presents the maximum level reached from participants, the number of participants that tried it, and the maximum level proposed by the system.

Set	Exercise name	Max level reached	N. Users trying it	Max Level
1	Alternative-attention vegetable	12	13	12
	Divided-attention face/sound	12	13	12
	Attention-selective-memorization arrow	12	13	13
	Attention-selective fruit	14	13	14
	Visual-memory fruit	12	13	12
	Visual-memory vegetable	10	13	10
	Working-memory arrow	8	11	9
2	Alternative-attention cardinal point	12	8	12
	Divided-attention arrow/sound	12	6	12
	Attention-selective-memorization face	13	9	13
	Attention-selective cardinal point	14	8	14
	Memory face	12	8	12
	Visual-spatial-memory fruit	11	8	11
	Working-memory face	7	6	9
	Planning-ZooSafariVisit	10	7	10

4. Results

The data collected in the training though our system are summarised in Table 2. This Table presents for each exercise the maximum level reached from participants in the training, the number of participants that tried it, and the maximum level proposed by the system. This table shows that mainly the Working-memory exercises were not completed by any patient.

Tables 3 and Table 4 report, respectively, the result of the training with the first and with the second sets of exercises. These tables present for each cognitive function stimulated (first column) the exercises difficulty, the average trials carried out for each exercise at each difficulty level, the number of participants that attempted that level, the average time in minutes spent at that difficulty level, and the minimum number of trial needed to reach that difficulty level determined by the MS-rehab system.

Finally, Table 5 presents the detailed impact of the training process on single participants showing the Equivalent Scores in RAO subtests before and after the training. An improvement can be observed for several participants. We considered an Equivalent Scores of three or four as normal, of two as borderline, and of 0 or 1 as deficit. Positive results are highlighted in green, negative in red, and white cells indicate that the score did not change. We considered a difference of at least 2 equivalent scores with respect to the results obtained in the initial assessment as an improvement or worsening. The yellow columns, identified by t1, show the initial assessment results and the columns identified by t2 show the final assessment results.

Table 3. This table presents the details of the exercises administered in the first set. For each exercise at each difficulty level we report: the average number of attempts done by all the participants, the number of participants that reached that level; the average time spent at that level (in minutes), and the minimum number of attempts needed to pass that level established by the system.

	Difficulty Level	Average number of attempts	Number of participants at that level	Average time spent (minutes)	Min. number of attempts to pass
Alternating-attention vegetable	Easy	11	13	9	6
	Medium	11.1	11	8	8
	Difficult	9.6	10	6	10
Divided-attention face/sound	Easy	16.2	13	162	8
	Medium	14	9	89	8
	Difficult	8.6	9	64	6
Attention-selective -with-memorization arrow	Easy	15.8	13	29	6
	Medium	15.5	11	41	8
	Difficult	22.6	10	69	10
Attention-selective fruit	Easy	10.5	13	4	8
	Medium	10.8	12	7	8
	Difficult	12.6	11	13	12
Visual-memory fruit	Easy	7.8	13	1	6
	Medium	11.2	11	2	8
	Difficult	12.1	11	3	8
Visual-memory vegetable	Easy	8.3	13	1	8
	Medium	27	11	6	6
	Difficult	26.5	8	9	6
Working-memory arrow	Easy	9.1	11	8	6
	Medium	23.3	9	25	6
	Difficult	7.3	3	13	7

5. Discussion

The presented results allow us to give an answer to the main questions we highlighted in the introduction. First, we demonstrated that it is feasible to design a multidomain training program which include exercises based on automatic planning for cognitive rehabilitation of MS patients and that they can effectively use these exercises.

Considering the complete training program, a comparison between results presented in Table 3 and Table 4, which report the differences between the first and of the second sets of exercises, highlights that participants involved in the second set of exercises learned from their previous training activity with the first set. Furthermore, comparing the exercises that stimulate the same cognitive ability in the same two tables, we can see how attention and memory exercises done in the second set needed fewer attempts than the same typology of exercises in the first set. This was possible because the exercises were the same changing the set of used stimuli only. Note that, this happen despite the fact that stimuli used in the second set were more complex (vegetables vs cardinal points; face/sound vs arrow/sound; arrow vs cardinal point), indicating that subjects learned in the first part of the training.

