JDDM
Journal of Dynamic Decision Making

Open Access
Original Research

Some Evidence for the Effectiveness of a Brief Error Management Training in Complex, Dynamic, and Uncertain Situations

Joanna Hermida and C. Dominik Güss

Department of Psychology, University of North Florida, Jacksonville, USA

Previous research highlights the need for developing techniques to improve decision making in uncertain situations. The current study explores the effects of a brief training program on complex problem solving (CPS) and dynamic decision making (DDM) performance in two computer-simulated tasks with different task characteristics, ChocoFine ($N = 76$) and WinFire ($N = 99$). Half of the participants in each simulation group received brief training on 16 frequent CPS and DDM errors. We hypothesized that participants who received training on potential errors would show better performance, report fewer errors, and show fewer behavioral errors compared to participants who did not receive error training prior to the start of the simulated tasks. The results showed that participants in both error training groups had better performance scores. Participants who received training had fewer self-reported errors compared to the no-training group overall, and in the ChocoFine simulation, but not the WinFire simulation. Regarding behavioral errors, status-quo bias was related to weaker performance in both simulations. These findings have implications for leaders prone to the status-quo bias and for organizations that could implement training programs for DDM and CPS.

Keywords: dynamic decision making, complex problem solving, human error, microworlds, self-reflection, status-quo bias

Making successful decisions is a key challenge in many domains of life – for instance, in business and politics, domains that are unpredictable, complex, and ever-changing. It is well established that faulty decision making and human error in these domains can lead to catastrophic consequences (e.g., Reason, 1990; 2000). However, "improving decision making and learning in dynamic environments presents a major challenge to researchers" (Karakul & Qudrat-Ullah, 2008, p. 20). One way to study faulty decision making and human error is the use of microworlds, i.e., computer-simulated dynamic problem-solving tasks. When working on these tasks, participants make decisions, including faulty decisions, and learn from their mistakes. In these microworld experiments, all participants' decisions and systems data are saved in log files, allowing a controlled approach to study of decision making. It is also possible to systematically study the effect of error-management training (EMT) in order to improve decision making and reduce errors. If proven successful, such decision-making and EMT could then be implemented, adapted, and tested in various real-life domains. Ultimately, such training could help to save millions of dollars for businesses and organizations, potentially prevent more severe outcomes during natural disasters for emergency planners and government officials, and save lives when politicians make health-policy related decisions.

Some findings of experimental studies on dynamic decision making (DDM) and error analysis have been applied to specific life domains: For example, the field of police decision making (Harman et al., 2019), studying police officers' decision to shoot or not in a given instance, and how cognitive and affective processing influenced these decisions and decision errors. In another example in the field of healthcare management, DDM research has demonstrated improved treatment strategies and reduced treatment failure of patients with type 2 diabetes mellitus (Meyer et al., 2014).

DDM involves making decisions in complex, uncertain, and dynamic environments (Dörner, 1996). Complexity refers to many interconnected variables. Uncertainty refers to lacking information and unpredictable developments. Dynamics refers to the ever-changing nature of these environments. DDM can be defined as the process of overcoming obstacles between a current state and a desired goal state via a multistep process involving an individual's cognitive, emotional, and social abilities in a novel and dynamic environment (Dörner & Funke, 2017; Frensch & Funke, 1995; Güss et al., 2015). Other definitions explain DDM as a series of interdependent decisions in novel and ever-changing environments (Brehmer, 1996; Fischer et al., 2012; Gonzalez et al., 2005). In such environments, the outcome is not only dependent on the individual's decisions, but also on the changes occurring in various aspects of the dynamic environments (Brehmer, 1996; Fischer et al., 2012).

An ongoing challenge for researchers is to uncover the underlying factors that affect performance in DDM tasks. Some of these factors are cognitive biases and errors (Kahnemann & Renshon, 2009), as laboratory

Corresponding author: C. Dominik Güss, ORCiD: 0000-0002-2289-1794, Department of Psychology, University of North Florida, Jacksonville, United States of America. e-mail: dguess@unf.edu

and field studies have shown (e.g., Dhami et al., 2019; Samuelson & Zeckhauser, 1988). Examples include inaccurate perceptions of the problem, poorly defined goals, confirmation bias, among many others (Dörner, 1996; Dörner & Güss, 2022; Güss & Dörner, 2011; Kahneman & Renshon, 2009).

The goal of this study is to further investigate how individuals can improve their DDM skills and thus increase their performance. To this end, we conducted a training program on cognitive biases and errors regarding all the steps of the DDM process. The main goal was to investigate whether, after providing such an error list, human errors decreased, and performance improved in two different DDM and CPS tasks.

Whereas many studies have focused on isolated cognitive biases (e.g., anchoring, Bahník et al., 2019; desirability bias, Russo & Corbin, 2016), the current study focuses on cognitive biases and errors in relation to *all* the decision-making and action steps required in complex, dynamic, and uncertain environments. The requirements of these environments are different from the static requirements of simple choice tasks. It is not possible to focus on only one aspect of a situation, for example, on defining goals or on gathering information. In these environments, the decision maker must deal with *all* of the steps at the same time.

These steps are - although different researchers sometimes use different terminology: (1) problem identification; (2) goal definition; (3) information gathering; (4) elaboration and prediction; (5) planning, decision-making, and action; and (6) outcome evaluation and self-reflection (Edwards, 1962; Güss & Dörner, 2011; Keleş & Yazgan, 2022; Klein, 1999; Sternberg, 1986). When discussing each step, we will also refer to some cognitive biases and errors (for a complete list of the errors identified and analyzed in the current studies see Appendix A).

(1) *Problem identification*: Realizing that a problem exists is the prerequisite for all other steps and one error is simply denying it (e.g., an alcoholic not realizing or accepting having an addiction problem). If one realizes that a problem exists (see also situational awareness, Nicholson & O'Hare, 2014), then the task is to define the key elements/aspects of the problem as one can easily get lost in irrelevant aspects of a big problem (Dörner, 1996). How a problem is identified and represented affects the further decision-making process (e.g., Billings & Hermann, 1998).

(2) *Goal definition*: A second demand is to identify problem-solving goals. Since the main goals are often vague (e.g., make profit; extinguish forest fires), it is imperative to develop sub-goals that help accomplish the main goal (e.g., increasing market presence and launching new products for a company to make profit; clear forest to avoid further spreading of fires). An error would be not to define goals. Through the process of identifying a problem and creating goals to reach a possible solution, decision-makers can begin to learn their strengths and weaknesses and adjust their sub-goals as they see fit (Grant et al., 2002; Sanders & McKeown, 2008).

(3) *Information gathering*: Decision makers, with their problems and goals in mind, are then faced with the task of gathering additional information that is relevant to their established goals. Gathering further information allows decision makers to determine whether causal relationships change over time and how these changes occur (Ramnarayan et al., 1997). For example, top managers must gather information about the market, target clients, potential competitors, and so on. Fire fighters must gather information about wind direction and strength when fighting forest fires. Dynamic situations are ever-changing, and the acquisition of new information remains a constant task throughout all decision-making stages (Osman & Palencia, 2019). One error related to information gathering is to not analyze the causes of a problem and to not forecast possible consequences (Dörner, 1996; Stenmark et al., 2010). Another error is entrenchment. Individuals who experience entrenchment may spend too much time gathering all kinds of information, especially related to irrelevant aspects of the problem.

(4) *Elaboration and prediction*: In the elaboration and prediction stage, decision makers may begin inferring aspects of their environment and how certain variables interact (Brehmer & Dörner, 1993; Güss et al., 2015). Keeping in mind that simple heuristics may not lead to optimal results in novel and unstable environments, decision makers begin to realize that previously successful protocols are not suitable and as a result, become aware of their limitations in understanding the problem situation (Dodson & Schacter, 2002). For example, decision makers may fail to consider time developments (Güss & Dörner, 2011) and may neglect to weigh their options in terms of possible long-term consequences of their actions (Lafond et al., 2012). Choice deferment in decision making provides additional explanation for why a decision maker may spend a large amount of time in this stage without moving forward (Chernev et al., 2015; Tversky & Shafir, 1992).

(5) *Planning, decision-making, and action*: Complex, dynamic, and non-transparent situations offer a multitude of possible choices (see also choice overload, e.g., Chernev et al., 2015). One must develop a promising plan that then can be implemented and lead to success. Research has shown that when the number of choices increase, the likelihood to defer making a choice increases (e.g., Carroll et al., 2011). Also, the more important a task is, the more likely it is that poeple defer making a choice (Krijnen et al., 2015). Sometimes the plan is to do nothing or stick with what one has

done before. This error has been called status-quo bias (e.g., Samuelson & Zeckhauser, 1988). Leaving a situation and confronting uncertainty is more difficult and disadvantageous than not doing anything or sticking with what one has decided or done before (see also Kahneman et al., 1991). The military strategist von Clausewitz (1873) named this sticking to old rituals "methodism" (*Methodismus*; see Dörner & Meck 2022), making decisions based on experience and selecting "an always recurring proceeding out of several possible ones" (von Clausewitz, 1873, p. 63).

