The International Journal of Indian Psychology
ISSN 2348-5396 (Online) | ISSN: 2349-3429 (Print)
Volume 8, Issue 3, July- Sep, 2020
DIP: 18.01.064/20200803, DOI: 10.25215/0803.064
http://www.ijip.in

The International Journal of INDIAN PSYCHOLOGY

Research Paper

"What is the significance of feedback and reinforcement learning mechanisms in aiding dynamic decision making?" - proposing a revised hybrid reinforcement learning initiated instance-based model of dynamic decision making

Nilosmita Banerjee[1]*

ABSTRACT

Aim: The current study aimed to propose a new hybrid-model called Reinforcement-Learning initiated Instance-based Model which acknowledges the significance of feedback in promoting learning and memory-mechanisms over a period of time to aid Dynamic Decision-Making (DDM), a crucial aspect disregarded by a previous DDM-model called Instance-based Learning Model proposed by Gonzalez et.al.,(2003) as also test this hybrid-model in a novel Financial Expenditure DDM-environment with varying income-periods like income/employment and retirement-phase. ***Method:*** In a mixed-measures experimental-design the current study subjected participants *(n=48)* to a financial DDM-task wherein they undertook discretionary-expenditures during varying income-periods (income-phase/retirement-phase), across two conditions—feedback and no-feedback in two separate decision-making sessions (DMS) 1&2 in each condition. ***Results:*** Results indicated that participants in feedback-condition (DMS-1&2) outperformed in the financial DDM-task as compared to those in no-feedback (DMS-1&2) condition. Furthermore, participants in feedback-condition in DMS-1 performed significantly better in DMS-2 without explicit-feedback, while those in no-feedback condition in DMS-1 deteriorated in DMS-2 in DDM-task without feedback, suggesting that feedback facilitated development of memory-instances during the learning-process in DMS-1 which aided DDM in similar DDM-environment (DMS-2). ***Conclusion:*** The current study highlighted the significance of feedback in aiding learning and memory-mechanisms in DDM-settings.

Keywords: *DDM, Instance-based Learning Model, Reinforcement-Learning, Feedback*

Have you ever faced a hard time selecting a fast-food corner to eat from at a food-court in a shopping mall? Or felt overwhelmed with multiple brand options for a single product to choose from, in a supermarket? Although these situations typify simple occurrences faced in everyday life, what these questions actually represent are real-

[1]MSc student in Cognitive Computational Neuroscience, Dept. of Psychology, University of Sheffield (United Kingdom)
*Responding Author

Received: July 08, 2020; Revision Received: August 21, 2020; Accepted: August 23, 2020

world, complex, high-risk, everchanging, dynamic-environments in which we make decisions (Gonzalez, 2014) and decisions made in such dynamic-settings is referred to as Dynamic Decision-Making (DDM; Gonzalez et.al., 2017).

Dynamic Decision Making (DDM)

DDM has been defined as undertaking multiple, inter-dependent decisions in real-time wherein the decision-making environment is impacted both autonomously and as a consequence of the decisions made (Edwards, 1962). In a DDM-environment, the decision-maker is concerned with controlling a dynamic system which is complex, unfolds in unpredictable ways and is often (if not always) subject to temporal-constraints (Jagacinski & Miller, 1978). Most decisions undertaken in naturalistic-settings such as managing choice-explosion faced during grocery shopping in supermarkets (e.g., Gonzalez, 2014), selecting an optimal route when driving, undertaking financial decisions like stock-market investment decisions made during constant price changes (Brehmer,1992), as also decisions made in specialised dynamic-settings such as military commanding, air-traffic controlling or health management in hospitals etc., count as dynamic-environments in which DDM is undertaken (Gray, 2002).

Studying human DDM in real-world dynamic-settings has been a challenge for DDM-researchers (Funke, 1988;1995). Inorder to cope with this difficulty DDM-researchers have resorted to using virtual simulations of real-world dynamic-settings in computer simulated tasks called microworlds (Martin et.al., 2004), which are built by incorporating DDM-characteristics and are used for investigating human DDM (e.g., Rapoport, 1966; Brehmer & Dorner, 1993; Brehmer & Nahlinder, 2007). Although virtual laboratory tasks, these microworlds present decision-makers with similar dynamic-constraints faced in real-world DDM-environments (e.g., Erbert, 1972) and therefore their decision-performance in these tasks is considered to provide adequate insight into decision-makers actual decision-making tendencies and performance in real DDM-settings (e.g., Kerstholt, 1994;1996). A key finding past microworld-based laboratory DDM studies highlighted about human DDM is that, overall decision-makers performance in DDM-settings is extremely poor (Anzai, 1984). In explaining why this might be, past researchers state that this poor performance can be attributed to the lack of mental-models and required cognitive-capacity to deal with DDM-characteristics in such environments, which leads to decision-making incompetency in-turn causing poor performance (see Kerstholt and Raaijmakers, 1997 for a review).

Although past researchers focused on building upon these initial findings by solely focusing on factors that lead to poor decision-performance in dynamic-settings, (e.g., Sterman, 1994), recent DDM researchers state that highlighting sub-optimal performance in DDM-environments or microworld tasks, do not shed light on the decision-processes and decision-rules decision-makers employ in undertaking DDM (Gonzalez & Dutt, 2011a,b). It also does not enable us to determine how these decision-processes and decision-rules can aid decision-makers to gain expertise in a specific DDM-environment, further equipping them with the decision-making skills required to manoeuvre in similar DDM-environments (Hotaling et.al., 2015). Post this acknowledgement, recent research efforts have shifted their focus from factors impacting DDM to focusing on investigating the above stated aspects (Gonzalez, 2005). In investigating the main reasons for poor decision-making performance as accounted in past studies, recent researchers state that the main reason for poor performance in these past studies was not because decision-makers lacked mental-models

and the cognitive-capacity to deal with the DDM-characteristics as suggested by past researchers (e.g., Sterman, 1989a,b) but it was mainly due to lack of repeated-exposure to these DDM-environments (Gonzalez et.al., 2003). Researchers further state that inorder to perform well, decision-makers need to establish a causal relationship between the decision-choice and decision-outcome—a crucial aspect required to build mental-models of DDM-environments (Brown et.al., 2009) and that this can be only established following repeated-exposure to DDM-environments (Gonzalez & Quesada, 2003). To test the impact of repeated-exposure on decision-performance in DDM-settings, Gonzalez et.al., conducted a series of behavioural studies wherein they repeatedly exposed participants to a microworld task. Results of this study indicated that as predicted, over a period of time consistent repeated-exposure led to improved decision-performance (Gonzalez et.al., submitted for publication).

Gonzalez et.al., (2003) explained that this change in decision-performance was an outcome of repeated-exposure to the DDM-environments, which helped decision-makers learn the decision-choice—decision-outcome relationship and aided in building mental-models of the DDM-environment, which were stored as instances in their memory and was used to further guide their decision-performance in future trials or similar DDM-situations. Gonzalez et.al., (2003) thus recognised the role of learning and memory-mechanisms in DDM and for the first time conceptualised this in a theoretical-framework called Instance-Based Learning Theory (IBLT). They further proposed a model based on IBLT called Instance-Based Learning Model (IBLM) which provided a detailed step-by-step explanation of how decision-makers employ memory-based decision-processes and use learning-based decision-rules to build the mental- models (memory-instances) of decision-choice—decision-outcome relationship.

Although, IBLT/IBLM have been long used to explain human DDM, a critical analysis of IBLT and IBLM has highlighted two major crucial gaps pertaining to—a.) the proposed decision-rule IBLM prescribes that decision-makers use to establish decision-choice—decision-outcome link and b.) undermining the role of feedback as a crucial step in aiding learning in DDM. An extensive search of the literature indicated that past research has neither identified nor addressed these gaps in IBLT/IBLM. Thus, this study will be the first to both identify and address these gaps. In doing so the study aims to extend beyond the current IBLT theoretical-framework and draw explanations from a recently emerging theory in DDM research called Reinforcement-Learning theory (RLT; Sutton & Barto, 1998) to address the identified gaps and propose a new hybrid-model which combines concepts from both RLT and IBLT. Finally, we aim to test this model in a novel DDM-environment focusing on financial discretionary-expenditure decision-making across differing income periods of life.

Instance Based Learning Theory (IBLT)

Although the significance of cognitive-processes of learning and memory in DDM and complex tasks had been appreciated previously (Simon, 1955; Edwards, 1961), a formal recognition of these cognitive-processes in DDM was only acknowledged after Gonzalez et.al., (2003) proposed the theoretical-framework called IBLT. The IBLT proposes that with repeated-exposure and practice decision-makers learn to employ a heuristic-based decision-rule (HBDR) to select decision-choices. Based on a satisficing-policy which entails searching through available alternatives until a sub-optimal decision-choice threshold is met

(e.g., Simon, 1957; Simon & March, 1958; Ken, 2000; Colman, 2006), HBDR assists decision-makers to form a knowledge-base of high-utility decision-choices which is stored in the decision-makers memory in the form of instances. Drawing influence from the decision sciences, Case-Based Decision Theory (Reisbeck & Schank, 1989), Gonzalez et.al., (2003) state that these memory- instances which form the instance-based knowledge-based consist of a tripartite-structure and are composed of three core elements—situation—which refers to a set of environmental cues, decision—which refers to the decision-choice selected and action taken and utility—which refers to the goodness of the selected decision-choice (instances = Situation-Decision-Utility [SDU]; e.g., Ram, 1993). IBLT further proposes that following consecutive interactions with similar DDM-settings, decision-makers engage in memory-based decision-processes which involves the stages of recognition, evaluating/judgement, decision-choice and decision- execution, stages which are explained in the IBLM.

IBLM

Based on IBLT, IBLM is a closed-loop model of DDM (see Fig.1) which explains and illustrates the steps decision-makers undertake post employing HBDR to learn and store optimal, high-utility decision-outcomes, in the memory-based decision processes in a similar DDM-environment. The IBLM states that in a similar DDM-environment the decision-maker begins with the first step of recognition—of environmental cues in the similar DDM-setting, which if appraised as a typical cue (i.e., matches an already stored memory-instance of a similar decision situation from the past), then goes on into the next step of judgment/evaluation— wherein the decision-maker assesses the most optimal past outcome with the highest utility in a similar DDM-context. Following this they proceed to the last step—decision-choice wherein the decision-maker chooses the selected high-utility decision-choice from the past stored memory-instance/SDU before executing the decision-choice. Another step called feedback is added in the IBLM as a part of the memory-based decision-process, which is not explained or acknowledged in the original IBLT theoretical-framework. The IBLM states that feedback is a part of the memory-based decision-processes and is only used to refine or modify an already stored memory-instance/SDU, provided the stored memory-instance/SDU does not generate the predicted high-utility decision-outcome as the past situation (which qualified it the first time to form the memory-instance/SDU).