Table 4. This table presents the details of the exercises administered in the second set. For each exercise at each difficulty level we report: the average number of attempts done by all the participants, the number of participants that reached that level; the average time spent at that level (in minutes), and the minimum number of attempts needed to pass that level established by the system.

	Difficulty Level	Average number of attempts	Number of participants at that level	Average time spent (minutes)	Min. number of attempts to pass
Alternating-attention cardinal point	Easy	6.4	8	5	6
	Medium	7.7	7	5	8
	Difficult	7.5	4	5	10
Divided-attention arrow/sound	Easy	8.3	6	26	8
	Medium	5.8	4	25	8
	Difficult	12	2	664	6
Attention-selective -with-memorization face	Easy	8	9	23	6
	Medium	10	9	30	8
	Difficult	25.3	6	66	10
Attention-selective cardinal point	Easy	11.5	8	7	8
	Medium	15.2	6	20	8
	Difficult	24.4	5	38	12
Memory-long -time face	Easy	11.3	8	4	8
	Medium	14.6	7	9	8
	Difficult	9,5	2	8	6
Visual-spatial -memory fruit	Easy	8,6	8	1	8
	Medium	17.8	8	4	6
	Difficult	10	7	3	6
Working-memory face	Easy	6.7	6	8	6
	Medium	11.4	5	24	6
	Difficult	1.5	2	1	6
Planning Zoo Safari Visit	Easy	11.6	7	10	6
	Medium	16.5	4	18	6
	Difficult	20.5	2	19	6

Table 5. This Table presents for each patient the impact of the training process with respect to the equivalent scores obtained in the RAO battery (Amato et al., 2006) (t1 initial assessment and t2 final assessment). Yellow rows indicate the results in the initial assessment, positive results are presented in green, negative in red, and neutral in white (differences of two units are considered only). SRT-LTS: Selective Reminding Test Long Term Storage; SRT-CLTR: Selective Reminding Test Consistent Long Term Retrieval; SRT-D: Selective Reminding Test Delayed; SPART: Spatial Recall Test; SDMT: Symbol Digit Modality Test; PASAT 3: Paced Auditory Serial Addition Test 3 sec; PASAT 2: Paced Auditory Serial Addition Test 2 sec.

	participants - MS type																			
Test name	1 - PP		2 - PP		3 - PP		4 - RR		5 - RR		6 - RR		7 - PP		8 - RR		9 - RR		10 - RR	
	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
SRT-LTS	4	4	2	4	3	4	2	0	0	4	3	4	4	4	4	4	0	0	4	4
SRT-CLTR	4	4	2	4	3	4	4	0	1	3	3	2	4	4	2	4	0	0	4	4
SRT-D	4	4	2	4	4	4	3	3	3	3	4	2	4	4	4	4	0	0	4	4
SPART	2	4	3	4	0	1	4	4	4	4	4	4	0	4	4	4	0	4	0	1
SDMT	4	4	4	4	4	4	4	4	4	2	4	4	3	4	2	3	0	0	1	4
PASAT 3	4	4	2	2	4	4	4	4	4	4	4	4	4	4	4	3	2	4	0	4
PASAT 2	4	4	3	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	3	4

Interestingly, in the case of executive functions (the ZooSafari Visit), although this planning exercise was scheduled in the second set of exercises, it required twice the number of attempts with respect to the minimum necessary for other exercises. This observation showed that planning based exercises are in general more complex with respect to the others. It means that the choice of including them in the second set was correct.

Second, and most importantly, addressing the last question posed in the introduction, we showed that a multidomain cognitive rehabilitation program including planning based exercises can have a positive impact on MS patients. Indeed, considering the single individual performance, it appears that after training there was a general improvement, and two other participants resulted without deficit in the RAO test, as a result a total of 5 participants can be classified in the first phenotype, as shown in Table 1.

Analysing in more details the results presented in Table 5, participants 1, 2, 5, 7, 8, 9, 10 obtained at least in one subtest higher equivalent scores with respect to the initial assessment: participants 1, 2 and 8 starting from a borderline evaluation in some subtest obtained normal equivalent scores in all of them; participant 2 in 3 subtests and participants 1 and 8 in one subtest; participants 5, 7, 9, and 10 starting from a deficit in one or more RAO subtests reached an equivalent score of 4 after training. Also, the fact that many participants showed stable assessments may be interpreted as a positive result, considering the progressive nature of MS.