(6) *Outcome evaluation and self-reflection*: As decision makers encounter faults in their understanding of the problem or see the negative consequences of some of their actions, they must then begin to revise their strategic approach or formulate new strategies. Not engaging in monitoring and self-reflection is an error. Self-reflective strategies refer to an evaluation of one's thoughts, feelings, and behaviors (Grant et al., 2002). Research has shown that individuals who engage in self-reflection about their decision-making performed better than participants who engaged in less self-reflection (e.g., Donovan et al., 2015). Individuals who self-reflect more than others have a more accurate mental representation of their progress and better strategic control in pursuit of their sub-goals and main goals (Donovan et al., 2015; Locke & Latham, 2002; Osman, 2010). The use of self-reflection in DDM tasks allows individuals to relate new information to past knowledge and assists in the understanding of new ideas (Sanders & McKeown, 2008).

Microworlds and DDM

How can DDM be studied in complex, uncertain, and dynamic tasks? One frequently used method is microworlds. Microworlds are computer-simulated problem situations, whose purpose is to immerse the participant in a specific situation, for example the simulation of a company. Participants are expected to formulate and initiate decisions that in turn change the simulated environment, essentially creating an endless cycle of cause and effect influences between the participants' decisions and the targeted problem environment (Funke, 2001; Gonzalez et al., 2017). A microworld consists of a complex, uncertain, and dynamic problem environment in which performance is assessed by automatically recording and saving each decision the participant makes, along with varying changes in the system (Güss et al. 2015).

The use of microworld simulations to analyze DDM performance allows researchers to develop complex theories involving human thought and behavior (Dörner, 1999; Dörner et al., 1999; Dörner & Güss, 2013). Using microworlds also allows researchers to identify which strategies in each DDM task will be more likely to lead to success or failure (Güss et al., 2015; Schoppek & Fischer, 2017). A major advancement in recent years in DDM research and microworlds has been to simplify the environments while maintaining the integrity of dynamic complexity (Funke & Greiff, 2017; Gonzalez et al., 2017; Greiff et al., 2015) yet researchers have criticized the simplicity of some of those microworlds (e.g., Funke et al., 2017; Güss et al., 2017). To address the possible influence of task characteristics, we chose two different microworlds for the current studies: ChocoFine (Dörner, 2000) and WinFire (Gerdes et al. 1993; Schaub, 2019). ChocoFine can be described as highly complex (over 1,000 simulated variables), highly uncertain due to the multitude of variables and their interactions, and low in dynamics as changes are shown only when proceeding to the next month. WinFire can be described as moderate in complexity (over 30 simulated variables), moderate in uncertainty because locations where fires start are unknown, and high in dynamics as fires spread quickly. More details about the simulations are provided in sections Study 1 and Study 2.

Human Error and DDM Training

Past research has shown the value of active exploration and error management training (EMT) that aids in learning and performance (Keith & Frese, 2008); however, previous studies utilized often simple tasks in their methodology, rather than designs that feature complex and dynamic characteristics, i.e., microworld simulations (Heimbeck et al., 2003). EMT is characterized by both error encouragement (e.g., encouraging mistakes as a part of learning) and active exploration (e.g., trainees initiate, direct, and regulate their own learning while training) (Keith & Frese, 2008). Heimbeck and colleagues (2003) advocate for a positive attitude towards error training along with a guided approach in the beginning of the learning process. Positive attitudes towards errors yield long-lasting learning compared to error avoidance attitudes. According to Heimbeck and colleagues (2003), the addition of error training and guided behaviors revealed higher performance compared to groups that did not receive EMT.

In recent years, researchers have identified that exposure to errors in training can improve the ability to detect them and to manage any stress associated with goal acquisition (Damm et al., 2011; Loh et al., 2013). Subjects who particpate in EMT show an increase in self-regulatory behavior involving two components: (1) Deploying emotional control to reduce adverse emotional reactions to errors (Güss & Starker, 2023), and (2) active engagement in metacognitive activities (i.e., planning, monitoring, and self-reflection). Metacognitive activities increase with EMT because the decision maker is forced to think through their errors and therefore consider the causes of those errors (Ivancic & Hekseth, 2000).

As previously noted, individuals who progress through the DDM process (e.g., from problem identification to self-reflection) may experience cognitive biases and human errors (Dörner, 1996; Güss et al., 2015; Ramnarayan et al., 1997). Research has suggested that training individuals can diminish the influences of these common human errors and biases (Hedge & Kavanagh, 1988; Gully et al., 2002; Donovan et al., 2015). In one study, for example, 110 students were trained for up to 7 hours to work with 5 different computer-based complex problem-solving situations (Kretzschmar & Süss, 2015). Then their performance was assessed in a sixth simulation. The training group participants compared to the control group participants were significantly better at obtaining knowledge; yet there was no difference between the two groups regarding solving the problem, i.e., knowledge application. One reason for the non-significant findings could be that the training group did not receive feedback, or any instruction related to self-reflection or possible error. Another reason could be that some errors may be apparent in one simulated situation but not in another.

Hypotheses

The goal of the current studies was to demonstrate how error training in regard to the DDM process can aid and facilitate performance in novel and complex environments. In the present study, we investigate the effects of EMT on DDM performance across two distinct microworld simulations, ChocoFine (Study 1) and WinFire (Study 2). We are investigating if a general list of errors in DDM and CPS tasks can be helpful for participants in two tasks with completely different task characteristics and demands. Using two different simulations also adds to the possible generalizability of the research findings. We hypothesized that participants in the training conditions and across both computer simulations will exhibit higher performance compared to participants in the control conditions.

We also made specific predictions regarding the two simulations. In both simulations, the goals are clearly defined, i.e., saving forest and cities in WinFire, and making profit in ChocoFine. Therefore, we do not expect differences between the two simulations regarding self-reported goal definition errors. Due to the high number of variables, ChocoFine is more non-transparent than WinFire. Therefore, we expect more self-reported errors in ChocoFine regarding information collection, regarding elaboration and prediction, and regarding planning, decision-making, and action. We did not make predictions regarding the other two problem-solving steps: Problem identification, and outcome evaluation and self-reflection.

We also predict that participants in the training conditions of both computer simulations will report fewer self-reported errors than participants in the control conditions. Finally, as an indicator of validity, we predict that participants who commit more behavior errors in both computer simulations will display lower performance scores than those participants who commit fewer behavioral errors.

Study 1: ChocoFine

The main goal of Study 1 was to investigate performance differences between training and non-training group in the *ChocoFine* simulation (Dörner, 2000). It is important to view decision making in relation to the task demands as the progression of the DDM process depends largely on the characteristics of the task environment and the objectives. ChocoFine is a computer simulation in which individuals take the role of the CEO position of a chocolate-producing factory. Although ChocoFine is considered highly complex, the simulation is low regarding dynamics meaning that the environment changes only when participants decide to move on to the following month. ChocoFine is medium in time pressure, as participants who complete the simulation can move on to the following month on their own accord – although they are given a specific time to work on a specific number of months, e.g., 1 hour for 12 months. The level of uncertainty experienced in the ChocoFine computer simulation task is very high, considering the plethora of variables participants must utilize (over 700 possible input and information variables); it is also not apparent right away which variables cause either an increase or decrease in profit.

Method

Eighty-three undergraduate students were recruited as participants from the University of North Florida, 19 men, 63 women, and one participant identified as 'other'. Participants' ages ranged from 18 to 40 years old (M = 21.20, SD = 3.67). Regarding ethnicity, 65.1% of participants identified as White, 9.6% as Black, 13.3% as Hispanic, 7.2% as Asian, and the remaining 4.8% as 'Other'. There were 45 participants in the experimental, training group and 38 participants in the control, no-training group. Participants were randomly assigned to one of two conditions. No pattern of relationship was determined between gender and condition, $\chi^2(2) = 1.90$, $p = .39$. There was also no significant age difference between the experimental condition (M = 20.84, SD = 2.80) and the control condition (M = 21.63, SD = 4.49), $t(81) = 0.97$, $p = .33$. A total of five participants (3 in the experimental condition, 2 in the control condition) were identified as extreme outliers as having a total performance score smaller than $-\$1,790,000$ and not included in further analyses. The performance scores, i.e., total capital in month 12, ranged then between $-\$471,908$ to $\$3,086,693$. Finally, data of two participants were not automatically saved properly (only until month 4 and month 9) and could therefore not be included. The two not-saved data sets were in the experimental condition. Sometimes participants click accidentally on "end simulation". Thus, complete data sets of 76 participants were used for the following analyses.