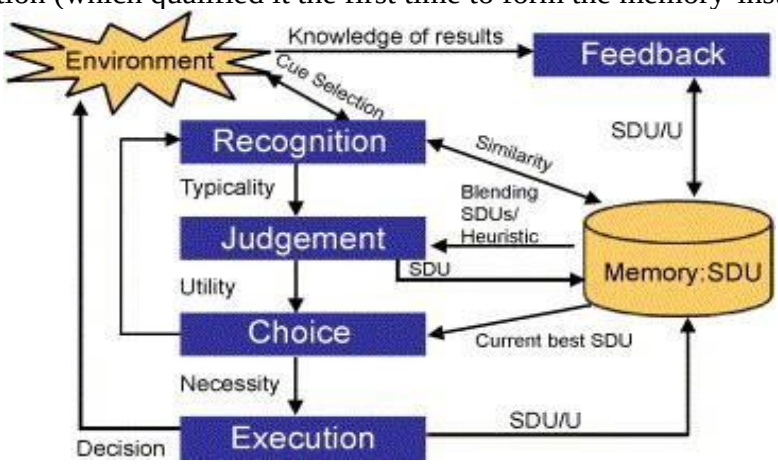


Figure.1: The IBLM closed-loop learning circuit illustrating the main steps comprising of the memory-based decision-processes in DDM in the order in which they are undertaken by the decision-maker in a DDM environment as proposed by Gonzalez et.al., (2003)

Gonzales et.al., (2003) conducted a study to test the propositions of IBLT/IBLM wherein they subjected participants to a microworld task called the Water Purification Plant (WPP), a simulated water-distribution system. The participants participated in the WPP task for a total of 18 trials for 3 consecutive days wherein the researchers provided them with the overall end goal of WPP. Gonzalez and colleagues stated that WPP did not present the participants with any feedback. Results indicated that post repeated-exposure of WPP, participants overall decision-performance significantly improved from day-1 to day-3, thereby showing evidence of learning. Gonzalez et.al., (2003) stated that the repeated-exposure to WPP enabled the participants to engage in HBDR which helped them to learn the optimal decision-outcomes by establishing the decision-choice—decision-outcome link. In future trials in the consecutive days, participants demonstrated the use of similar high-utility options that worked best in the previous days indicating that in similar DDM-situations decision-makers recognized typical environmental cues, judged/evaluated, chose and executed/selected the already stored high- utility decision-choices from their memory/SDU. Although, overall this study provided evidence for the main propositions of IBLT pertaining to repeated-exposure to DDM-environments leading to accumulation and formation of high-utility instances and retrieval, evaluation and selection of these high-utility decision-choices in similar DDM-settings, it did not provide, test or explain the role of feedback in DDM-environments.

Gaps highlighted in IBLT/IBLM

Although IBLT/IBLM provides a comprehensive explanation of the learning-based decision-rule and memory-based decision-processes employed by decision-makers in DDM-settings, a critical analysis of the IBLT/IBML has raised two main gaps pertaining to the proposed heuristic-based decision-rule (HBDR) and the undermined role of feedback in DDM- environments.

Issue I: Proposed Learning-based Decision-Rule in IBLT/IBLM

The IBLT/IBLM proposes that the learning-based decision-rule employed by a decision-maker during their initial interaction with the DDM-environment to generate decision-outcomes is based on a Heuristic-Based Decision-Rule (HBDR; Gonzalez et.al., 2003) which is based on a satisficing-policy wherein the decision-maker sets a sub-optimal decision- outcome threshold and searches alternatives only till the point this threshold is met, before executing the decision-choice (Richardson, 2017). They repeat this until they have stored enough instances of successful decision-outcomes in their memory, before shifting from relying on decision-rule for generating decision-outcomes to relying on memory-based decision-processes for generating decision-outcomes (Gonzalez et.al., 2003). Although IBLM attempts to explain how decision-makers use HBDR to learn the decision-outcome—decision- choice link and to generate high-utility decision-choices, there is a problem with this proposition which pertains to the use of heuristics to facilitate learning.

Heuristics and systematic-errors

Heuristics refer to a set of metal-shortcuts decision-makers employ during decision- making instead of engaging in long laborious alternative search and judgements (Dawson & Arkes, 1987). According to the adaptive decision-making hypothesis (Payne et.al., 1993) researchers state that heuristics are useful as they save the decision-makers sparing time towards laborious choice selection in decision-making situations (Tversky & Kahneman, 1981). However, the downside of this is that heuristics generate qualitatively poor, sub-

optimal, low-utility decision-outcomes (Kahneman et.al., 1982). This mainly happens because heuristics often lead to cognitive-biases caused due to ignoring crucial information (e.g., base- rate fallacy; Bar-Hillel, 1980) required to build the decision-choice—decision-outcome link in DDM-environments, inturn producing systematic-errors leading to low-utility, sub-optimal and poor decision-outcomes (e.g., Arvai et.al., 2004). A study with a microworld business-firm task presented evidence for this when they found that entrepreneurs using heuristic-strategies to manage the business-firm made poor, low-utility and even erroneous decisions pertaining to management, production and supply decisions in the task (Burmeister & Schade, 2007).

Researchers further state that in dynamic-settings without any active feedback, these negative aspects of heuristic-strategies exacerbate (Hogarth, 1981) as lack of feedback denies decision-makers the opportunity to identify any discrepancy between their decision-choice made and decision-outcome incurred, due to the employment of the heuristic-strategies (Simon, 1956; Kleinmuntz & Thomas, 1987; Plous, 1993). This further disrupts the development of the memory-instances consisting the DDM-environment mental-models or the decision-choice— decision-outcome link (e.g., Brewer, 1987). Cronin et.al (2008) found evidence for this in a study wherein they subjected participants to a dynamic stocks and flow task and found that participants performed poorly in this DDM task inspite of repeated-exposure coupled with no- feedback, because of their use of heuristics, which impeded their learning as the decision-makers failed to establish the decision-choice—decision-outcome link as also failed to form memory-instances consisting the mental-models of the DDM task further leading to poor performance.

In summary, this indicates that although heuristic-strategies save time, decision-rules based on heuristics-strategies do not promote learning the generation of high-utility decision-outcomes due to the above discussed reasons, in DDM-settings (e.g., Fagley & Miller, 1990; Nicholls, 1999) further suggesting that instead of HBDR as proposed in IBLM, a different decision-rule relying on the feedback initiated learning in the DDM-environment, helps decision-makers generate high-utility decision-outcomes.

Issue 2: Undermining the Role of Feedback

The second issue highlighted in IBLT/IBML pertains to the subverted status given to the role of feedback in the learning and memory-mechanisms in DDM-environments. Inspite of the fact that Gonzalez et.al., (2003) have explicitly admitted that the IBLM provides little if any information on how feedback might be accounted for in learning in DDM-environments there has been no active attempts to rectify this in the IBLM, till date. While according to IBLM, feedback is only used to refine an already stored memory-instances/SDU's, recent research exploring the impact of feedback on DDM state otherwise.

Feedback in DDM

The undermined role of feedback in IBLT/IBLM can be attributed to a large-body of past research investigating the role of feedback in DDM-environments that IBLT/IBLM is built on, which has consistently disparaged the effectiveness of feedback in DDM-environments (e.g., Astwood et.al., 2008). Past researchers investigating the impact of feedback in DDM-environments and decision-outcomes state that decision-makers tend to misperceive feedback in DDM-settings, which limits the effectiveness of feedback to promote learning

that can help decision-makers to establish the decision-choice—decision-outcome link (Sterman, 1989a,b).

Sterman (1989a,b) in two studies provided evidence for this when he found that inspite of repeated-exposure coupled with feedback, participants performed poorly and generated low-utility decision-outcomes in managerial decision-making simulated microworld inventory distribution system task.

Although, evidence for misperception of feedback was accounted by DDM studies post Sterman (1989a,b) (e.g., Diehl et.al., 1995), a critical analysis of these past studies indicated that while participants in these past studies were provided with feedback, the same was provided at potential temporal-delays (Sterman, 1989a,b) or the feedback provided was structurally complex (Diehl et.al., 1995) or a combination of both (Kluger & DeNisi, 1996;1998;2000). Researchers state that in order to increase the effectiveness of a given feedback to facilitate learning and aid memory-based decision-processes in DDM-settings, decision-makers should be provided with the feedback in immediate succession post the decision-choice (Osman et.al., 2008). A lack of either an immediate feedback (e.g., Astwood et.al., 2008) and/or a temporally-delayed feedback (e.g., Sterman, 1989a,b) does not allow the decision-maker to incorporate the feedback into their decision-choice, as they fail to account for the decision-choices that caused them in the first place (Gonzalez et.al., 2005). A study found evidence for this, wherein participants showed improved decision-performance post receiving consistent feedback in a simulated microworld Beer Factory task wherein they had to undertake production, supply and distribution decisions (Fu & Gonzalez, 2006).
Furthermore, structurally complex feedback (e.g., Diehl et.al., 1995) which adds to the complexity of a DDM-environment rather than simply providing the information to establish the decision-choice—decision-outcome link, is also detrimental to the effectiveness of feedback in DDM-contexts (see Hsiao & Richardson, 1999 for a review; Hsiao, 2000; Atkins et.al., 2002). Researchers state that most effective feedbacks provide the decision-maker with the task-goal and the current decision-outcome which would enable the decision-maker to detect any discrepancy between the task-goal and the information on the decision-outcome generated, further helping decision-makers learn the most optimal policy to bridge any detected discrepancy (Sengupta & Abdel-Hamid, 1993).

Role of Feedback in learning and memory-mechanisms in DDM

The very recognition towards the significance of feedback in DDM has led researchers to acknowledge that the role of feedback as extremely central to both the learning and memory-mechanisms involved in DDM-environments in the recent times (Burns & Vollmeyer, 2002). In the context of learning-processes in DDM-environments, recent researchers state that providing consistent and reliable feedback during the repeated-exposure is pivotal, as the presence or absence of feedback can dramatically affect the accuracy of the decision-outcome (Berry & Broadbent, 1984;1988). For example, a study conducted by Brand et.al., (2009) provided evidence for this wherein they subjected participants to two separate versions of Game of Dice Task (GDT), one which they called original GDT which provided constant feedback after each trial and one which they called modified GDT which provided no-feedback to the participants after each trial and found that participants in the original GDT had higher net scores than participants in the modified GDT, thereby indicating the significance of feedback in producing high-utility and optimal decision-outcomes (Brand et.al., 2009). In conclusion, researchers, state that feedback impacts learning in a DDM-

environment in three important ways—a.) it enables decision-makers to learn the distinction between high and low- utility decision-outcome by identifying any observed discrepancy between the task-goal and decision-outcome, b.) following an identification of a low-utility decision-outcome decision- makers engage in corrective measures to alter their choice selection till an optimal decision- outcome is reached and finally c.) undertaking this process of fine-tuning of choice selection, enables decision-makers optimize and generate more high-utility decision in successive encounters (Einhorn & Hogarth, 1978;1981;1986; Hogarth, 1981; Kleinmuntz, 1985).

Furthermore, researchers state that while these high-utility decision-outcomes are being undertaken and decision-makers are learning the decision-choice—decision-outcome relationship, they simultaneously tend to store these in their memory-instances/SDU's (Klein et.al., 2010). This indicates that feedback not only helps in refining an already established memory-instance/SDU's as proposed in IBLM (Gonzalez et.al., 2003), but also aids in creating those memory-instance/SDU's in the first place (Klein, 2015). Rhodes and Jacoby (2007) found evidence for this across three experiments. In experiment 1&2 they found that memory of items that participants were repeatedly exposed to in a previous criterion-recognition decision-making task, impacted their decision-performance in a similar decision-making environment. In the third experiment they found that providing performance feedback following each trial, in an initial criterion-recognition decision-making session (block-1&2) not only improved their decision-performance in those decision-making sessions but also positively impacted participants decision-performance in a similar decision-making session in the future (block-3&4) even without performance feedback. Furthermore, they subjected a group of participants to a no-feedback condition (block-1&2) and found that their performance in the criterion-recognition decision-making task degraded across the blocks which provided no-feedback. However, this improved significantly in a similar decision-making environment (block-3&4) when the same participants were subjected to feedback. In explaining these results Rhodes and Jacoby (2007) stated that, providing feedback was crucial as it first facilitated learning and allowed the participants to respond optimally when feedback was provided as also further enabled them to simultaneously store these optimal decision-choices in their memory during this feedback-initiated learning-process, inorder to use them in a similar decision- making environment, even in the absence of feedback. A lack of feedback on the other hand denied the decision-maker the opportunity to undertake an optimization-based learning of the decision-choice—decision-outcome link in the first place as a result of which decision-makers could not store them in the memory to further guide them in a similar decision-making session (c.f., Estes & Maddox, 1995; Verde & Rottello, 2007).

In conclusion recent literature has started acknowledging that feedback impacts both learning and memory-mechanisms in decision-making and this takes place in a closed-loop circuit (see Fig.2)—the proposition being that as the decision-maker learns and establishes the decision-choice—decision-outcome link aided by constant feedback, they simultaneously build memory-instances/SDU's in their memory to use in future similar DDM-settings (e.g., Kahneman & Klein, 2009).