As a final consideration, we summarise the guidelines that came out from our studies for the design of cognitive rehabilitation multidomain programs for MS patients:

(1) To schedule initial training sessions for presenting the system interface. Indeed, some patients can be unfamiliar with computerized systems in general which can have a negative impact on both training and motivations.
(2) To provide a personalized feedback to patients after each rehabilitation exercise. We noticed that patients are interested to this information, which they try to use to understand what was going wrong and to improve their performance in the subsequent exercises. This features also concern planning based exercises.
(3) To provide a personalized feedback (alarm signals) to administrators after some sessions of performed exercises (the number of sessions may be personalized). This personalized alarms may help the administrators to follow the progress and difficulties of patients enabling prompt intervention.
(4) To provide mechanisms for remote monitoring and manual set up the difficulty level of exercises. As an example, these features can be used for maintaining the difficulty at a given level avoiding feeling of frustration. MS-rehab supports these mechanisms for all the exercises including those based on planning.
(5) To add a two-week personal contact with the patients. We observed that this contact has a positive impact on commitment and motivation of the user.
(6) To perform a careful tuning of the administered exercises, and in particular of those based on planning which are more complex. This tuning activity should address the interface, the difficulty level progression mechanism and the feedback given to users.
(7) To carefully design the training program with increasing levels of difficulty, through exercises meant to firstly train simpler cognitive abilities and, subsequently, more complex ones (such as planning).

6. Conclusion

Although the presented results are far from being statistically significant, they allow us to give some preliminary answers to the research questions we have posed at the beginning of the paper. The asynchronous approach for integrating a planner seems to be more effective, especially to avoid efficiency problems that may arise during the execution making the interaction slow with possible time out situations. The advantage of having an immediate (synchronous) checking of their actions seems to be not crucial for subject, with respect to an accurate performance estimation, computed using the planner, which can be obtained in both the approaches.

The automatic generation of planning problems during the training was demonstrated feasible, indeed the ZooSafari Visit uses this approach and relevant problems were not reported neither in the usability test nor in the training with individuals affected by MS. The impact of the introduction of a planning based exercise in the training was positive. We also demonstrated that exercises based on automatic planning can be successfully introduced in cognitive training with participants, and that the impact of the multidomain rehabilitation project on them was positive.

The current study has some limitations:

- The small sample size does not allow appropriate generalizations, and therefore further studies are required.
- The simple interaction mechanism provided in the ZooSafari Visit, using arrows in the keywords, was adequate for all the participants. However, integrating a joystick like in the Rehacom system (Naeeni Davarani et al., 2022) would probably increase the number of possible users with a MS diagnosis.
- We only used a first level assessment specific for MS (Amato et al., 2006), instead of a deeper one addressing executive functions.

Estimating the specific impact of planning based exercises on single patients, also considering real life activities, will be an objective of our future work.

In our ongoing research we are addressing the following issues: developing and testing new planning based cognitive training exercises with an high ecological validity, like Week-end-in-Rome (Gaspari & Donnici, 2019); improving the feedback efficacy of these exercises, for example studying the impact of different mechanism for scoring the performance like those used in planning competitions (Do & Tran, 2013); exploring more sophisticated training tasks which requires the additional expressive power of PDDL 2.1 including fluents and plan metrics (Fox & Long, 2003); investigating the use of the Expressive Numeric Heuristic Search Planner (ENHSP) (Scala et al., 2016) that supports the above features of PDDL 2.1; testing the developed planning based exercises with different categories of users, with the goals of making them available for the cognitive improvement of healthy aging people, for example.

Acknowledgments

We would like to thank Dott. Sergio Stecchi: thanks to his idea this project came out. We also would like to thank all the AISM Bologna staff involved in the project: Adolfo Balma, Anna Fiorenza, Roberta Gollini, Carlo Mestitz, Manuela Panico, and all the participants, their contribution was fundamental for the success of this project. Finally, we would like to thank the anonymous referee for their comments that helped to improve this paper.