Simulation instructions. In both conditions (training vs. no-training) participants received a typed handout with in-depth instructions highlighting key commands for the ChocoFine simulation, alongside screenshots for easier visual comprehension.

Error training handout. Participants received an error training handout itemizing the six identified DDM steps with each DDM step containing two to five possible errors (e.g., "Elaboration and Prediction: Not considering time developments: We think in the here and now and do not consider time developments and situational changes happening over time"; or "Evaluation of Outcome and Self-Reflection: No monitoring and self-reflection: We think sometimes that if something is going well then it does not deserve further reflection", see Appendix A for complete list of errors). There was a total of 16 possible errors across all DDM stages. The list is based on previous research where researchers identified situations in the simulations MORO, CHOCOFINE, and WINFIRE where errors occurred (Dörner & Güss, 2022). The researchers described 24 errors and explained why specific decision making in these specific situations can be regarded as an error. We selected 16 out of those 24 errors, as the most applicable and relevant for the two simulations used in this study, and we also excluded errors referring to group processes, since we conducted the current study in individual settings. Even though the list includes some of the most widely shown errors (see also Dörner & Güss, 2022; Güss, Tuason & Orduna, 2015), it is most likely not a complete list of errors in complex and uncertain situations.

The experimenter read the list aloud with the participants and expanded on the common human errors associated with each of the DDM steps before the start of the simulation in the experimental, training group. The control, no-training group, only received the handout and explanation of the error training handout *after* completing the simulation game.

It is important to note, that for complex and uncertain situations, it is difficult to make general, always applicable, prescriptions on what to do and how to proceed. What constitutes an error depends on the circumstances. It might be desirable to act quicky and underplan, for example when a fire is approaching a town in WinFire; under other circumstances, one might be required to step back, and spend time to develop a detailed plan, for example when developing a strategy to deal with competitors in ChocoFine.

ChocoFine simulation. The ChocoFine micro-world simulation used in the current study is both highly complex and dynamic, with over a thousand variables (Dörner, 2000). The main interface of the program consists of three screens: (1) the main screen, (2) the production screen, and (3) the marketing screen (see Figure 1). The main screen shown in the background of Figure 1 shows the different kinds of chocolates produced by the company such as milk or bitter chocolates. The blue bars show how many chocolates are in stock and green bars show the demand. The red bar shows the total revenue of the company. The production screen, opened here in front of the main screen, shows the six machines of the company and which kind of chocolates can be produced on each of them. For example, machine 1 can produce milk, bitter, and mocha chocolates. For each day, or week, or month, participants can indicate which chocolates should be produced on which machines.

In the ChocoFine micro-world participants are taking the CEO postion of a small chocolate company. As CEO, participants must make decisions in the fields related to advertisement, marketing, production, and hiring/firing personnel. The complexity and uncertainty involved in ChocoFine requires participants to pick and choose from a multitude of possible decisions in each of these domains for each proceeding month.

The participants were instructed to manage production, marketing, personnel, and sales within the company. ChocoFine is described as a top management virtual game and considered as a complex simulation. It was originally developed in 1993 as a tool for business domains (Gerdes et al., 1993) at the University of Bamberg in Germany and has been revised several times.

Behavioral Changes in ChocoFine. (1) Performance was assessed as the total capital decision-makers accrued in month 12 at the end of the ChocoFine simulation. Although there are no limits for capital earnings, performance scores ranged between $-\$471,908$ to $\$3,086,693$ in the current sample. The overall mean was $\$1,132,355$ ($SD = 770,800$), the median for the sample was $\$1,157,057$.

Behavioral performance errors were calculated from the automatically saved log-file data as has been done in other studies on CPS and DDM (e.g., Güss et al., 2015; Stadler et al., 2019).

(2) Avoidance of DM versus Underplanning and Actionism: The emphasis on avoidance versus actionism of decision-making areas was operationalized based on the amount of money spent during each month for three decision-making areas: Advertising, market research, and transportation of goods. Regarding advertising, participants had the option of choosing general brand advertising or specific advertising for each type of chocolate or specific product profile components (i.e., luxurious or organic chocolate). Under the market research option, participants had access to analyze and purchase information related to their own products and clients, as well as their competitors' products and clients. Finally, under the transportation decision-making area, participants had expenses related to, for example, purchasing trucks for transportation of chocolate products. Besides changing production numbers, participants could focus therefore on none, one, two, or all three of these areas per month.

Since both data on the lower end (avoidance) and the higher end (Underplanning and Actionism) of the distribution stand for errors, we standardized the variable "Avoidance versus Actionism", $AA'_i = (AA_i - AA_{mean})^2 * 10^{-9}$, so that it can be used as a continuous variable with high scores indicating both errors.

(3) Status-quo bias/Methodism vs. Flexibility: The variations in each decision-making area related to methodism and flexibility of behavior were operationalized and extracted from the quantity of changes

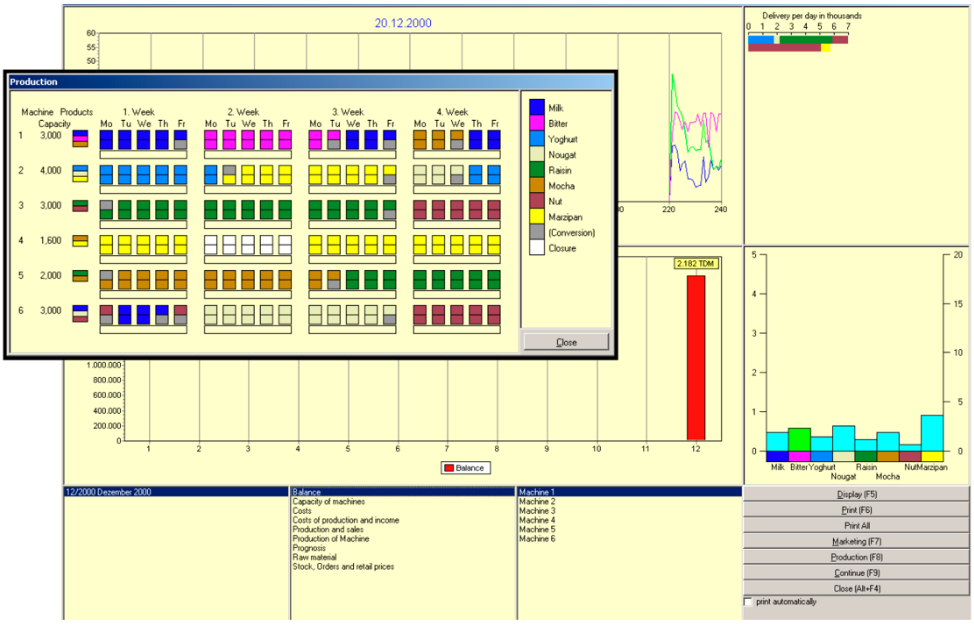

Figure 1. Screenshot of ChocoFine: Main Screen (Back) and Production Screen (Front).

made from one month to the next month in relation to the total number of months concluded in each of the three areas of decision making: Personnel, advertising, market research, and transportation of goods. Changes were coded using a binary method (e.g., "0" if no change occurred from one month to the next month, or "1" if a change occurred). The range was from 4 to 28. Low scores stand for few changes and methodism, applying past decisions to the current situation; high scores stand for many changes and indicate flexibility.

(4) Depth of information processing: The depth of exploration was operationalized as time spent working on the first two months of the ChocoFine simulation. Participants controlled proceeding from one month to the next month of the simulation by clicking the "continue" button after they made all decisions for the current month. The first month requires participants to adjust to their screens therefore, we considered the first two months as a holistic representation of exploration. Although time does not directly measure depth of information processing, past research has shown that experts spend more time initially analyzing a complex and dynamic situation than novices (Güss et al., 2017; Kobus et al., 2001). This initial deep processing has been a predictor of performance in these studies.

Demographic survey. Participants in both conditions received a demographic questionnaire after concluding the simulation assessing age, gender, and ethnicity.

The study was approved by the Institutional Review Board (#1200309-2). Participants were first instructed to sign an informed consent handout upon entering the computer lab. Participants in both conditions were given a detailed 3-page handout of the simulation instructions that illustrated the goal of the simulation as well as the different types of commands or screens they can use. The experimenter then verbally clarified the simulation instructions and answered any questions the participants had before beginning the trial version of ChocoFine. In both conditions, participants were given a one-month trial version to complete for approximately 10 minutes. This helped participants to familiarize themselves with the simulation. After the trial month, participants were given approximately 5 minutes to ask questions about the commands or troubleshooting the program before beginning the "true" experiment version of ChocoFine. This ChocoFine simulation took a total of 60-minutes to complete the 12 months. Data and decisions for every participant were automatically saved in log-files. Upon concluding the study, all participants completed the demographic questionnaire.