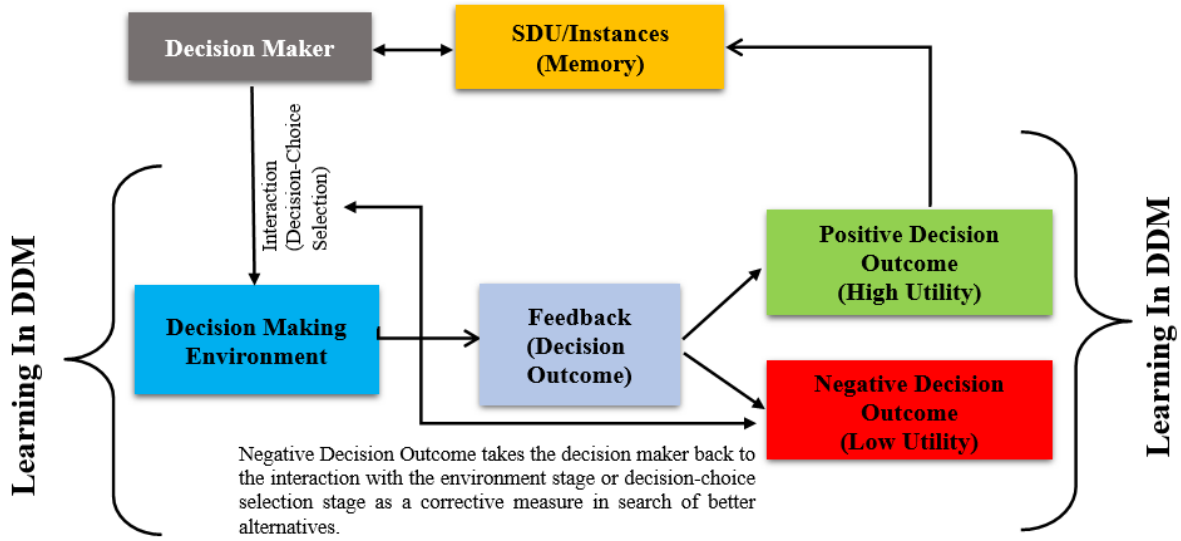


Figure. 2: Diagrammatic representation of how feedback is used in construction of the Memory- instances/SDU's and Learning in DDM

Need for a new Model

These gaps highlighted in IBLT/IBLM indicates a need to propose a model which addresses these gaps. An extensive search of the past literature has indicated that these gaps have neither been identified nor been addressed and thus this study will be the first in addressing these identified gaps by proposing a new hybrid-model of DDM. In doing so we draw influence from the explanations given by Klein (2015) (see Fig.3) who states that optimal decision-outcomes following repeated-exposure to DDM-environment can be generated by feedback initiated optimization focused decision-rules and not a HBDR.

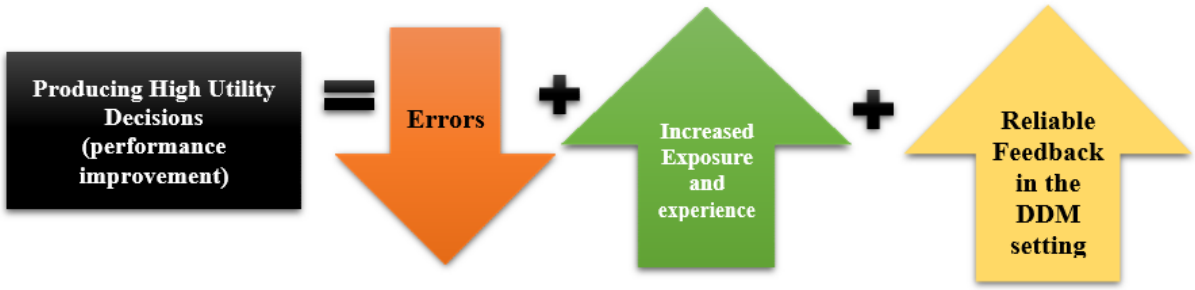


Figure. 3: Conditions which can improve the decision outcomes in the DDM environments as proposed by Klein (2015)

Reinforcement-Learning Theory (RLT)

An extensive search of the literature indicated that a newly emerging theoretical account gaining popularity in DDM research, the Reinforcement-Learning Theory (RLT), prescribes a decision-rule which acknowledges the role of both—a.) undertaking a decision-rule based on an optimisation-policy to produce high-utility decision-outcomes and b.) the role of

feedback- initiated learning in producing these optimal decision-outcomes (Sutton & Barto, 1998). Based on the principles of operant-conditioning, RL is a simple mechanism which states that decision- makers learn to repeat and store those decision-choices which lead to positive-reinforcement (rewards) in their memory while those decision-choices yielding negative-reinforcement (punishment) tend to be terminated (Skinner, 1939; Skinner & Ferster, 1957; see Fig.4).

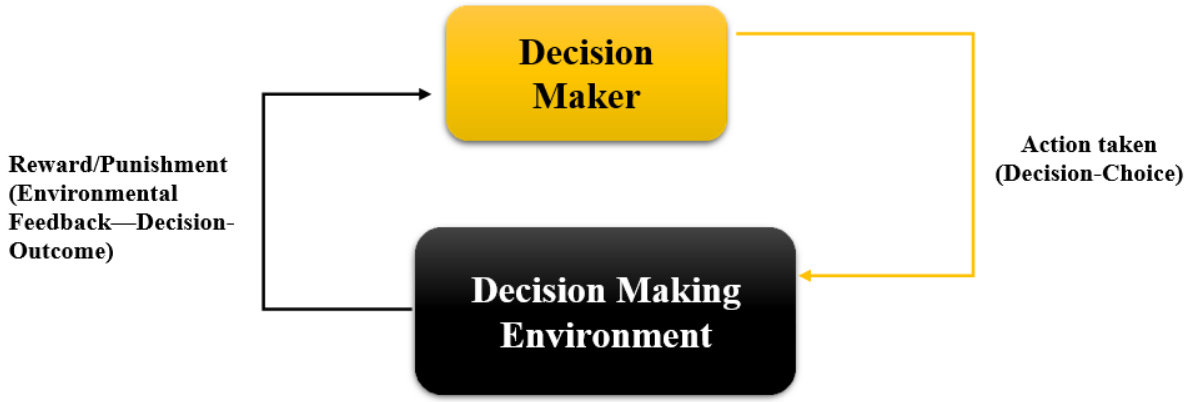


Figure. 4: Diagrammatic representation of the Reinforcement Learning Circuit

Researchers state that as a form of learning RL-mechanisms are most suited to a wide range of dynamic tasks and settings. Repeated-exposure to DDM-environment coupled with feedback in the form of decision-outcomes, helps decision-makers to learn in maximizing their future-goals and minimizing costs by producing optimal decision-outcomes over a period of time (Dimitrakakis & Ortner, 2018). These are then stored in their memory and used in similar instances in future decision-making situations (Busemeyer & Bruza, 2012; Busemeyer & Pleskac, 2009). A particular type of RL algorithm called the model-based RL which produces accurate accounts of human behaviour in DDM-environments (e.g., Anzai, 1984) has demonstrated how people update their model of the environment (i.e., memory) post the reward-punishment cycles during the repeated-exposure in DDM-environment (Simon & Daw, 2011), and has been used in behavioural studies investigating the use of RL-mechanisms in DDM-settings. A series of studies conducted by Gureckis and Love (2009a,b) and Otto and Love (2010) which used this model-based RL in DDM-tasks, found that feedbacks about rewards and punishments received while performing on a simulated DDM-task helped decision-makers optimize their decision-outcomes, and produced high-utility decision- outcomes on later trials following a reward-based decision-outcome. Additionally, they found that decision-makers trying to estimate the dynamics of the environment by experiencing it, create an association between an observed decision-outcome and an earlier decision-choice utilizing memory, facilitated by familiar environmental-cues, to produce high-utility decision- outcomes in similar tasks (Gureckis & Love, 2009a,b; Otto & Love, 2010) . Furthermore, researchers accounting for the robustness of RL-mechanisms in real-world DDM-settings have found that RL-based decision-rules lead to high-quality approximate solutions to large-scale stochastic-planning problems faced in real-world DDM industrial and government settings (Barto, 2004).

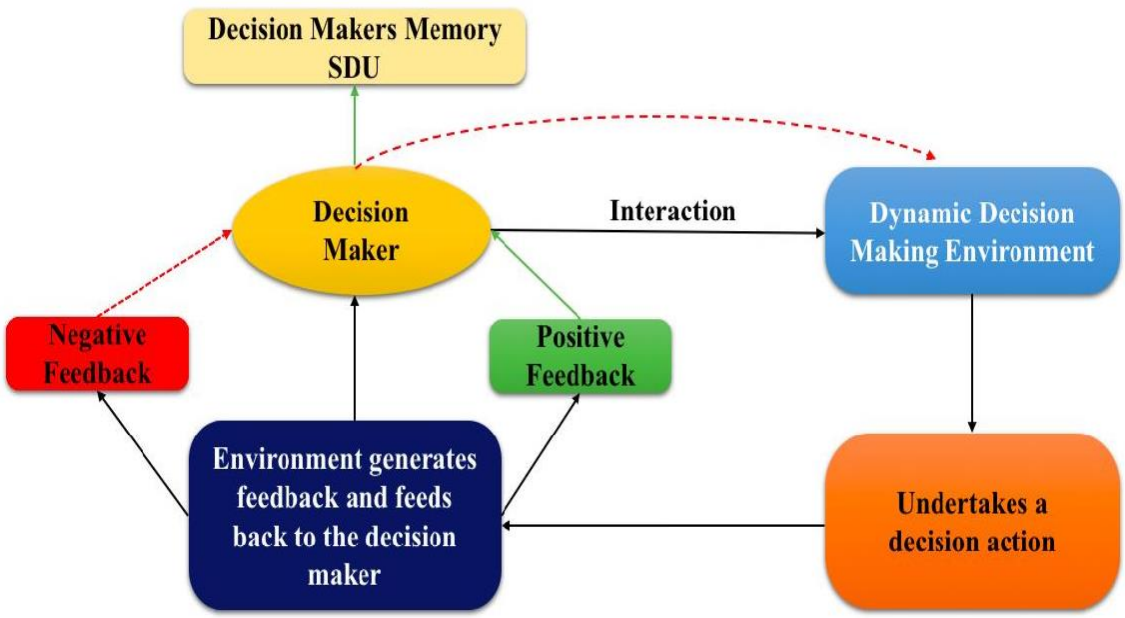
Reinforcement Learning Initiated Instance-Based Model (RLIIBM) of DDM

Proposing a Hybrid Model Of DDM—RLIIBM: RL+IBM

The proposition that RL-based-decision-rules (RLDR) should be used in combination with Instance-based models (IBM) has only recently been acknowledged by fields like artificial intelligence and machine learning, which state that RL-based principles best explain how experience-based learning and memory-instances built and stored while accumulating these experiences, enables the production of high-utility decision-outcomes in IBM (Meyer et.al., 2014). Dutt (2011) explained this in an example of a computational model of IBM which used an RLDR (and not a HBDR) to predict learning and performance in future trials in a dynamic repeated-binary choice task which required participants to select between two choices repeatedly over multiple trials and wherein each alternative affected participant overall earnings.

Drawing influence from these recommendations newly emerging interdisciplinary research prescribes, it seems plausible to propose a hybrid-model which undertakes a combinatorial approach, thereby proposing a model which incorporates RL-IBM mechanisms to address the gaps highlighted in IBLM pertaining to decision-rule and the undermined role of feedback in aiding learning and memory-mechanisms in DDM-settings. Thus, in this paper we propose the Reinforcement Learning Initiated Instance Based Model (RLIIBM), a two staged model comprising of—a.) RL-based decision-rule which is based on a feedback initiated optimization mechanism to facilitate learning the associations between decision-choice and decision-outcome and creating memory-instances/SDU's of high-utility decision-outcomes and b.) IBM based decision-processes based on recognition, judgment, choice and decision- execution, which aims at explaining human DDM (see Fig.5 for a diagrammatic representation of RLIIBM).