Funding

This project was partially supported by AIMS Bologna.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Amato, M. P., Portaccio, E., Goretti, B., Zipoli, V., Ricchiuti, L., De Caro, M. F., Patti, F., Vecchio, R., Sorbi, S., & Trojano, M. (2006). The Rao's Brief Repeatable Battery and Stroop Test: normative values with age, education and gender corrections in an Italian population. *Multiple Sclerosis Journal*, 12(6), 787-793. https://doi.org/10.1177/1352458506070933

Bäckström, C., & Nebel, B. (1995). Complexity results for SAS+ planning. *Computational Intelligence*, *11*(4), 625-655. https://doi.org/10.1111/j.1467-8640.1995.tb00052.x

Baschieri, D., Gaspari, M., & Zini, F. (2018). A planning-based serious game for cognitive rehabilitation in multiple sclerosis. Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good. New York, NY, USA. ACM. (pp. 214–219). https://doi.org/10.1145/3284869.3284916

Benedict, R. H., & Zivadinov, R. (2011). Risk factors for and management of cognitive dysfunction in multiple sclerosis. *Nature reviews. Neurology*, *7*(6), 332–342. https://doi.org/10.1038/nrneurol.2011.61

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*(1), 1–16. https://doi.org/10.1111/j.2044-8260.1982.tb01421.x

Brown, J. D. (2002). The Cronbach alpha reliability estimate. *JALT Testing & Evaluation SIG Newsletter*, *6*(1).

Chiaravalloti, N. D., & DeLuca, J. (2008). Cognitive impairment in multiple sclerosis. *The Lancet. Neurology*, *7*(12), 1139–1151. https://doi.org/10.1016/S1474-4422(08)70259-X

Compston, A., & Coles, A. (2008). Multiple sclerosis. *Lancet*, *372*(9648), 1502–1517. https://doi.org/10.1016/S0140-6736(08)61620-7

De Meo, E., Portaccio, E., Giorgio, A., Ruano, L., Goretti, B., Niccolai, C., Patti, F., Chisari, C. G., Gallo, P., Grossi, P., Ghezzi, A., Roscio, M., Mattioli, F., Stampatori, C., Simone, M., Viterbo, R. G., Bonacchi, R., Rocca, M. A., De Stefano, N., Filippi, M., … Amato, M. P. (2021). Identifying the distinct cognitive phenotypes in multiple sclerosis. *JAMA neurology*, *78*(4), 414–425. https://doi.org/10.1001/jamaneurol.2020.4920

Do, M., & Tran, M. (2013). Blocksworld: An iPad puzzle game. *In Proceeding of the 3rd International Planning in Games Workshop, ICAPS 2013* (pp. 35-39).

Fowler M. (2004). *UML distilled: a brief guide to the standard object modeling language.* Addison-Wesley Professional.

Fox, M., & Long, D. (2003). PDDL2.1: An extension to PDDL for expressing temporal planning domains. *Journal of Artificial Intelligence Research*, *20*, 61-124. https://doi.org/10.1613/jair.1129

Funke, J., & Kruger, T. (1995). *Plan-A-Day*: Konzeption eines modifizierbaren Instruments zur Führungskräfte-Auswahl sowie erste empirische Befunde [Plan-A-Day: design of a modifiable instrument for selecting managerial personnel and first empricial results]. In J. Funke, & A. Fritz, (Eds.), Neue Konzepte und Instrumente zur Planungsdiagnostik (pp. 97–120). Bonn (DE): Deutscher Psychologen Verlag.

Gaspari, M., & Donnici, M. (2019). Weekend in Rome: a cognitive training exercise based on planning. SAT@SMC.

Gaspari, M., Zini, F., & Stecchi, S., (2023). Enhancing cognitive rehabilitation in multiple sclerosis with a disease-specific tool. *Disability and Rehabilitation: Assistive Technology, 18*(3), 313-326. https://doi.org/10.1080/17483107.2020.184943.