The experimental condition, however, included a 10-minute experimenter led training after receiving a detailed handout on the instructions of the ChocoFine simulation and before starting the 1-hour simulation. Participants in the experimental condition received the detailed list of errors after the trial version. The experimenter read with the participants through this 2-page list (see Appendix A) and answered related questions. This same list was given to the participants in the control group after they completed the "real" ChocoFine simulation – and to the experimental group again as well. Both experimental and control group were then asked to circle all the errors they thought they might have done while working on the simulation.

Results

Comparison of Training vs. No-training. An independent samples t-test was conducted to compare performance in ChocoFine between the experimental and control condition. As mentioned before, five participants were identified as extreme outliers and excluded from the performance analysis and for two participants no data were saved. The results revealed a marginally significant difference between both conditions. Participants who received error training compared to those who did not receive error training acquired more capital and thus, showed better total performance after 12 months of ChocoFine. The results indicate a medium effect size and were statistically significant using a one-sided test (see Table 1).

An independent samples t-test was also conducted to compare total self-reported errors in ChocoFine between the experimental and control conditions. The results revealed a significant difference between training and no-training groups. Participants who received error training reported fewer errors compared to the participants in the no-training group (see Table 1).

We then conducted independent samples t-tests to identify which condition (training vs. no-training) had more errors in the main six steps of DDM: Problem identification, goal definition, elaboration and prediction, planning, DM, and action, and self-reflection. Significance of the .05 level p-value was Bonferroni adjusted to minimize Type 1 error to .0083. Therefore, some of the initial findings for a p-value of .05 were not significant anymore. The results revealed a significant difference between training conditions and the total self-reported errors in the first step of the DDM process: Problem identification. Participants in the no-training group identified more errors related to problem identification than the training group. Participants in the no-training condition identified more errors for goal definition compared to individuals who received error training. Lastly, participants in the no-training group reported more errors related to evaluation of outcome and self-reflection compared to the training group.

However, no significant differences were found between the no-training group and training group and the three self-reported errors related to information gathering, elaboration and prediction, and planning, decision-making, and action (see Table 1).

Performance and Behavioral Errors/Strategies: Independent samples t-tests were conducted to compare behavioral errors identified in the ChocoFine simulation between training and no-training groups. The results indicated no significant differences in all three behavioral errors/strategies (i.e., depth of information processing, methodism vs. flexibility, and avoidance vs. underplanning and actionism (see Table 1).

Multiple linear regression analyses were conducted to predict performance based on the three behavioral errors/strategies. No significant regression was found, $F(3, 72) = 1.95$, $p = .13$ with an R^2 of .08. Depth of information processing was not a significant predictor, $\beta = .09$, $t = .76$, $p = .447$. Avoidance vs. Underplanning and Actionism was also not a significant predictor, $\beta = -.09$, $t = -.81$, $p = .42$. Methodism vs. Flexibility was a significant predictor, $\beta = .26$, $t = 2.30$, $p = .03$. The more flexible, i.e., the more behavior changes, the higher was the performance.

Performance and Self-Reported Errors. A Pearson correlation was conducted between performance and overall self-reported errors in the ChocoFine simulation. Preliminary analyses were performed to ensure no violation of the assumption of normality, linearity, and homoscedasticity. Although the result goes in the expected direction, it revealed no significant relationship between total self-reported errors ($M = 5.49$, $SD = 2.42$) and ChocoFine performance ($M = 1,132,355$, $SD = 770,800$), $r(74) = -.18$, $p = .13$.

Multiple regression analyses were conducted to regress performance on the 16 self-reported errors. Results were not significant, $R^2 = .21$, $F(16, 58) = .97$, $p = .51$. The only significant predictor was problem identification/methodism, $\beta = -.35$, $t = -2.30$, $p = .03$.

Discussion

The main goal of Study 1 was to investigate performance differences between training and non-training groups in ChocoFine. This short training referred to 16 errors related to DDM and CPS steps. First, there were only marginal significant performance differences between the two groups ($p = .08$). The training group performed slightly better than the non-training group. Second, the training group self-reported significantly fewer errors after the simulation and fewer errors to problem identification and self-reflection. As previous research has shown EMT can foster more active processing and self-reflection (e.g., Ivancic & Hekseth, 1998). Yet, the total number of self-reported errors did not correlate significantly with performance. Third, the two training groups did not differ in the three behavioral errors and strategies as assessed in the log files. Regression analyses showed that low status-quo bias and high flexibility in decision making predicted performance, a finding also shown in other research on DDM (e.g., Güss et al., 2017).

ChocoFine is a highly complex, yet slow developing simulation. ChocoFine is also a business simulation and applying business knowledge is advantageous when working on this problem (e.g., see a comparison of business experts and novices, Güss et al., 2017). In study 2, we wanted to administer the same training program, but in a dynamic situation that requires no background knowledge and that is more dynamic.

Study 2: WinFire

The goal of Study 2 was to extend the findings from Study 1 regarding the effects of EMT on performance, yet using a simulation that requires less background knowledge and that shows different characteristics. For Study 2, we used the WinFire simulation. WinFire is a computer simulation task in which the participants' main objective is to extinguish forest fires while simultaneously saving neighboring towns as well as the forest itself (Gerdes et al., 1993; Schaub, 2019). WinFire is described as moderate in complexity with few

Table 1. Descriptive Statistics, ChocoFine Performance, and Self-Reported Error Comparisons Across Training Conditions.

Outcome	Training group *M (SD)*	No-training group *M (SD)*	95% CI	*t*	*df*	*p*	*d*
Performance	1,270,133 (699,701)	962,159 (829,475)	[−657,471; 41,523]	−1.76	74	.08	.40
Self-reported errors							
Problem identification	0.60 (0.54)	0.85 (0.44)	[0.03, 0.48]	2.29	74	.03	.52
Goal definition	0.45 (0.55)	0.68 (0.53)	[−.03, 0.47]	1.79	74	.08	.41
Information gathering	0.76 (0.66)	0.82 (0.58)	[−0.22, 0.35]	0.43	74	.67	.10
Elaboration and prediction	1.05 (0.85)	1.29 (0.84)	[−0.14, 0.64]	1.26	74	.21	.29
Planning, DDM, and action	1.62 (0.79)	1.74 (1.05)	[−0.32, 0.55]	0.53	60.1	.60	.13
Evaluation of outcomes and self-reflection	0.48 (0.55)	0.76 (0.50)	[0.05, 0.53]	2.40	73.1	.02	.55
Total SR errors	4.95 (2.06)	6.15 (2.80)	[0.08, 2.31]	2.14	74	.04	.49
Behavioral errors/strategies							
Depth of information proc.	21.50 (11.02)	23.38 (11.66)	[−0.62, 0.29]	0.72	74	.47	.17
Methodism vs. Flexibility	18.00 (6.06)	16.74 (5.01)	[−0.23, 0.68]	−0.98	74	.33	.23
Avoidance vs. Actionism (standardized)	90.23 (246.71)	96.51 (270.29)	[−112.07, 124.65]	0.11	74	.91	.02

variables for the participants to work with (e.g., the use of trucks and helicopters to extinguish fires; and obstacles such as water levels, wind speed and direction, and unknown emerging fires). Although WinFire is considered moderate in complexity, the simulation is highly dynamic in nature. The situation in which participants find themselves in WinFire changes frequently with every decision made by the participant. For instance, the speed and direction in which the fires spread may increase based on a participant's decision to utilize fire trucks, as they are slower to extinguish fires than helicopters. WinFire is highly dynamic, even without direct intervention from the participant. Fires can start anywhere at any time. The core strategies of the DDM process one expects to detect throughout the WinFire simulation consist of assessing situations rapidly and identifying crucial situations, prioritizing, flexibility in the planning of resource allocation, and quick long-term decision-making to evade further escalation of the problem (i.e., rapid spread of wildfire; Güss et al., 2015).

Method

A total of 111 undergraduate students were recruited from the University of North Florida; 21 men, 88 women, and two participants identified as 'other'. Participant's ages ranged from 18 to 49 years ($M = 21.54$, $SD = 5.05$). Participants' ethnicities were identified as 64.9% White, 12.6% as Black, 15.5% as Hispanic, 4.5% as Asian, and the remaining 2.7% as 'Other'. There were 59 participants in the experimental, training group. There were 52 participants in the control, no-training group. Participants were randomly assigned to either condition. No pattern of relationship was found between gender and condition, $\chi^2(2) = 0.80$, $p = .67$, meaning there was a similar distribution in gender across both conditions. There were also no significant age difference between the no-training condition ($M = 21.77$, $SD = 5.12$) and the training condition ($M = 21.34$, $SD = 5.02$), $t(109) = -0.45$, $p = .66$. A total of 12 participants were excluded from the analysis as their log files only contained partial or extensive missing data regarding total performance measures perhaps due to the participants accidentally exiting the simulation before the time was over. Eight of those participants were in the no-training condition, four were in the training condition.