Stage 1: Decision Rule– Closed Loop Reinforcement Learning based decision-rule used in DDM to learn the decision-outcome—decision-choice link



Stage 2: Decision Process– Closed Loop Instance-based memory guided decision-processes of the DDM in a similar DDM-environment

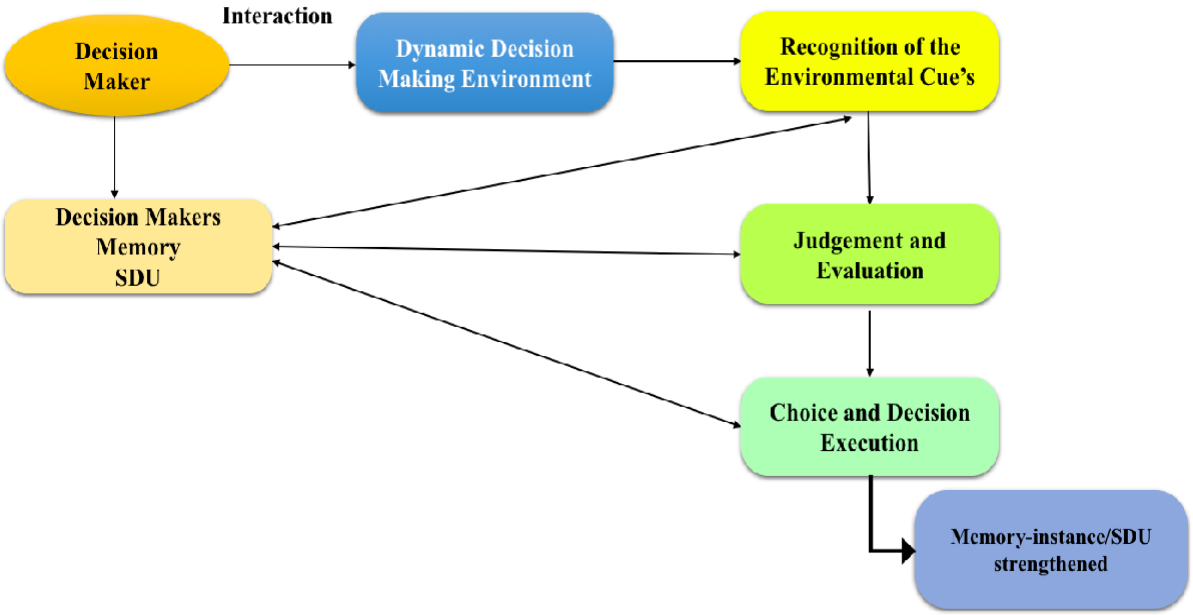


Figure. 5: Reinforcement-Learning Initiated Instance-Based Model (RLIIBM) of DDM

Predictions of RLIIBM

The following are the predictions for the current model proposed in this study:

RLDR and IBM Decision-Process

i.) Decision-makers subjected to DDM-environments providing feedback during repeated-exposure will learn to optimize their decision-outcomes, establish the decision-choice—decision-outcome link and simultaneously develop a high- utility memory-instances/SDU following repeated-exposure to the DDM- environment.

ii.) In a similar DDM-environment the decision-maker will move from learning- based decision-rule to memory-based decision-processes (recognition, judgement, choice and execution) wherein even without feedback they will continue to undertake high-utility decision-outcomes as their initial exposure to feedback in a previous similar DDM-environment has helped them learn the optimal policy to produce high-utility decision-outcomes.

HBDR and IBM Decision-Process

iii.) Decision-makers subjected to DDM-environments providing no-feedback during repeated-exposure will engage in HBDR (set a sub-optimal threshold and arrive at a decision) and will produce low-utility decision-outcomes in spite of repeated-exposure to DDM-environment.

iv.) Such decision-makers when subjected to similar DDM-environments post engaging in HBDR will continue to produce low-utility decision-outcomes as lack of feedback initially

had impaired their ability to form memory-instances in the first place, which they could have relied on to make decisions in similar DDM-environments.

Present Study—Testing RLIIBM in a novel DDM-environment

The present study attempts to test the proposed hybrid-model—RLIIBM in a novel DDM-environment which focuses on understanding how decision-makers balance their discretionary- expenditures (DE) along with the mandatory-expenditures (ME) incurred in everyday life across different financial periods of life such as employment/income-phase and retirement- phase. Research investigating human financial expenditure decision-making so far has only been undertaken by behavioural economics researchers who explain expenditure behaviour throughout different life-cycles such as employment and retirement-phase using economic theoretical-frameworks such as the life-cycle hypothesis (LCH; Ando & Modigliani, 1963). Researchers so far have accounted for the impact of gender (Schubert et.al., 1999) level of financial-literacy (Mandell, 2009), age (Chen & Sun, 2011), differing levels of specific cognitive-abilities like numeracy (Agarwal & Mazumder, 2013), etc. on financial expenditure decision-making in DDM-environments. However, the role of cognitive-processes such as learning and memory on expenditure decisions in dynamic-settings, have not been investigated yet.

On the other hand, DDM research pertaining to human financial decisions in DDM- settings undertaken so far, has only focused on understanding human performance in stock market investment (Gonzalez et.al., 2005), financial allocation and resource management in simulated business-firms (Donovan et.al., 2015) using IBLT/IBLM (Gonzalez, 2004). The decision-rules and decision-processes decision-makers employ in financial expenditure decision-making in dynamic-settings with differing income-periods (e.g., employment and retirement) has not been investigated yet. Thus, testing our new hybrid-model in this novel DDM-setting will help us understand the role of cognitive-processes such as learning and memory in human financial expenditure decision-making in dynamic-settings with varied income-flows.

RLIIBM predictions in the Financial Discretionary-Expenditure DDM-environment

According to the economic theoretical-framework of LCH of Expenditure behaviour, high-utility and optimal expenditure decisions focus on undertaking lower DE's as compared to ME's, specifically during income-phase inorder to maximize their savings for the retirement- phase when they have no-income (Shefrin & Thaler, 1988). Therefore, in the context of RLIIBM, undertaking low DE's inorder to maximize savings for retirement is a high-utility decision-outcome. Thus, according to RLIIBM it is predicted that DDM-environments which provide decision-makers with feedback pertaining to DE's and their overall savings, will help decision-makers optimize their decision-choices, thereby generating increasingly high-utility optimal decision-outcomes resulting in undertaking overall low-DE. It is further predicted that these high-utility decision-choice—decision-outcome memory-instances/SDU pertaining to undertaking low-DE will be simultaenously formed and stored in the decision-makers memory (Stage1: RLDR). Further, in a similar DDM-environment, even without explicit feedback updates about their DE's and savings-status, it is predicted that participants will move from a learning-based decision-rule to a memory-based decision-process (recognition, judgement, choice and execution) state wherein decision-makers will utilize the memory-instances/SDU's of these high-utility

decision-outcome and utilize it in a similar DDM-environment and are thus predicted to continue undertaking lower DE decisions (Stage2: IBM Decision-Processes).
On the contrary, it is predicted that decision-makers subjected to DDM-environments with no-feedback updates pertaining to DE's and savings-status, will engage in undertaking HBDR to undertake DE decisions. The lack of feedback on their performance will prohibit decision- makers to monitor their performance thereby leading to generation of low-utility decision- outcomes and thus result in undertaking high DE. Finally, it is predicted that lack of feedback initially, will impair the decision-makers ability to form high-utility decision-outcome memory-instances/SDU's to be utilized in a similar DDM-environment and thus decision-makers will continue to perform worse and continue to generate low-utility outcomes by undertaking even higher DE.

METHODOLOGY

Participants

Forty-eight University of Sheffield students were randomly selected to participate in the current study (Females=36 and Males=12, Mean Age=22.16 years, SD=5.69, Range=18-51). As the current study required participants to perform a computer task, the only inclusion criteria of the study were to recruit participants with normal or corrected to normal vision.

Design

A mixed-measures experimental design was used in the current study with one between-subjects independent variable (IV) and one within-subjects IV's, each of which had two levels. The between-subjects IV was Condition with two levels—a.) Feedback and b.) No-feedback and the within-subjects IV was Decision-Making Sessions (DMS) with two levels—a.) DMS- 1 and b.) DMS-2. All participants were randomly allocated to either the feedback or no- feedback condition. Each participant performed in both the DMS-1 and DMS-2 of the condition (feedback and no-feedback) they were randomly allocated to. The dependent variable in the present study was the amount of Discretionary-Expenditure (DE) undertaken by the participants. A 2X2 mixed-measure ANOVA was performed to assess the effect of Condition (feedback and no-feedback) and the impact of DMS (DMS-1 and DMS-2) on the DE decisions undertaken. Analysis of the data was carried out using SPSS version 25.

Materials

Stimulus

The Investment and Savings Task (IST)

A Financial Decision-Making task called the Investment and Savings Task (IST) was created and programmed for the current study in PsychoPy v1.82.01 (Peirce, 2007;2009;2018). This task was administered on HP 15-inch laptop.

Phases, Life-cycles/ Trials and Salary in IST

The IST consisted of 20 simulated life-cycles/trials, 15 out of which were the Income-Phase and the rest 5 were the Retirement-Phase. The IST would begin with general and task-specific instructions, following which a screen with the Phase (e.g., Income-Phase or Retirement-Phase; see Fig.1) would be displayed on the computer-screen infront of the participant. The IST would start with the Income-Phase, during which the participant would earn a consistent amount of salary worth £10,000 (i.e., from Life-cycle/trial 1-15). After the 15th Life-cycle, the Retirement- Phase would begin from the 16th Life-cycle/trail until the

end of the IST. During this phase the participants received a salary amount of £ 0 which would be displayed on the computer-screen (see Fig.2).

Figure.1: Income-Phase and Retirement-Phase screens in the Investment and Savings Task (IST)

Figure.2: Life-Cycle and Salary screens in the Investment and Savings Task (IST): Life-cycle 1-15 displayed a salary of £10,000 (Income-Phase) and Life-cycle 16-20 displayed a salary of £0 (Retirement-Phase)

There were two main types of expenditures, participants would undertake in IST namely—Mandatory-Expenditure (ME) and Discretionary-Expenditure (DE). The ME comprised of expenditures participants incurred in their everyday life (e.g., bills, transport, rent, heating etc.), while DE comprised of leisure expenditures participants incurred (e.g., vacations, shopping, recreation etc.). Participants had no control on their MEs in the IST. A randomly generated amount which would be displayed on the screen would be deducted from their incurred salary in every life-cycle under ME. However, the participants had to choose how much they wished to spend in DE section by selecting a value of their choice in every life-cycle/trial by adjusting a slider presented on the computer-screen within a range of £1000 (minimum value) to £10,000 (maximum value; see Fig.3). The DE amount was in 1000's and the currency type for the DE was in British Pounds (£).

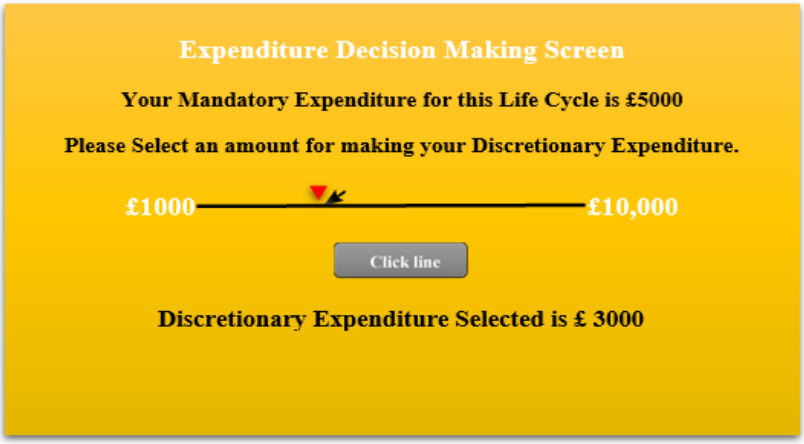


Figure. 3: Mandatory and Discretionary Expenditure Decision-Making screens in the Investment and Savings Task (IST)

Conditions and Decision-Making Sessions in IST

The IST consisted of two main conditions—Feedback-condition and No-Feedback condition. In Feedback-condition participants were presented with a Feedback-screen after undertaking DE in each life-cycle/trial. The Feedback-screen updated the participants with the total amount of money they had in their savings-account post the automatic deduction of the ME and the deduction of the selected DE, provided them with a message about their saving- status (e.g., in debt or not) and finally reminded them about the main goal of the IST which was maximizing savings (see Fig.4). In the No-Feedback condition, participants went on to the next life-cycle without any feedback-screen (see Fig.5).