Gironi, M., Arnò, C., Comi, G., Penton-Rol, G., & Furlan, R. (2016). Multiple sclerosis and neurodegenerative diseases. In D. Boraschi & G. Penton-Rol (Eds.), *Immune Rebalancing* (pp. 63-84). Academic Press. https://doi.org/10.1016/B978-0-12-803302-9.00004-X

Haslum, P., Lipovetzky, N., Magazzeni, D., & Muise, C. (2019). An introduction to the planning domain definition language. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 13*(2), 1-187. https://doi.org/10.2200/S00900ED2V01Y201902AIM042

Holt, D. V., Rodewald, K., Rentrop, M., Funke, J., Weisbrod, M., & Kaiser, S. (2011). The Plan-a-Day approach to measuring planning ability in patients with schizophrenia. *Journal of the International Neuropsychological Society, 17*(2), 327-335. https://doi.org/10.1017/S1355617710001712

Hou, Y., Dan, X., Babbar, M., Wei, Y., Hasselbalch, S. G., Croteau, D. L., & Bohr, V. A. (2019). Ageing as a risk factor for neurodegenerative disease. *Nature Reviews. Neurology, 15*(10), 565- 581. https://doi.org/10.1038/s41582-019-0244-7

Hung, C. W., Chen, Y. C., Hsieh, W. L., Chiou, S. H., & Kao, C. L. (2010). Ageing and neurodegenerative diseases. *Ageing Research Reviews, 9*(1), S36-46. https://doi.org/10.1016/j.arr.2010.08.006

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research, 216*(1), 1–8. https://doi.org/10.1016/j.bbr.2010.07.029

Kalmar, J. H., Gaudino, E. A., Moore, N. B., Halper, J., & DeLuca, J. (2008). The relationship between cognitive deficits and everyday functional activities in multiple sclerosis. *Neuropsychology. 22*(4), 442–449. https://doi.org/10.1037/0894-4105.22.4.442

Menif, A., Guettier, C., & Cazenave, T. (2013). Planning and execution control architecture for infantry serious gaming. *In Proceeding of the 3rd International Planning in Games Workshop, ICAPS 2013.*

Messinis, L., Nasios, G., Kosmidis, M. H., Zampakis, P., Malefaki, S., Ntoskou, K., Nousia, A., Bakirtzis, C., Grigoriadis, N., Gourzis, P., & Papathanasopoulos, P. (2017). Efficacy of a computer-assisted cognitive rehabilitation intervention in relapsing-remitting multiple sclerosis patients: A multicenter randomized controlled trial. *Behavioural Neurology*, 5919841. https://doi.org/10.1155/2017/5919841

Naeeni Davarani, M., Arian Darestani, A., Hassani-Abharian, P., Vaseghi, S., Zarrindast, M. R., & Nasehi, M. (2022). RehaCom rehabilitation training improves a wide-range of cognitive functions in multiple sclerosis patients. *Applied Neuropsychology: Adult, 29*(2), 262-272. https://doi.org/10.1080/23279095.2020.1747070

Nau, D. S. (2007). Current trends in automated planning. *AI Magazine, 28*(4), 43. https://doi.org/10.1609/aimag.v28i4.2067

Nguyen, L., Murphy, K., & Andrews, G. (2019). Immediate and long-term efficacy of executive functions cognitive training in older adults: A systematic review and meta-analysis. *Psychological Bulletin, 145*(7), 698–733. https://doi.org/10.1037/bul0000196

Oosterman, J. M., Wijers, M., & Kessels, R. P. (2013). Planning or something else? Examining neuropsychological predictors of Zoo Map performance. *Applied Neuropsychology: Adult, 20*(2), 103–109.

Pellier, D., & Fiorino, H. (2018). PDDL4J: a planning domain description library for java. *Journal of Experimental & Theoretical Artificial Intelligence, 30*(1), 143-176. https://doi.org/10.1080/0952813X.2017.1409278

Ríos Cruz, S. G., Olivares Pérez, T., Hernández Expósito, S., Bolívar Barón, H. D., Dowens, M., & Betancort Montesinos, M. (2020). Efficacy of a computer-based cognitive training program to enhance planning skills in 5 to 7-year-old normally-developing children. *Applied Neuropsychology. Child, 9*(1), 21–30. https://doi.org/10.1080/21622965.2018.1503959

Scala, E., Haslum, P., Thiébaux, S., & Ramírez, M. (2016). Interval-based relaxation for general numeric planning. *European Conference on Artificial Intelligence, ECAI 2016* (pp. 655-663).

IOS Press.
Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Bury St. Edmunds, Suffolk: Thames Valley Test Company.
Zelinski, E. M. (2009). Far transfer in cognitive training of older adults. *Restorative Neurology and Neuroscience*, *27*(5), 455-471. https://doi.org/10.3233/RNN-2009-0495