Simulation instructions. In both conditions (training vs. no-training) participants received a typed 3-page handout with in-depth instructions highlighting key commands for the WinFire simulation, alongside screenshots for easier visual comprehension.

Error training handout. Participants received an error training handout that discussed and expanded on the common human errors associated with each of the DDM steps discussed previously in Experiment 1 (e.g., "Elaboration and Prediction: Not considering time developments: We think in the here and now and do not consider time developments and situational changes happening over time"; or "Evaluation of Outcome and Self-Reflection: No monitoring and self-reflection: We think sometimes that if something is going well then it does not deserve further reflection", see Appendix A for a complete list of errors).

WinFire simulation. The microworld used in Experiment 2 is titled WinFire (Schaub, 2019). Participants completing the WinFire simulation assume the role of

Figure 2. Main Screen of the WinFire Simulation with Explanations.

commander in chief of a fire department to protect the villages and the forest from approaching fires (see Figure 2). Participants have the option to issue a series of commands to several fire trucks and helicopters in their effort to save the villages as well as the forest. Quick decisions and multitasking are a necessary component of WinFire to avoid fires from spreading. Performance was assessed as the total percentage of the saved, i.e., not burnt, area at the end of the simulation. Performance scores ranged from 0% (low performance) to 100% (high performance).

WinFire Behavioral Changes. (1) Performance: Performance was operationalized as the total percentage of the forest and villages saved from the forest fires. Total performance at second 600 was chosen as the performance variable for the following analyses as the simulation ended after 10 minutes. A score of 100 indicates the forest and area was saved in its entirety, while a score of 0 indicates that none of area was saved and everything burnt down.

The following behavioral measures operationalized from the automatically saved log-files were assessed similarly to ChocoFine.

(2) Avoidance of DM vs. Underplanning and Actionism: Avoidance or underplanning and actionism errors are found under the planning, decision-making, and action phase of DDM. The central focus of the decision-making areas was operationalized as choosing among several types of commands for any given unit (e.g., fire trucks and helicopters). Participants had several options to command their units to control the forest fire. The first option that participants could use in the WinFire simulation is the "Extinguish" command, which allows participants to smother the fire after moving their unit to a specified area. Participants could also use the "Patrol" command, where units can patrol a specific area of the simulation. The "Search" command option allows units to independently seek neighboring fires in the nearby designated area and drive to them. Finally, the "Clear" command is used when the burning forest fire is too extensive to extinguish. The participant will use controlled burning of the area to prevent fires from expanding. Finally, decision-makers can request further information to inquire about water storage for each unit and receive information on the state of a specific area of the forest to determine the percentage of the forest that is currently burning. Thus, the possible range is between zero to four types of decisions per fire truck or helicopter unit in a given time interval. Behavior was coded in the first minute of the simulation, divided into four different time intervals (0-15 seconds, 16-30 seconds, 31-45 seconds, and 46-60 seconds), and then summed up.

Avoidance of decision making refers to being hesitant in making decisions. Therefore, participants who employed avoidance in the WinFire simulation will have fewer instances of behaviors in any of the identified commands (i.e., Extinguish, Patrol, Search, or Clear). On the other hand, participants who seemingly rush through the simulation and act right away without first planning their actions is another identified error, called Underplanning and Actionism. An example of underplanning may involve a decision-maker dictating too many commands across all units. These behavioral actions indicate no clear plan in the WinFire simulation. Since both data on the lower end (avoidance) and the higher end (Underplanning and Actionism) of the distribution stand for errors, we standardized the variable "Avoidance versus Actionism", $AA'_i = (AA_i - AA_{mean})^2$, so that it can be used as a continuous variable with high scores indicating both errors.

(3) Methodism vs. Flexibility: The variations in each decision-making area were operationalized from the total number of commands that participants chose in the first, second, third, fourth quarter of the first minute of the WinFire simulation. For example, decision-making changes were assessed after the first fire begins and when the "extinguish", "patrol", "search", and/or "clear" commands were initiated. Behavioral changes were assessed in the WinFire simulation based on the total number of commands (i.e., Extinguish, Patrol, Search, and Clear) from one interval of time to the next interval of time. Changes were coded using a binary coding method (e.g., "1" indicates a change and a "0" indicates no behavioral change). The total number of behavioral changes will be assessed from the aggregate of interval one, interval two, interval three, and interval four of the first minute. It should be noted that although these numbers may not indicate an absolute change for specific units (i.e., helicopters or fire trucks), they do indicate a holistic change in strategy.

Status-quo bias/Methodism error refers to adopting non-changing strategies for much of the simulation. For example, in the first interval of time (0-15 seconds), a participant sends two trucks to specific locations and gives one truck an "Extinguish" command. During the second interval (e.g., 15-30 seconds), the participant then proceeds to only choose the "Extinguish" command for a single truck rather than utilizing both trucks to extinguish neighboring fires. On the opposite side of methodism is flexibility, participants who illustrate errors of potentially too much flexibility dictate too many behavioral changes during the computer simulation. Flexibility per se is a good thing. It refers to changing one's behavioral strategy, rather than using the same procedures in the following interval of time. For example, participants use two trucks for a given burning area, then utilize a helicopter for another specific area and simultaneously use the "Extinguish" and "Patrol" commands. The following interval, these trucks may be sent to another area instead, now given the commands to "search" and/or to "clear". This example shows a highly flexible adaptive approach. In sum, the lower the number, the fewer behavior changes happened, and the more methodism errors were shown. The higher the number, the more flexible was the strategic approach.

(4) Depth of information processing (Incomplete Information Gathering): An indicator of depth of information processing is information gathering. Information gathering was calculated for the first minute of the computer simulation as the frequency that participants actively inquired for further information, for example, the current percentage of the burning forest or the percentage of water in each unit. The higher the number, the more information was gathered and the deeper was the information processing.

Demographic survey. Finally, participants in both conditions received a demographic questionnaire after concluding the simulation assessing age, gender, and ethnicity (see Appendix C).

Participants in Experiment 2 followed the same procedures as Experiment 1. Participants were first instructed to complete an informed consent form before receiving the simulation instructions, working on a short test game, and then working on the "real" simulation. Participants in the control condition only received the error training handout at the end of the simulation whereas participants in the experimental condition were given error training before the start of the simulation for review and after the simulation as well. After the simulation every participant was instructed to encircle all the errors, they thought they did while working on the simulation.

The major difference in Experiment 2 was using the quick action computer simulation WinFire. Participants completing the WinFire simulation first received a detailed 3-page handout describing the simulation and commands. Participants then completed a separate trial version of the game which lasted approximately five minutes to familiarize themselves with the screen and commands. The "real" analyzed version of the WinFire simulation was then completed for ten minutes. Participants across both conditions circled their errors on the error training handout immediately following the end of the simulation. Upon concluding the study, all participants were asked to fill out a brief demographic questionnaire.

Results

Training vs. No-training Comparison. At first, we compared performance in WinFire between the experimental/training and control/no-training group. Since the performance variable was skewed and not normally distributed, the non-parametric Mann-Whitney U test was conducted. The Mann-Whitney U test result revealed that WinFire performance was significantly lower in the control, no-training condition ($Mdn = 84.09$, $n = 43$) compared to the experimental, training condition ($Mdn = 97.40$, $n = 56$), $U = 566.50$, $z = -4.52$, $p < .001$, $r = -.45$. These results reveal a large effect size (see Table 2).

To identify which condition (training vs. no-training) identified more errors between the main six self-reported errors in DDM (i.e., problem identification, goal definition, information gathering, elaboration and prediction, planning, DM, and action, evaluation of outcome and self-reflection) independent samples t-tests were conducted. The results revealed no significant difference between training conditions in the total self-reported errors (see Table 2).

Performance and Behavioral Errors. Independent samples t-tests were conducted to compare behavioral errors identified in the WinFire simulation between training and no-training groups. The results indicated no significant differences in depth of information processing. Regarding behavioral errors or strategies, the training group, however, demonstrated significantly

Table 2. Descriptive Statistics, WinFire Performance, and Self-Reported Error Comparisons Across Training Conditions.