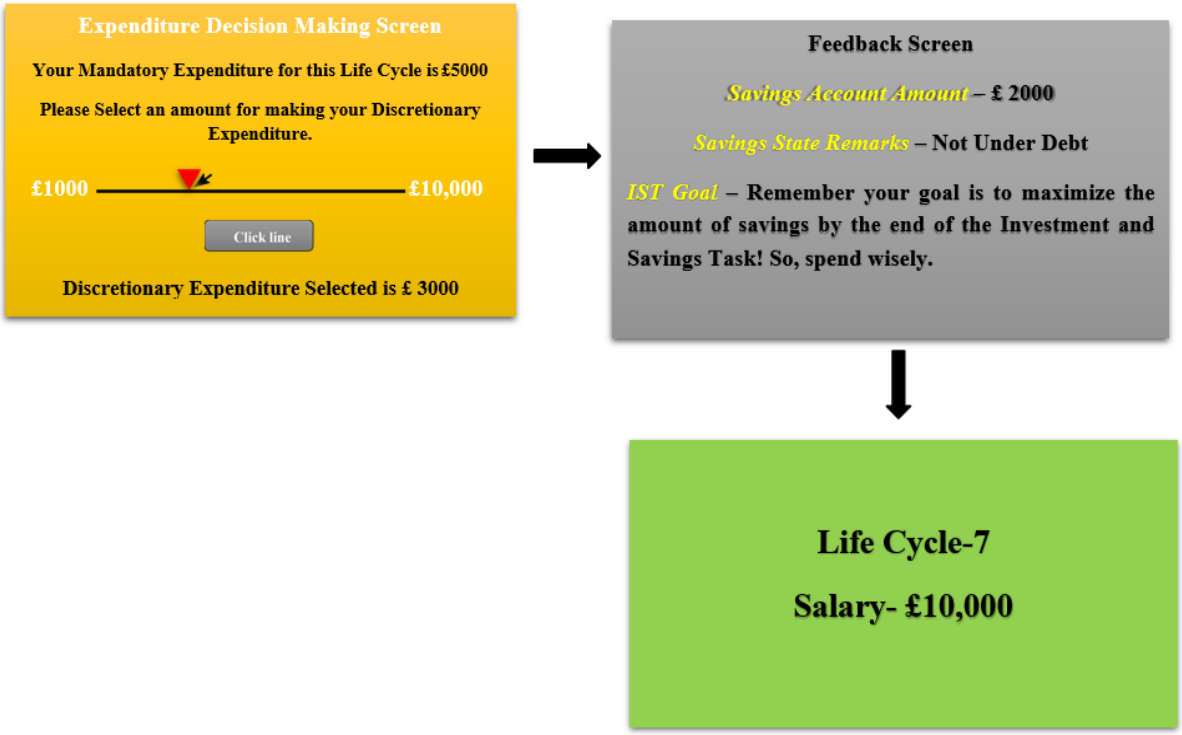


Figure. 4: Conditions Screens in Investment and Savings Task (IST): In the Feedback Condition the participants received the feedback screen after selecting the Discretionary expenditure and then went on to the next Life-cycles/trials in the IST

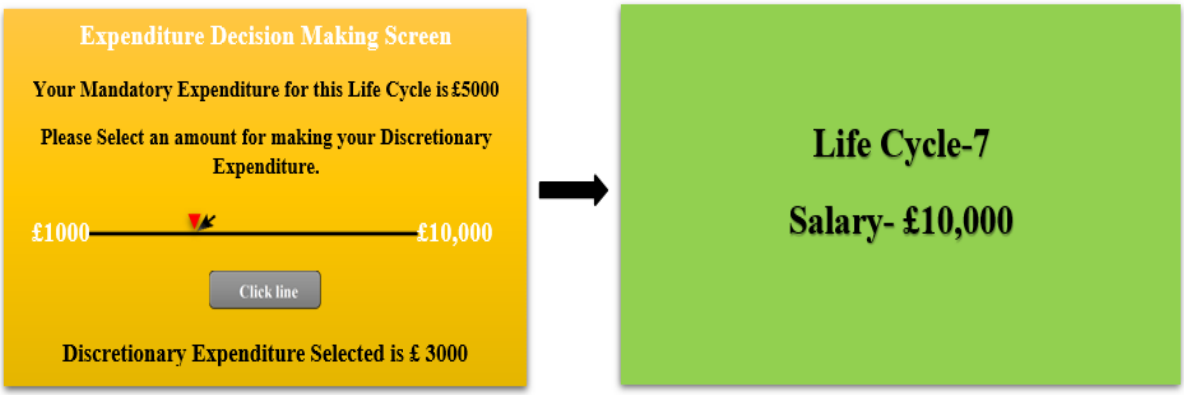


Figure. 5: Conditions Screens in Investment and Savings Task (IST): In the No-Feedback Condition the participants received no feedback screen after selecting the Discretionary expenditure and went on to the next Life-cycles/trials directly in the IST

There were two decision-making sessions (DMS) the participants played in the IST in each condition (Feedback and No-feedback). In the DMS-1 in feedback-condition participants were provided with the feedback-screen after every life-cycle/trial in the IST (see Fig.6). In the DMS-2 in feedback-condition participants were provided with no-feedback screen after every life-cycle/trial in the IST (see Fig.7). In the DMS-1 in no-feedback condition participants were provided with no-feedback screen after every life-cycle/trial in the IST (see Fig.8). In the DMS- 2 in no-feedback condition participants were again provided with no-feedback screen after every life-cycle/trial in the IST (see Fig.9).

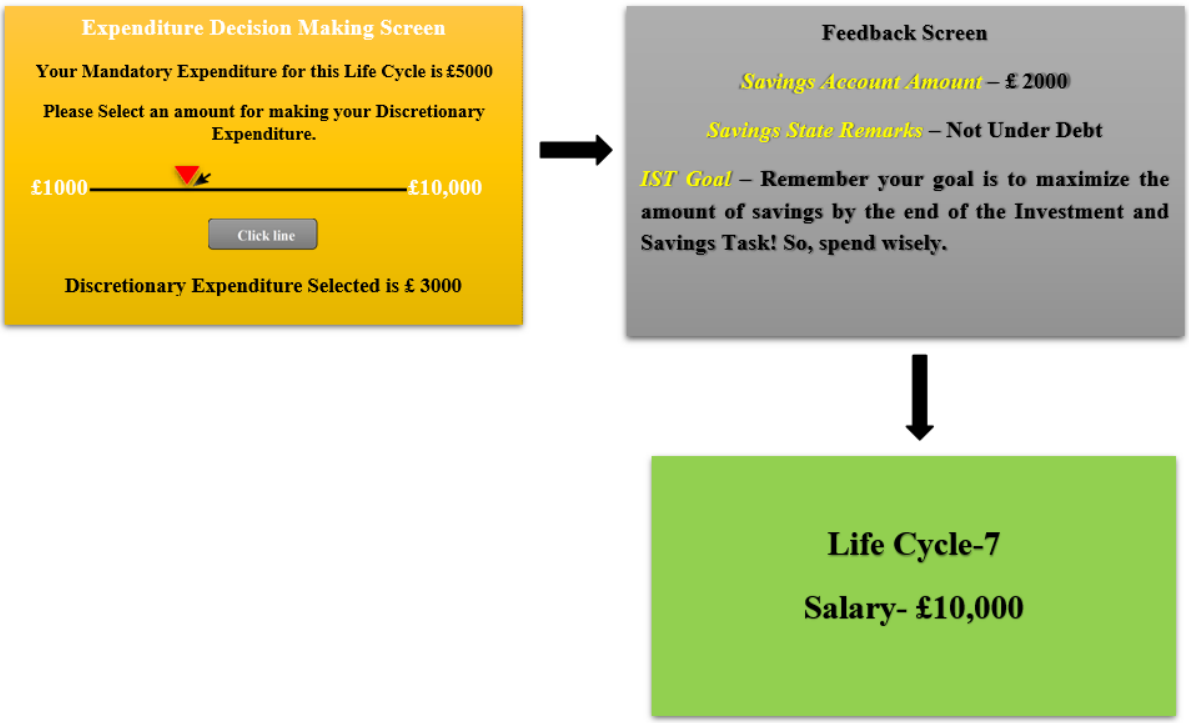


Figure. 6: Decision Making Session Screens in Investment and Savings Task (IST): In the Decision-Making Session (DMS) 1 in Feedback Condition the participants received a feedback screen after selecting the Discretionary expenditure and then went on to the next Life-cycles/trials in the IST for all 20 life-cycles/trials of DMS-1

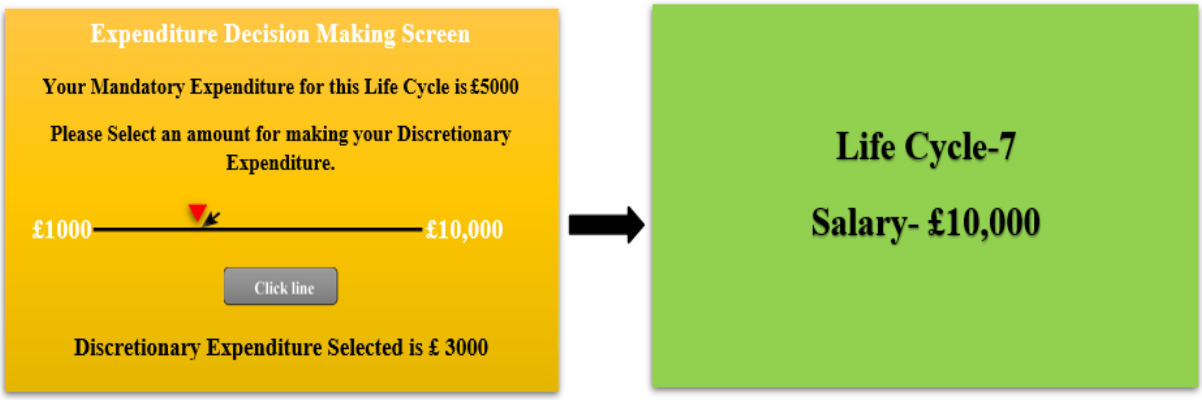


Figure. 7: Decision Making Session Screens in Investment and Savings Task (IST): In the Decision-Making Session 2 in Feedback Condition the participants received no feedback screen after selecting the Discretionary expenditure and then went on to the next Life-cycles/trials in the IST for all 20 life-cycles/trials of DMS-2