Outcomes	Training group M (SD)	No-training group M (SD)	95% CI	t	df	p	d
Self-reported errors							
Problem identification	0.63 (0.56)	0.65 (0.53)	[−0.19, 0.25]	0.24	97	.81	.05
Goal definition	0.43 (0.53)	0.58 (0.54)	[−0.06, 0.37]	1.40	97	.17	.28
Information gathering	0.70 (0.66)	0.79 (0.60)	[−0.16, 0.35]	0.73	97	.47	.15
Elaboration and prediction	1.05 (0.82)	1.28 (0.88)	[−0.12, 0.57]	1.31	97	.19	.27
Planning, DM, and action	1.34 (0.92)	1.28 (0.77)	[−0.41, 0.28]	−0.35	97	.73	.07
Evaluation of outcomes	0.45 (0.54)	0.53 (0.55)	[−0.13, 0.31]	0.80	97	.42	.16
Total SR errors	4.58 (2.14)	5.12 (2.13)	[−0.33, 1.39]	1.22	97	.23	.25
Behavioral Errors/Strategies							
Depth of information proc.	13.16 (8.82)	10.53 (7.35)	[−5.93, 0.68]	−1.58	97	.12	.32
Methodism vs. Flexibility	4.29 (3.01)	2.65 (1.80)	[−2.60, −0.67]	−3.36	92.01	.001	.64
Avoidance vs. Actionism (standardized)	162.52 (300.20)	32.58 (32.47)	[−210.90, −49.00]	−3.22	56.67	.002	.57

more flexibility, and showed significantly less avoidance in decision making (see Table 2).

Multiple linear regression analyses were conducted to predict performance based on the three behavioral errors/strategies. Six outliers with performance scores below 38% of saved forest were removed. A significant regression was found, $F(3,89) = 3.39$, $p = .022$ with an R^2 of .10. Depth of information processing was not a significant predictor, $\beta = .05$, $t = .34$, $p = .73$. Avoidance vs. Underplanning and Actionism was also not a significant predictor, $\beta = -.02$, $t = -.18$, $p = .86$. Methodism vs. Flexibility was a significant predictor, $\beta = .30$, $t = 2.06$, $p = .04$. The more flexible and the fewer status-quo bias, i.e., the more behavior changes, the higher was the performance.

Performance and Self-Reported Errors. The relationship between performance and total self-reported errors after completion of the simulation was investigated using a Spearman correlation. The results indicated that there was a weak, negative correlation between total performance in the WinFire simulation ($M = 84.24$, $SD = 19.43$) and participant's total self-reported errors after completion of the simulation ($M = 4.82$, $SD = 2.14$), $rs(97) = -.14$, $p = .15$, $p = .07$. The higher the overall performance in the WinFire simulation, the lower is participants' numbers of total self-reported errors after the simulation.

Multiple regression analyses were conducted to regress performance on the 16 self-reported errors (again excluding the 6 outliers). Results were not significant, $R^2 = .13$, $F(16, 76) = .76$, $p = .72$. Not considering side effects and future developments was the only significant predictor, $\beta = -.30$, $t = -2.23$, $p = .03$.

Discussion

The main goal of Study 2 was to investigate performance differences between training and non-training groups in WinFire. WinFire requires less background knowledge compared to ChocoFine, is less complex, but more dynamic. Results showed significantly better performance in EMT training compared to the non-training group. Second, there were no significant differences regarding self-reported errors between training group and non-training group. The total number of self-reported errors correlated marginally significantly with performance. The higher the overall performance in the WinFire simulation, the lower participants' numbers of total self-reported errors after the simulation. Third, regarding the three behavioral errors and strategies as assessed in the log files the training group showed less status-quo bias and more flexibility/adaptivity and less avoidance. Previous research reports that real-time decision-making, such as in WinFire, requires individuals to follow a feedforward strategy (i.e., decisions that involve making predictions of a future state (Brehmer, 1996; Gonzalez, 2005). Therefore, environments that are ever-changing force decision-makers to embrace change and modify their strategies accordingly throughout the progression of the task. Regression analyses in WinFire showed that low status-quo bias and high flexibility in decision making predicted performance, a finding also shown in previous research on DDM (e.g., Güss et al., 2017).

Comparing Self-reported Errors in ChocoFine and WinFire

We also compared self-reported errors regarding the six steps of decision making and problem solving in ChocoFine and WinFire. Due to the different task characteristics, we expected participants to self-report more errors in ChocoFine compared to WinFire regarding the three steps: information collection, elaboration and prediction, and planning, decision-making, and action. We did not expect differences regarding the other three steps: Goal definition, problem identification, and outcome evaluation and self-reflection.

We run a 2 x 2 MANOVA with one independent variable being simulation (ChocoFine versus WinFire) and the other independent variable being experimental condition (Experimental training group versus control group). The six dependent variables were the totals of the self-reported errors in the six DDM steps. The descriptive statistics are reported in Tables 1 and 2.

Using Wilk's Lamda, there was no overall significant effect of simulation on the number of self-reported errors in the six DDM steps, $\lambda = 0.943$, $F(6, 166) = 1.675$, $p = .13$, $\eta_{p^2} = .057$. Separate uni-variate ANOVAs on the six outcome variables revealed only one significant difference, i.e., more self-reported errors regarding planning, decision making, and action in ChocoFine compared to WinFire.

There was also no overall significant effect of training on number of self-reported errors in the six DDM steps, $\lambda = 0.941$, $F(6, 166) = 1.743$, $p = .114$, $\eta_{p^2} = .059$. Separate uni-variate ANOVAs on the six outcome variables revealed two significant differences, i.e., more self-reported errors regarding goal definition and evaluation of outcomes in the control group compared to the training group (see Table 3). These findings support only partially the hypothesis that participants in the training conditions of both computer simulations will report fewer self-reported errors than participants in the control conditions.

General Discussion

The goal of the current study was to illustrate how a brief error management training (EMT) could facilitate performance in DDM and CPS environments. We expected that individuals in the experimental, training condition would have better performance compared to the control, no-training condition. We then investigated whether self-reported errors and behavioral errors observed in the log files could explain performance differences, and we investigated differences between the two simulations.

Microworlds vary in their complexity, uncertainty, and dynamics; therefore, it is important to analyze how DDM differs in different conditions (see also Greiff et al., 2014). Our results indicated that training had a significant positive influence on performance in the WinFire simulation and a marginally significant positive effect in the ChocoFine simulation. Considering that the two microworlds have different task characteristics, it is important to note that a general EMT had a stronger effect in the more transparent WinFire simulation.

Regarding self-reported errors, there was mixed support for our hypothesis. In the ChocoFine simulation self-reported errors were related to the training condition. The training group reported fewer errors related to the steps of problem identification, and outcome evaluation and self-reflection. Overall, there were fewer self-reported errors in the training group regarding goal definition and outcome evaluation and self-reflection compared to the control group. However, the findings in the WinFire study did not support our hypothesis. The training group did not self-report fewer errors.

Correlations between performance and total self-reported errors in both WinFire and ChocoFine show that the higher the performance, the lower the self-reported errors. Yet, these correlations were not significant. This finding shows that participants' perception of their own DDM errors does not relate to their objective task performance.

One possible explanation for the variability in our findings might be the difference in time constraints between the WinFire and ChocoFine simulations. The self-reported errors task that participants completed at the end of the simulation necessitated that they greatly reflect on their actions compared to their goals and overall performance in the simulations. The participants spent one hour completing the ChocoFine simulation mainly because it had an extensive list of behaviors and actions they could modify and change to their liking, while the WinFire microworld took only ten minutes and had only a few possible commands. Therefore, the absence of fewer time constraints experienced in the ChocoFine simulation possibly allowed for more use of self-reflective strategies than the WinFire. On the other hand, participants who worked in the WinFire simulation were under higher time pressure.

Additionally, we expected behavioral errors to predict performance and to differ between training and no-training groups. We operationalized three behavioral errors and strategies and coded them from the automatically saved log files of participants in the two simulations. In the ChocoFine study, only status-quo bias correlated negatively with performance. In the fast-paced WinFire simulation, the training group showed less often status-quo bias, more adaptivity and less avoidance. Regression analyses for both ChocoFine and Winfire showed that status-quo bias negatively predicted performance. In both situations, not doing anything and sticking to old plans and decisions was detrimental. Instead, adjusting to the ever-changing situation and changing behaviors was advantageous. These results support previous findings suggesting that dynamic environments are ever-changing and thus, performance is dependent on the decision-makers subsequent decisions (Dörner, 1996; Edwards, 1962; Güss, 2011).

The findings on the status-quo bias versus flexibility are very relevant for organizations. "Managers and organizations should be prepared and proactive to overcome the biases, to avoid becoming trapped in the vicious cycle of rigidity, and to cope effectively with the uncertainties of a dynamic environment" (Shimizu & Hitt, 2004, p. 44). Research has demonstrated that

Table 3. Separate univariate ANOVAs on the outcome variables, i.e., self-reported errors on DDM steps for ChocoFine and WinFire.