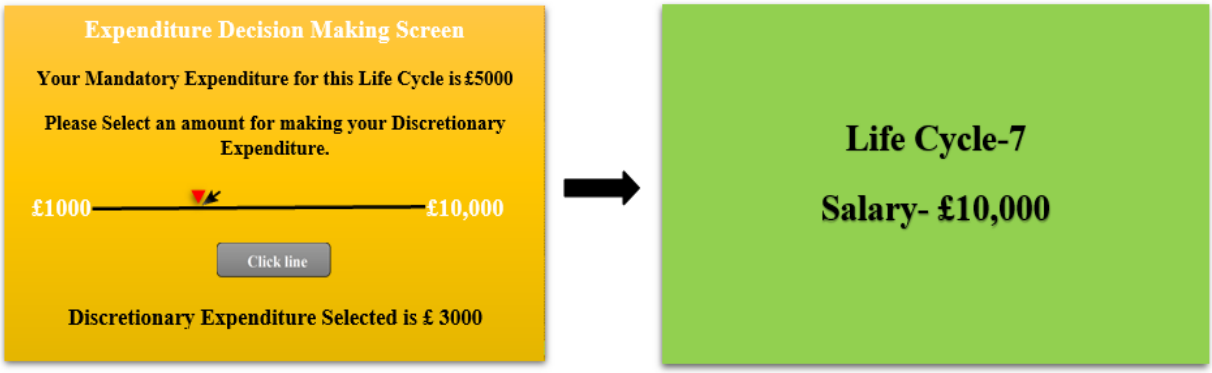


Figure. 8: Decision Making Session Screens in Investment and Savings Task (IST): In the Decision-Making Session 1 in No-Feedback Condition the participants received no feedback screen after selecting the Discretionary expenditure and then went on to the next Life-cycles/trials in the IST for all 20 life-cycles/trials of DMS-1

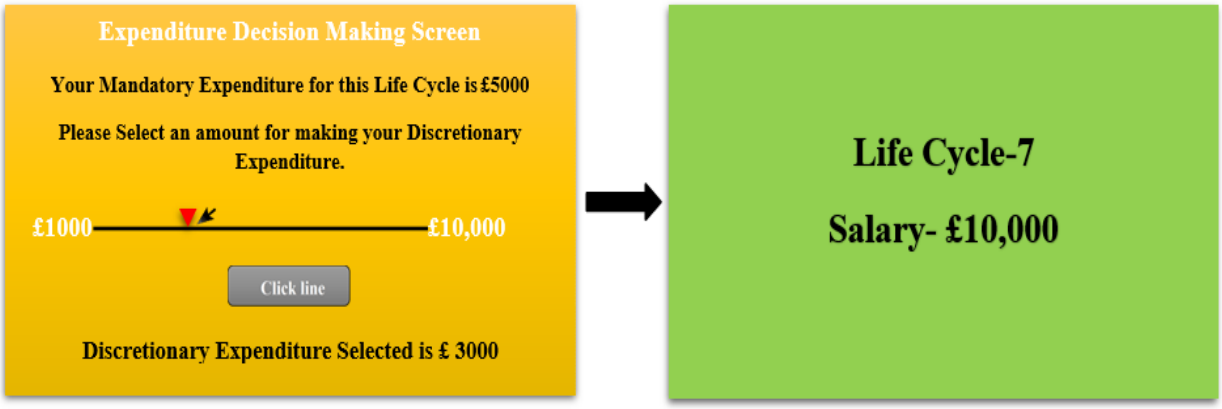


Figure. 9: Decision Making Session Screens in Investment and Savings Task (IST): In the Decision Making Session 2 in No-Feedback Condition the participants received no feedback screen after selecting the Discretionary expenditure and then went on to the next Life-cycles/trials in the IST for all 20 life-cycles/trials of DMS-2

Temporal elements in IST

The instruction-screens and the expenditure decision-making screens were self-paced, and the participants could skip to the next screen by pressing the spacebar-key. The screens indicating which phase the participant entered (e.g., Income-phase or Retirement-phase), life-cycle/trail number and the salary amount screen were not self-paced as these would be presented for a total of 2000 ms infront of the participant after which the next screen would automatically appear.

Procedure

The participants were first given the information sheet and a consent form to declare their voluntary participation for the current study. Following providing consent, the participants were randomly allocated to either the Feedback or No-Feedback condition in DMS-1. After finishing performing on all 20 life-cycles/trials in DSM-1 in the allocated condition (feedback or no-feedback), the participants were given a reasonable break of 10 minutes. Following this they began the DSM-2 in the same condition they were assigned to, in DMS-1. After finishing performing on all 20 life-cycles/trials in DSM-2, the participants were given a debrief form and thanked for their participation.

RESULTS

The decision-performance score in the Investment and Savings Task (IST) was calculated trial- wise thus there were 20 scores as the IST had 20 life-cycles/trials per decision-making session in each feedback and no-feedback condition. The average Discretionary-Expenditure (DE) undertaken across each trial for each participant in both Decision-Making Session-1 (DMS-1) and Decision-Making Session-2 (DMS-2) and in both feedback and no-feedback condition was indicative of the decision-performance of participants in IST. The Mean, SD and Range of the average DE across DMS-1 and DMS-2 in both the Feedback and No-Feedback condition are reported in Table 1.

Table 1: Mean, SD and Range of the average discretionary expenditure spent across Decision Making Session 1 (DMS-1) and Decision-Making Session 2 (DMS-2) in the Feedback and No-Feedback Condition. The Discretionary Expenditure amount is in 1000's and the currency type for the Discretionary Expenditure is in British Pounds (£).

Condition	Decision-Making Session 1(DMS-1) (DMS-2)			Decision-Making Session 2		
	Mean	SD	Range	Mean	SD	Range
Feedback	1.71	.309	1.08	1.28	.201	.79
No-feedback	2.80	.394	1.38	2.89	.278	1.08

Inspection of the histograms illustrated that the data was normally distributed. Since all other assumptions were met, the data was analysed using a parametric statistical analysis of Mixed- Measure ANOVA.

A 2X2 mixed-measures ANOVA with two factors was conducted. The first Factor was the Decision-Making Sessions (DMS) factor which had two levels namely Decision-Making Session 1 (DMS-1) and Decision-Making Session 2 (DMS-2). The second factor was called

Conditions and this factor had two levels namely the Feedback condition and the No-Feedback condition.

The results showed that there was a significant main effect of Decision-Making Sessions on the amount of DE undertaken (F(1,38)=9.10, p=.005, ηp2=.193). This shows that 19.3% of the variance in the amount of DE undertaken can be attributed to the Decision-Making Session the participants were in. Overall participants in the DMS-1 of DMS factor undertook higher DE (Mean=2.25) than when in DMS-2 (Mean=2.09; see Fig.1) of DMS.

There was also a significant main effect of the second factor named Condition on the amount of DE undertaken (F (1,38)=295.58, p< .001, ηp2=.886). This shows that 88.6% of the variance in the DE can be attributed to the Condition the participants were in. Overall participants in Feedback-condition of the Condition factor had undertaken lower DE (Mean=1.49) than those in No-feedback condition of the Condition factor (Mean=2.85; see Fig.1).

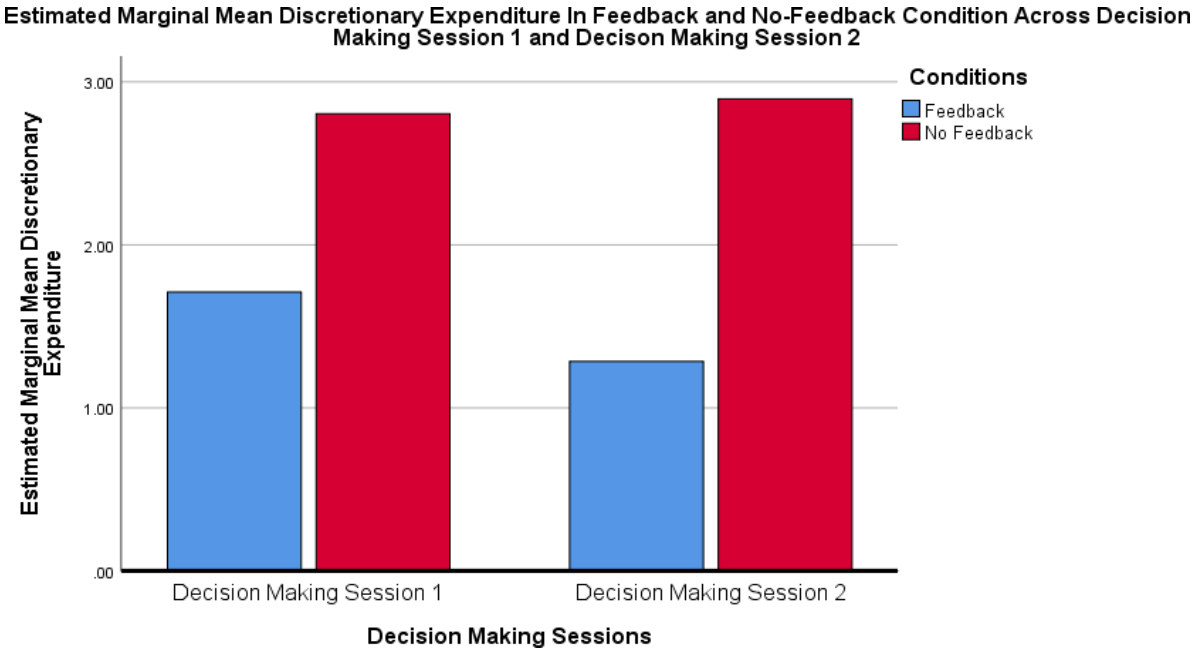


Figure.1: Estimated Marginal Mean Discretionary Expenditure spent across in Decision Making Session 1 (DMS-1) and Decision-Making Session 2 (DMS-2) in the Feedback and No-Feedback Condition. All Discretionary Expenditure values are to be considered in 1000's and the currency type for the Discretionary Expenditure is in British Pounds (£).

In addition, there was a significant interaction between the DMS and the Conditions factors on the amount of DE undertaken (F(1,38=21.88, p< .001, ηp2=.365).

Bonferroni corrected post-hoc t-tests showed that participants in DMS-1 in Feedback Condition undertook significantly lower DE compared to participants in DMS-1 of No-Feedback Condition (t(38)=-9.75, p< .001). Participants in DMS-2 in Feedback Condition also performed significantly better by undertaking significantly lower DE compared to participants in DMS-2 in No-Feedback (t(38)=-20.94, p< .001).

Furthermore, participants in DMS-2 in Feedback Condition performed significantly better than in DMS-1 in Feedback Condition by undertaking significantly lesser DE in the second decision-making session ($t(19)=5.10$, $p< .001$). Although participants in DMS-2 in No-Feedback Condition undertook higher DE as compared to the DE undertaken in their DMS-1 of No-Feedback Condition, this difference was not significant ($t(df)= -1.26$, $p=.222$).

These results suggest that the amount of DE undertaken differed between the Conditions (Feedback and No-feedback) as participants undertook lower DE in feedback-condition as compared to participants in no-feedback condition across both DMS-1&2. The results also suggested that the amount of DE undertaken varied across Decision-Making Sessions (DMS- 1&2) within each condition as participant in DMS-2 in feedback-condition undertook significantly lower-DE as compared to the DE undertaken in DMS-1, while the participants DMS-2 in no-feedback condition undertook higher-DE as compared to the DE they undertook in DMS-1, although this difference was not significant.

DISCUSSION

The aim of the current study was to propose a new hybrid-model called Reinforcement-Learning initiated Instance-Based Model (RLIIBM) of DDM which combines Reinforcement- Learning (RL) based feedback initiated decision-rule and Instance-Based Models (IBM) memory-based decision-processes to address the highlighted gaps in IBLT/IBLM pertaining to the use of Heuristics-Based Decision-Rule (HBDR) and the undermined role of feedback in the decision-processes, and the study further aimed to test this new hybrid-model in a novel financial expenditure DDM-environment. Results of the current study supported the predictions proposed by RLIIBM pertaining to Reinforcement-Learning Decision-Rule (RLDR) and HBDR, in the novel financial DDM- environment with regards to Discretionary-Expenditure (DE) decision-making. As predicted decision-makers subjected to feedback-condition (or DDM-environment) in Investment and Savings Task (IST) undertook significantly lower-levels of DE as compared to decision- makers subjected to the no-feedback condition, in both DMS-1&2. This provided support for the RLDR prediction of RLIIBM which states that providing consistent feedback promotes learning the generation of optimal, high-utility decision-outcome (optimization-policy) over a period of time during the repeated-exposure by establishing the decision-choice—decision- outcome relationship, further supporting the propositions of Reinforcement-Learning Theory as well (RLT; Sutton & Barto, 1998), the theoretical—framework on which RLDR is based. Additionally, this finding was also consistent with past research investigating the effectiveness of feedback in dynamic-contexts, as well (Einhorn & Hogarth, 1978;1981;1986) which stated that DDM-environments providing consistent feedback helps decision-makers to undertake feedback initiated active decision-performance monitoring by identifying any discrepancy between the decision-outcome and task-goal and engaging in corrective measures in successive encounters (i.e., optimize) post the identification of any discrepancy, thereby facilitating learning in the DDM-environments (Kleinmuntz, 1985).