	Simulation				Training			
Self-reported errors:	*F*	*df*	*p*	η_{p^2}	*F*	*df*	*p*	η_{p^2}
Problem identification	1.14	1,171	.288	< .01	3.09	1,171	.08	.02
Goal definition	.51	1,171	.475	< .01	5.15	1,171	.03	.03
Information gathering	.26	1,171	.611	< .01	.65	1,171	.42	< .01
Elaboration and prediction	.001	1,171	.972	< .01	3.30	1,171	.07	.02
Planning, DDM, and action	7.34	1,171	.007	.04	.04	1,171	.84	< .01
Evaluation of outcomes and self-reflection	2.49	1,171	.117	.01	5.24	1,171	.02	.03

Note. None of the interactions was statistically significant.

the tendency to show status-quo bias increases with experience, as knowledge in their domain also increases (Burmeister & Schade, 2007). A recent study reveals that a higher status within a business organization can influence a decision-makers's susceptibility to status quo biases (see Sana Chiu et al., 2020). Thus, especially leaders in organizations have to be aware of the possible traps their expertise can pose by predisposing them to falling for the status quo bias.

One limitation of our EMT is related to its brief implementation. We provided participants with a list of errors, their definitions, and some examples. The training referred to knowledge acquisition. It did not include practicing. Participants had to do the actual transfer in the specific WinFire and ChocoFine simulations. For future studies, the training could also facilitate the transfer of knowledge by showing and explaining specific short video clips of errors in microworlds. Such an approach might lead to stronger findings. It is nevertheless noteworthy that even our EMT focusing on knowledge acquisition showed some positive effects in the two simulations.

Another limitation is related to the list of errors itself. A general description of errors, based on research findings of previous DDM studies and as proposed in the current research, provides general strategies to "pick" from. Yet, they sometimes seem contradictory (e.g., underplanning versus overplanning). In a similar way, proverbs can provide sometimes contradictory advice. For example, "Sleep on it" suggests that people wait and reflect before making an important decision. Other proverbs such as "Never put off till tomorrow what you can do today" suggest that people seize an opportunity and decide and act right away. The important task is to analyze the specific situational circumstances and to determine which proverb advice or relevant error avoidance will most likely lead to the most adaptive and promising outcomes. This specific skill of cognitive analysis and self-reflection and the avoidance of the status-quo bias has also been called strategic flexibility (Brozovic, 2018) or cognitive flexibility (Laureiro-Martínez & Brusoni, 2018, p. 1031).

A major limitation of our study is the generalizability of our findings due to the large use of a college student population with a disproportionally high number of Whites compared to other ethnic groups. However, descriptive results showed a large distribution between age and gender, and most of the participants were also employed. Past research has demonstrated that college students typically engage in more risk taking (Salameh et al., 2014). Future research could utilize a sample of participants outside of a college demographic to account for maturational changes in middle-aged and older adult individuals and further increase generalizability (although age did not correlate with ChocoFine performance, $r(74) = .07$, $p = .56$, nor did age correlate with WinFire performance, $r(97) = -.14$, $p = .17$).

Another limitation is that we could not operationalize the self-reported errors in a way to assess them in the behavioral log files. How can lack of self-reflection be assessed in the log-files? Future research could include videos of screen developments and participant reactions and/or thinking-aloud protocols (e.g., Güss et al., 2010) in addition to the saved log files to assess more of these errors.

A final limitation is related to the results, which are based on aggregate averages of participant decisions saved in log files. Future research could investigate errors over time and how strategies acquired in the short EMT are implemented and when they are implemented. This could be done, for example, by time-based or event-based log-linear data analyses or by instructing participants to think-aloud (Smith et al., 2022) or by interrupting participants during the decision-making process and asking them short questions.

In conclusion, the current study showed that an EMT can enhance performance in a highly complex and slow-paced task environment such as ChocoFine and in a less complex but highly dynamic and fast-paced task environment such as WinFire. Little status-quo bias and high adaptability were positively related to performance in both microworlds. We conclude that creating awareness of possible errors in these tasks can stimulate self-reflection and monitoring through EMT. The promotion of self-reflection through EMT ultimately increases the DDM and CPS performance. The results of the present study have practical applications for leaders who make decisions in stressful, complex, and dynamic work environments and might fall for their overconfidence in their own expertise. Organizations may benefit from utilizing an EMT training program that encourages self-regulatory practices in fast-paced and uncertain environments.

Acknowledgements: We are thankful to Dietrich Dörner and Harald Schaub for allowing us to use their ChocoFine and WinFire simulations. This research was supported in part by a grant from the University of North Florida's Delaney Presidential Professorship to the second author. We would also like to thank the reviewers of this article and the journal editor for their thoughtful and detailed comments and helpful suggestions.

Declaration of conflicting interests: The authors report there are no competing interests to declare.

Peer Review: This article has been reviewed by two anonymous peer reviewers before publication.

Handling editor: Wolfgang Schoppek

Citation: Hermida, J. & Güss, D. C. (2024). Some Evidence for the Effectiveness of a Brief Error Management Training in Complex, Dynamic, and Uncertain Situations. *Journal of Dynamic Decision Making, 10*, 1. https://doi.org/10.11588/jddm.2024.1.90705

Received: 15.09.2022
Accepted: 18.06.2023
Published: 22.05.2024

References

Bahník, Š., Houdek, P., Vrbová, L., & Hájek, J. (2019). Variations on anchoring: Sequential anchoring revisited. *Judgment and Decision Making, 14*(6), 711–720.

Billings, R. S., & Hermann, C. F. (1998). Problem identification in sequential policy decision making: The re-representation of problems. In D. A. Sylvan & J. F. Voss (Eds.), P*roblem representation in foreign policy decision-making* (pp. 53-79). Cambridge University Press.

Brehmer, B. (1996). Man as a stabiliser of systems: From static snapshots of judgement processes to dynamic decision making. *Thinking & Reasoning, 2*(2–3), 225–238. https://doi.org/10.1080/135467896394528

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits if the laboratory and the deep blue sea of field study. *Computers in Human Behavior*, *9*(2-3), 171–184. http://doi.org/10.1016/0747-5632(93)90005-D

Brozovic, D. (2018). Strategic flexibility: A review of the literature. *International Journal of Management Reviews, 20*(1), 3-31. https://doi.org/10.1111/ijmr.12111

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. https://doi.org/10.1016/j.jbusvent.2006.04.002

Carroll, L. S., White, M. P., & Pahl, S. (2011). The impact of excess choice on deferment of decisions to volunteer. *Judgment and Decision Making, 6(*7), 629.

Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. Journal of Consumer Psychology, 25(2), 333–358. https://doi.org/10.1016/j.jcps.2014.08.002

Chiu, S. C. S., Pathak, S., Hoskisson, R. E., & Johnson, R. A. (2022). Managerial commitment to the status quo and corporate divestiture: Can power motivate openness to change?. *The Leadership Quarterly, 33*(3), 101459. https://doi.org/10.1016/j.leaqua.2020.101459

Damm, L., Nachtergaële, C., Meskali, M., & Berthelon, C. (2011). The evaluation of traditional and early driver training with simulated accident scenarios. *Human Factors*, *53*(4), 323–337. https://doi.org/10.1177/0018720811413765

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition. The distinctiveness heuristic. *Journal of Memory and Language, 46*(4), 782-803. http://doi.org/10.1006/jmla.2001.2822

Donovan, S., Güss, C. D., & Naslund, D. (2015). Improving dynamic decision making through training and self-reflection. *Judgment and Decision Making, 10*(4), 284–295. https://doi.org/10.1017/S1930297500005118

Dörner, D. (1996). *The logic of failure: Recognizing and avoiding error in complex situations.* Metropolitan Books.

Dörner, D. (1999). *Bauplan für eine Seele* [Blueprint for a soul]. Rowohlt.

Dörner, D. (2000). SchokoFin [CHOCO FINE]. *Computer simulation*. Otto-Friedrich Univeristät Bamberg, Germany.

Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology, 8*, 1153. https://doi.org/10.3389/fpsyg.2017.01153

Dörner, D., & Güss, C. D. (2013). PSI: A computational architecture of cognition, motivation, and emotion. *Review of General Psychology, 17*(3), 297-317. https://doi.org/10.1037/a0032947

Dörner, D., & Güss, C. D. (2022). Human error in complex problem solving and dynamic decision making: A taxonomy of 24 errors and a theory. *Computers in Human Behavior Reports, 7,* 100222. https://doi.org/10.1016/j.chbr.2022.100222.