On the other hand, as predicted decision-makers subjected to no-feedback condition (or DDM- environment) in the IST undertook significantly higher DE's as compared to decision-makers in the feedback-condition, in both DMS-1&2. This provided evidence for the HBDR proposition of RLIIBM which stated that in the absence of corrective feedback decision-makers will engage in heuristic-strategies by setting a sub-optimal threshold of decision-choices (satisficing-policy) and generate sub-optimal, low-utility decision-

outcomes despite repeated- exposure. This finding aligned with past DDM studies investigating the impact of heuristic- strategies on decision-performance in DDM-settings (e.g., Kahneman et.al., 1982) which stated that the sub-optimal and poor performance post engaging in heuristic-strategies in the absence of feedback mainly occurs because of two reasons—1.) decision-makers utilizing heuristic-strategies tend to ignore crucial information (e.g., Bar-Hillel, 1980) (usually provided by the corrective feedback) which further leads to cognitive-biases and systematic-errors (Kleinmuntz & Thomas, 1987; Burmeister & Schade, 2007) and 2.) lack of active feedback exacerbates theses above mentioned negative aspects of heuristic-strategies (Hogarth, 1981) as lack of feedback denies decision-makers the opportunity to identify any discrepancy between the decision-choice made and decision-outcome incurred, caused due to the employment of heuristic-strategies (Simon, 1956; Kleinmuntz & Thomas, 1987; Plous, 1993) further disrupting the development of the memory-instances consisting of mental-models of DDM- environment as well as the decision-choice—decision-outcome relationship, which feedback aids to learn over a period of time during repeated-exposure (e.g., Cronin et.al., 2008).

Results of the current study also supported the predictions of RLIIBM pertaining to the RLDR- IBM—memory-based decision-processes in a similar financial DDM-setting. As predicted decision-makers given feedback in DMS-1 performed better in DMS-2 by undertaking significantly lesser DE's as compared to DMS-1, even without explicit feedback. These findings support the propositions of Instance-Based Learning Theory (IBLT) pertaining to the retrieval of memory-instances/SDU's of high-utility decision-outcomes in a similar DDM- environment to guide decision-making (Gonzalez et.al.,2003). These results also support the findings from past studies investigating the impact of feedback on learning, memory and decision-making performance (Estes & Maddox, 1995) and mirrors the findings from the Rhodes and Jacoby (2007) study wherein similar to the current study, post being subjected to feedback in an initial decision-making session, participants outperformed in another similar decision-making session even without being provided with any explicit feedback. Researchers state this mainly happens because decision-makers tend to store successful and high-utility decision-outcomes in their memory as and when they learn about them during the feedback-initiated learning-process, which they further use to guide their decision-making in a similar DDM situations (Klein et.al., 2010).

On the other hand, as predicted decision-makers given no-feedback in DMS-1 performed worse in DMS-2 by undertaking higher DE's as compared to DMS-1, however, this increase in DE's was not statistically significant. Although, the increase in the DE's in DMS-2 was not significantly higher in the no-feedback condition, decision-makers still undertook higher DE's in DMS-2 as compared to DMS-1. These findings corroborate with the results of Rhodes and Jacoby (2007) who demonstrated that providing decision-makers with no-feedback pertaining to their decision-performance deteriorated their performance in a similar decision-making environment. Researchers explain the lack of feedback in dynamic-settings denies decision- makers the opportunity to identify any discrepancy between their decision-choice made and decision-outcome incurred, due to the employment of heuristic-strategies (Kleinmuntz & Thomas, 1987) further disrupting the formation of memory-instances/SDU's consisting high- utility decision-choice—decision-outcome link which they could have used in similar DDM- settings (e.g., Cronin et.al., 2008).

An important question to pose at this stage is, if in the original Water Purification Plant (WPP) study conducted by Gonzalez et.al., (2003) to test the propositions of IBLM, provided participants with no-feedback and participants engaged in heuristic-strategies to generate decision-outcomes, how did the participants show learning over the 3 days? A closer analysis of the WPP task outlay revealed that although the participants were not provided with any explicit feedback on their decision-performance, the WPP task screen constantly updated the participant's water-distribution score after every decision-choice was made, on the top right corner. It is plausible to state that, participants used this water-distribution score update as feedback to assist their learning in WPP over 3 days further strengthening the argument pertaining to the central role of feedback in aiding learning and memory DDM (e.g., Einhorn et.al., 1979).

LIMITATIONS

Although this was the first study to identify the gaps in IBLT/IBLM and address the identified gaps by proposing RLIIBM, two key limitations have been identified in the current model. The first limitation pertains to acknowledging the DDM-characteristic of high temporal-constraints, decision-makers face in certain DDM-settings. The current study tested RLIIBM in a microworld task which overall had moderate-levels of temporal constraints. Previous research shows that despite repeated-exposure coupled with feedback, decision-makers under high- levels of temporal constraints perform poorly as compared to those under moderate or low temporal constraints (Gonzalez, 2004) as high-levels of temporal constraints constricts decision-makers information processing time required for processing the feedback inorder to undertake optimization of further decision-choices and facilitate learning (Greatrex, 2018).

The second limitation identified pertains to utilizing RLIIBM in explaining how learning in real-world DDM-environments which provide feedback at temporal delays or provides inconsistent feedback would take place, as RLIIBM stresses on the importance of consistent feedback during the repeated-exposure to the DDM-environment. As accounted previously, researchers state that lack of consistent feedback after every interaction with the DDM-environment during the repeated-exposure, will impede the decision-makers ability to account for the specific decision-choices that caused them in the first place, thereby causing them to suffer misperception of feedback (Sterman, 1989a,b), further compelling the decision-maker to employ heuristic-strategies to undertake decision-making and inturn resulting in worse decision-performance (Plous, 1993).

FUTURE DIRECTIONS

In order to address both the identified limitations in RLIIBM, we plan to develop a modified version of RLIIBM in a follow-up study in future, wherein we aim to replace the current prescribed Reinforcement-Learning Decision-Rule (RLDR) with a recently developing decision-rule which is a subset of RLDR called the Case-Based Heuristically Accelerated Reinforcement-Learning decision-rule (CB-HARL; Bianchi et.al., 2009). CB-HARL focuses on exploiting the concept of transfer learning which refers to reusing and abstracting a global knowledge-based human develop across multiple domains using RL-based mechanisms over a period of time, into the target domain (i.e., present DDM-environment) (Taylor & Stone, 2009; Canini et.al., 2010). This decision-rule acknowledges that engaging in RLDR can be difficult and less adaptive in temporally constrained DDM-settings and thus to increase the effectiveness of the Reinforcement-Learning (RL) based decision-rules in

such settings, CB- HARL utilizes the characteristic temporally accelerated searching mechanisms of heuristic- strategies to search the global knowledge-base for high-utility decision-choices from the past that have closest overlap with the current DDM-environment (Bianchi et.al., 2015). Along with addressing the limitation pertaining to temporal-constraints, replacing RLDR with CB-HARL will also address the limitation pertaining to the need of constant feedback as well. As CB- HARL exploits an already established global knowledge-base, it does not require the decision- maker to go through the entire process of feedback-initiated learning from the start, as in CB- HARL decision-makers can employ transfer learning instead and utilise the global knowledge- base to facilitate learning in the current DDM-context (e.g., Bianchi et.al., 2008; Bianchi & Mantaras, 2010; Celiberto et.al., 2011; Bianchi et.al., 2015).

CONCLUSION

In conclusion the current study aimed to propose a new hybrid-model of DDM to address the gaps pertaining to the usage of heuristic-based decision-rules (HBDR) and the undermined role of feedback in the decision-processes, highlighted in a past DDM model called the Instance- Based Learning Model. A noteworthy addition to the DDM literature, this is one of the first study to identify the highlighted gaps in IBLT/IBLM and present a new hybrid-model called Reinforcement-Learning initiated Instance-Based Model (RLIIBM), to address the above identified gaps, which builds upon the newly emerging theoretical framework of Reinforcement-Learning in combination with recently growing research acknowledging the significance of feedback in DDM-settings. Furthermore, this was one of the first study to test this new hybrid-model—RLIIBM in a novel financial DDM-environment which focuses on investigating discretionary-expenditure decision-making in dynamic-settings during varied income-flow periods such as during income-phase and retirement-phase.

The current study found empirical evidence supporting the predictions presented by the RLIIBM thereby highlighting the significance of feedback in aiding the learning processes with regards to the refinement of decision-choices over a period of time, to produce high-utility optimal decisions-outcomes, thereby demonstrating that a Reinforcement-Learning based decision-rule and not a HBDR as prescribed in the original IBMT/IBLM, explains how decision-makers gain decision-making proficiency in DDM-settings during their repeated- exposure to DDM-environments. This study further highlighted the significant role feedback-initiated learning plays in forming memories of successful decision-outcomes, which decision- makers can flexibly use in future, similar DDM-environments by referring back to their past memories of successful decision-outcomes to guide their decision-making in these similar environments.

REFERENCES

Agarwal, S., & Mazumder, B. (2013). Cognitive abilities and household financial decision making. American Economic Journal: Applied Economics, 5(1), 193-207.

Ando, A., & Modigliani, F. (1963). The" life cycle" hypothesis of saving: Aggregate implications and tests. The American economic review, 53(1), 55-84.

Anzai, Y. (1984). Cognitive control of real-time event-driven systems. Cognitive Science, 8(3), 221-254.

Arvai, J. L., Campbell, V. E., Baird, A., & Rivers, L. (2004). Teaching students to make better decisions about the environment: Lessons from the decision sciences. The Journal of Environmental Education, 36(1), 33-44.

Astwood, R. S., Van Buskirk, W. L., Cornejo, J. M., & Dalton, J. (2008, September). The impact of different feedback types on decision-making in simulation based training environments. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 52, No. 26, pp. 2062-2066). Sage CA: Los Angeles, CA: SAGE Publications.

Atkins, P. W., Wood, R. E., & Rutgers, P. J. (2002). The effects of feedback format on dynamic decision making. Organizational Behavior and Human Decision Processes, 88(2), 587-604.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. Acta Psychologica, 44(3), 211-233.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. The Quarterly Journal of Experimental Psychology Section A, 36(2), 209-231.

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. British journal of Psychology, 79(2), 251-272.

Bianchi, R. A., & de Mantaras, R. L. (2010, August). Case-Based Multiagent Reinforcement Learning: Cases as Heuristics for Selection of Actions. In ECAI (pp. 355- 360).

Bianchi, R. A., Celiberto Jr, L. A., Santos, P. E., Matsuura, J. P., & de Mantaras, R. L. (2015). Transferring knowledge as heuristics in reinforcement learning: A case-based approach. Artificial Intelligence, 226, 102-121.

Bianchi, R. A., Ribeiro, C. H., & Costa, A. H. (2008). Accelerating autonomous learning by using heuristic selection of actions. Journal of Heuristics, 14(2), 135-168.

Bianchi, R. A., Ros, R., & De Mantaras, R. L. (2009, July). Improving reinforcement learning by using case based heuristics. In International Conference on Case-Based Reasoning (pp. 75-89). Springer, Berlin, Heidelberg.

Brand, M., Laier, C., Pawlikowski, M., & Markowitsch, H. J. (2009). Decision making with and without feedback: the role of intelligence, strategies, executive functions, and cognitive styles. Journal of Clinical and Experimental Neuropsychology, 31(8), 984-998.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. Acta psychologica, 81(3), 211-241.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. Computers in Human Behavior, 9(2-3), 171-184.

Brehmer, B., & Naehlinder, S. (2007). Achieving what cannot be done: Coping with the time constants in a dynamic decision task by doing something else. Scandinavian journal of psychology, 48(5), 359-365.

Brewer, W. F. (1987). Schemas versus mental models in human memory.

Brown, A., Karthaus, C., Rehak, L. A., & Adams, B. (2009). The role of mental models in dynamic decision-making (No. DRDC-CR-2009-060). HUMANSYSTEMS INC GUELPH (ONTARIO).

Burmeister, K., & Schade, C. (2007). Are entrepreneurs' decisions more biased? An experimental investigation of the susceptibility to status quo bias. Journal of Business Venturing, 22(3), 340-362.

Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. The Quarterly Journal of Experimental Psychology: Section A, 55(1), 241-261.

Busemeyer, J. R., & Bruza, P. D. (2012). Quantum models of cognition and decision. Cambridge University Press.

Busemeyer, J. R., & Pleskac, T. J. (2009). Theoretical tools for understanding and aiding dynamic decision making. Journal of Mathematical Psychology, 53(3), 126-138.

Canini, K. R., Shashkov, M. M., & Griffiths, T. L. (2010, June). Modeling Transfer Learning in Human Categorization with the Hierarchical Dirichlet Process. In ICML (pp. 151-158).

Celiberto Jr, L. A., Matsuura, J. P., De Mantaras, R. L., & Bianchi, R. A. (2011, June). Using cases as heuristics in reinforcement learning: a transfer learning application. In Twenty-Second International Joint Conference on Artificial Intelligence.

Chen, S., Sun, Z., Tang, S., & Wu, D. (2011). Government intervention and investment efficiency: Evidence from China. Journal of Corporate Finance, 17(2), 259-271.

Colman, Andrew (2006). A Dictionary of Psychology. New York: Oxford University Press. p. 670.

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. Organizational behavior and Human decision Processes, 108(1), 116-130.

Dawson, N. V., & Arkes, H. R. (1987). Systematic errors in medical decision making. Journal of General Internal Medicine, 2(3), 183-187.

DeNisi, A. S., & Kluger, A. N. (2000). Feedback effectiveness: can 360-degree appraisals be improved? Academy of Management Perspectives, 14(1), 129-139.

Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. Organizational Behavior and Human Decision Processes, 62(2), 198-215.

Dimitrakakis, C., & Ortner, R. (2018). Decision Making Under Uncertainty and Reinforcement Learning.

Dutt, V. (2011). Explaining human behavior in dynamic tasks through reinforcement learning. Journal of Advances in Information Technology, 2(3), 177-188.

Dutt, V. (2011b). Instance-based learning: Integrating sampling and repeated decisions from experience. Psychological review, 118(4), 523.

Ebert, R. J. (1972). Human control of a two-variable decision system. Organizational Behavior and Human Performance, 7(2), 237-264.

Edwards, W. (1961). Behavioural decision theory. Annual review of psychology, 12(1), 473-498.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. Human factors, 4(2), 59-74.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. Psychological review, 85(5), 395.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioural decision theory: Processes of judgement and choice. Annual review of psychology, 32(1), 53-88.

Einhorn, H. J., & Hogarth, R. M. (1986). Decision making under ambiguity. Journal of Business, S225-S250.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. Psychological review, 86(5), 465.

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(5), 1075.

Fagley, N. S., & Miller, P. M. (1990). The effect of framing on choice: Interactions with risk-taking propensity, cognitive style, and sex. Personality and social psychology bulletin, 16(3), 496-510.

Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement.

Funke, J. (1988). Using simulation to study complex problem solving: A review of studies in the FRG. Simulation & Games, 19(3), 277-303.

Funke, J. (1995). Experimental research on complex problem solving. Complex problem solving: The European perspective, 243-268.

G. Barto, "Reinforcement learning in the real world," in Proc. of the 2004 IEEE International Joint Conference on Neural Networks, Amherst, 2004, pp. 1661 vol. 3.

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. Organizational Behavior and Human Decision Processes, 96(2), 142-154.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. Human Factors, 46(3), 449-460.

Gonzalez, C. (2014). Dynamic Decision Making: Implications for Recommender System Design. In DMRS (pp. 25-31).

Gonzalez, C., & Dutt, V. (2011a). A generic dynamic control task for behavioural research and education. Computers in Human Behavior, 27(5), 1904-1914.

Gonzalez, C., & Dutt, V. (2011b). Instance-based learning: Integrating sampling and repeated decisions from experience. Psychological review, 118(4), 523.

Gonzalez, C., & Quesada, J. (2003). Learning in dynamic decision making: The recognition process. Computational & Mathematical Organization Theory, 9(4), 287-304.

Gonzalez, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic decision making: Learning processes and new research directions. Human factors, 59(5), 713-721.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. Cognitive Science, 27(4), 591-635.

Gonzalez, C., Qudrat, H., & Lerch, F. J. Decision making in a dynamic, real-time environment: Learning under high cognitive workload. Manuscript submitted to Human Factors for publication.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. Computers in human behavior, 21(2), 273-286.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. Cognitive Science Quarterly, 2(2), 205-227.

Greatrex, D. C. (2018). Effects of Temporal Expectation on Complex Decision Making (Doctoral dissertation, University of Cambridge).

Guess, C. D., Donovan, S. J., & Naslund, D. (2015). Improving dynamic decision making through training and self-reflection.

Gureckis, T. M., & Love, B. C. (2009a). Learning in noise: Dynamic decision-making in a variable environment. Journal of Mathematical Psychology, 53(3), 180-193.

Gureckis, T. M., & Love, B. C. (2009b). Short-term gains, long-term pains: How cues about state aid learning in dynamic environments. Cognition, 113(3), 293-313.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. Psychological Bulletin, 90(2), 197.

Hotaling, J. M., Fakhari, P., & Busemeyer, J. R. (2015). Dynamic decision making. International encyclopaedia of the social & behavioural sciences, 709-714.

Hsiao, N., & Richardson, G. P. (1999, July). In search of theories of dynamic decision making: a literature review. In 17th International Conference of the System Dynamics Society, Wellington, New Zealand. Retrieved February (Vol. 19, p. 2012).

Jagacinski, R. J., & Miller, R. A. (1978). Describing the human operator's internal model of a dynamic system. Human Factors, 20(4), 425-433.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. American psychologist, 64(6), 515.

Kahneman, D., & Tversky, A. (1981). The simulation heuristic(No. TR-5). STANFORD UNIV CA DEPT OF PSYCHOLOGY.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (Eds.). (1982). Judgment under uncertainty: Heuristics and biases. Cambridge university press.

Kerstholt, J. (1994). The effect of time pressure on decision-making behaviour in a dynamic task environment. Acta Psychologica, 86(1), 89-104.

Kerstholt, J. (1996). The effect of information costs on strategy selection in dynamic tasks. Acta Psychologica, 94(3), 273-290.

Kerstholt, J. H., & Raaijmakers, J. G. (1997). Decision making in dynamic task environments. Decision making: Cognitive models and explanations, 205-217.

Klein, G. (2015). A naturalistic decision making perspective on studying intuitive decision making. Journal of Applied Research in Memory and Cognition, 4(3), 164-168.

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. Journal of Cognitive Engineering and Decision Making, 4(3), 186-209.

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. Management Science, 31(6), 680-702.

Kleinmuntz, D. N., & Thomas, J. B. (1987). The value of action and inference in dynamic decision making. Organizational Behavior and Human Decision Processes, 39(3), 341- 364.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological bulletin, 119(2), 254.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. Current directions in psychological science, 7(3), 67-72.

Mandell, L., & Hanson, K. O. (2009, January). The impact of financial education in high school and college on financial literacy and subsequent financial decision making. In American Economic Association Meetings (Vol. 4, p. 2009).

Manktelow, Ken (2000). Reasoning and Thinking. Hove: Psychology Press. p. 221.

March, J. G., & Simon, H. A. (1958). Cognitive limits on rationality. Organizations, 136-171.

Martin, M. K., Gonzalez, C., & Lebiere, C. (2004). Learning to make decisions in dynamic environments: ACT-R plays the beer game.

Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., & O'Connor, P. J. (2014). A machine learning approach to improving dynamic decision making. Information Systems Research, 25(2), 239-263.

Nicholls, N. (1999). Cognitive illusions, heuristics, and climate prediction. Bulletin of the American Meteorological Society, 80(7), 1385-1398.

Osman, M., Wilkinson, L., Beigi, M., Castaneda, C. S., & Jahanshahi, M. (2008). Patients with Parkinson's disease learn to control complex systems via procedural as well as non- procedural learning. Neuropsychologia, 46(9), 2355-2363.

Otto, A. R., & Love, B. C. (2010). You don't want to know what you're missing: When information about forgone rewards impedes dynamic decision making. Judgment and Decision Making, 5(1), 1.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The adaptive decision maker. Cambridge university press.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. Journal of Neuroscience Methods, 162 (1-2):8-13

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. Frontiers in neuroinformatics, 2, 10.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2018). PsychoPy2: Experiments in behavior made easy. Behavior research methods, 1-9.

Plous, S. (1993). The psychology of judgment and decision making. Mcgraw-Hill Book Company.

Ram, A. (1993). Indexing, elaboration and refinement: Incremental learning of explanatory cases. In Case-Based Learning (pp. 7-54). Springer, Boston, MA.

Rapoport, A. (1966). A study of human control in a stochastic multistage decision task. Behavioral Science, 11(1), 18-32.

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(2), 305.

Richardson, R. C. (2017). Heuristics and satisficing. A companion to cognitive science, 566-575.

Riesbeck, C. K., & Schank, R. C. (1989). Inside case-based reasoning. Hillsdale, NJ: Erlbaum (Lawrence).

Schubert, R., Brown, M., Gysler, M., & Brachinger, H. W. (1999). Financial decision-making: are women really more risk-averse? American Economic Review, 89(2), 381-385.

Sengupta, K., & Abdel-Hamid, T. K. (1993). Alternative conceptions of feedback in dynamic decision environments: an experimental investigation. Management Science, 39(4), 411-428.

Shefrin, H. M., & Thaler, R. H. (1988). The behavioural life-cycle hypothesis. Economic inquiry, 26(4), 609-643.

Simon, D. A., & Daw, N. D. (2011). Environmental statistics and the trade-off between model-based and TD learning in humans. In Advances in neural information processing systems (pp. 127-135).

Simon, H. A. (1955). A behavioural model of rational choice. The quarterly journal of economics, 69(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. Psychological review, 63(2), 129.

Simon, H. A. (1957). Models of man; social and rational.

Skinner, B. F. (1939). The alliteration in Shakespeare's sonnets: a study in literary behavior. The Psychological Record.

Sterman, J. D. (1989a). Modelling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. Management science, 35(3), 321-339.

Sterman, J. D. (1989b). Misperceptions of feedback in dynamic decision making. Organizational Behavior and Human Decision Processes, 43, 301–335.

Sterman, J. D. (1994). Learning in and about complex systems. System Dynamics Review, 10(2-3), 291-330.

Sutton, R. S., & Barto, A. G. (1998). Introduction to reinforcement learning (Vol. 135). Cambridge: MIT press.

Taylor, M. E., & Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. Journal of Machine Learning Research, 10(Jul), 1633-1685.

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision

Acknowledgements

The author appreciates all those who participated in the study and helped to facilitate the research process.

Conflict of Interest

The author declared no conflict of interest.

How to cite this article: Banerjee N (2020). "What is the significance of feedback and reinforcement learning mechanisms in aiding dynamic decision making?" — proposing a revised hybrid reinforcement learning initiated instance-based model of dynamic decision making. *International Journal of Indian Psychology*, *8*(3), 540-569. DIP:18.01.064/20200803, DOI:10.25215/0803.064