Dörner, D., & Meck, U. (2022). The red trousers. About confirmative thinking and perceptual defense in complex and uncertain domains of reality. *Journal of Dynmaic Decision Making, 8*(1), 1-14. https://doi.org/10.11588/jddm.2022.1.84578

Dörner, D., Schaub, H., & Strohschneider, S. (1999). Komplexes Problemlösen- Königsweg der Theoretischen Psychologie? [Complex problem solving—Via regia for theoretical psychology?]. *Psychologische Rundschau*, *50*(4), 198–205. https://doi.org/10.1026/0033-3042.50.4.198

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors, 4*(2), 59–73. https://doi.org/10.1177/001872086200 400201

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving, 4*, 19-42. http://doi.org/10.7771/1932-6246.1118

Frensch, P., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The*

European perspective (pp. 3-25). Lawrence Erlbaum Associates, 1995.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning, 7(*1), 69–89. https://doi.org/10.1080/13546780042000046

Funke, J., Fischer, A., & Holt, D. V. (2017). When less is less: solving multiple simple problems is not complex problem solving—A comment on Greiff et al. (2015). *Journal of Intelligence, 5*(1), 5. https://doi.org/10.3390/jintelligence5010005

Funke, J., & Greiff, S. (2017). Dynamic problem solving: Multiple-item testing based on minimally complex systems. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 427–443). Springer International Publishing. https://doi.org/10.1007/978-3-319-50030-0_25

Gerdes, J., Dörner, D. & Pfeiffer, E. (1993). *Interaktive Computersimulation "Winfire"* [The interactive computer simulation "Winfire"]. Otto-Friedrich-Universität Bamberg, Germany: Lehrstuhl Psychologie II.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*(4), 591–635. https://doi.org/10.1016/S0364-0213(03)00031-4

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Process, 96*(2), 142-154. https://doi.org/10.1016/j.obhdp.2004.11.002

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. https://doi.org/10.1016/j.chb.2004.02.014

Gonzalez, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic decision making: Learning processes and new research directions. *Human Factors*, *59*(5), 713–721. https://doi.org/10.1177/0018720817710347

Grant, A. M., Franklin, J., & Langford, P. (2002). The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality, 30*(8), 821-836. http://doi.org/10.2224/sbp.2002.30.8.821

Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2015). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning, 21*(3), 356–382. https://doi.org/10.1080/13546783.2014.989263

Greiff, S., Krkovic, K., Nagy, G., (2014). The systematic variation of test characteristics facilitates the understanding of task difficulty: A cognitive diagnostic modeling approach to complex problem solving. *Psychological Test and Assessment Modeling, 56*(1), 83-103.

Güss, C. D. (2011). Fire and ice: Testing a model on cultural values and complex problem solving. *Journal of Cross-Cultural Psychology, 42*(7), 1279-1298. https://doi.org/10.1177/0022022110383320

Güss, C. D., Devore Edelstein, H., Badibanga, A., & Bartow, S. (2017). Comparing business experts and novices in complex problem solving. *Journal of Intelligence, 5*(2), 20. https://doi.org/10.3390/jintelligence5020020

Güss, C. D., & Dörner, D. (2011). Cultural differences in dynamic decision-making strategies in a non-linear, time delayed task. *Cognitive Systems Research, 12*(3-4), 265-376. http://doi.org/10.1016/j.cogsys.2010.12.003

Güss, C. D., & Starker, U. (2023). The influence of emotion and emotion regulation on complex problem-solving performance. *Systems, 11*(6), 276. https://doi.org/10.3390/systems11060276

Güss, C. D., Tuason, M. T., & Gerhard, C. (2010). Cross-national comparisons of complex problem-solving strategies in two microworlds. *Cognitive Science, 34*(3), 489–520. https://doi.org/10.1111/j.1551-6709.2009.01087.x

Güss, C. D., Tuason, M. T., & Orduña, L. V. (2015). Strategies, tactics, and errors in dynamic decision making. *Journal of Dynamic Decision Making*, *1*(3), 1-14. https://doi.org/10.11588/JDDM.2015.1.13131

Gully, S. M., Payne, S. C., Koles, K. L. K., & Whiteman, J.-A. K. (2002). The impact of error training and individual differences on training outcomes: An attribute-treatment interaction perspective. *Journal of Applied Psychology*, *87*(1), 143–155. https://doi.org/10.1037/0021-9010.87.1.143

Harman, J. L., Zhang, D., & Greening, S. G. (2019). Basic processes in dynamic decision making: How experimental findings about risk, uncertainty, and emotion can contribute to police decision making. *Frontiers in Psychology, 10*: 2140. https://doi.org/10.3389/fpsyg.2019.02140

Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology*, *73*(1), 68–73. https://doi.org/10.1037/0021-9010.73.1.68

Heimbeck, D., Frese, M., Sonnentag, S., & Keith, N. (2003). Integrating errors into the training process: The function of error management instructions and the role of goal orientation. *Personnel Psychology, 56*(2), 333–361. https://doi.org/10.1111/j.1744-6570.2003.tb00153.x

Ivancic, B., & Hekseth, B. (2000). Learning from error in a driving simulation: effects on driving skill and confidence. *Ergonomics*, *43*(12), 1966–1984. https://doi.org/10.1080/00140130050201427

Kahneman, D., & Renshon, J. (2009). Hawkish biases. In A. T. Thrall & J. K. Cramer (Eds.), *American foreign policy and the politics of fear: Threat inflation since 9/11* (pp. 79-96). Routledge.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5*(1), 193-206. https://doi.org/10.1257/jep.5.1.193

Kanfer, R., Ackerman, P. L., & Heggestad, E. D. (1996). Motivational skills and self-regulation for learning: A trait perspective. *Learning and Individual Differences*, *8*(3), 185-209. https://doi.org/10.1016/S1041-6080(96)90014-X

Karakul, M., & Qudrat-Ullah, H. (2008). How to improve dynamic decision making? Practice and promise. In H. Qudrat-Ullah, J. M. Spector, & P.I. Davidsen (Eds.), *Complex decision making. Theory and practice* (pp. 3-24). Springer.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, *93*(1), 59–69. https://doi.org/10.1037/0021-9010.93.1.59

Keleș, T., & Yazgan, Y. (2022). Indicators of gifted students' strategic flexibility in non-routine problem solving. *International Journal of Mathematical Education in Science and Technology*, *53*(10), 2797-2818. https://doi.org/10.1080/0020739X.2022.2105760

Klein, G. (1999). Applied decision making. In P. A. Hancock (Ed.), *Human performance and ergonomics* (pp. 87-107). Academic Press.

Kobus, D. A., Proctor, S., & Holste, S. (2001). Effects of experience and uncertainty during dynamic decision making. *International Journal of Industrial Ergonomics, 28*(5), 275-290. https://doi.org/10.1016/S0169-8141(01)00022-1.

Kretzschmar, A., & Süß, H.-M. (2015). A study on the training of complex problem solving competence. *Journal of Dynamic Decision Making, 1*(4), 1–15. https://doi.org/10.11588/jddm.2015.1.15455

Krijnen, J. M. T., Zeelenberg, M., & Breugelmans, S. M. (2015). Decision importance as a cue for deferral. *Judgment and Decision Making, 10*(5), 407–415. http://doi.org/10.1017/S1930297500005556

Lafond, D., DuCharme, M. B., Gagnon, J.-F., & Tremblay, S. (2012). Support requirements for cognitive readiness in complex operations. Journal of Cognitive Engineering and Decision Making, 6(4), 393–426. https://doi.org/10.1177/1555343412446193

Laureiro-Martínez, D., & Brusoni, S. (2018). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal, 39*(4), 1031-1058. https://doi.org/10.1002/smj.2774

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. http://doi.org/10.1037/0003-066X.57.9.705

Loh, V., Andrews, S., Hesketh, B., & Griffin, B. (2013). The moderating effect of individual differences in error-management training: Who learns from mistakes? *Human Factors*, *55*(2), 435–448. https://doi.org/10.1177/0018720812451856

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. https://doi.org/10.1287/isre.2014.0513

Nicholson, B., & O'Hare, D. (2014). The effects of individual differences, prior experience and cognitive load on the transfer of dynamic decision-making performance. *Ergonomics, 57*(9), 1353-1365. https://doi.org/10.1080/00140139.2014.933884

Osman, M. (2010). Controlling uncertainty: a review of human behavior in complex dynamic environments. *Psychological Bulletin*, *136*(1), 65–86. http://doi.org/10.1037/a0017815

Osman, M., & Palencia, D. O. V. (2019). The future of problem solving research is not complexity, but dynamic uncertainty. *Journal of Dynamic Decision Making, 5*(11), 1-4. https://doi.org/10.11588/jddm.2019.1.69300

Ramnarayan, S., Strohschneider, S., & Schaub, H. (1997). Trappings of expertise and the pursuit of failure. *Simulation & Gaming*, *28*(1), 28-43. https://doi.org/10.1177/1046878197281004

Reason, J. (1990). *Human error.* Cambridge University Press. https://doi.org/10.1017/CBO9781139062367

Russo, J. E., & Corbin, J. C. (2016). Not by desire alone: The role of cognitive consistency in the desirability bias. *Judgment and Decision Making, 11*(5), 449–459. https://doi.org/10.1017/S1930297500004551

Salameh, P., Salamé, J., Waked, M., Barbour, B., Zeidan, N., & Baldi, I. (2014). Risk perception, motives, and behaviours in university students. *International Journal of Adolescent and Youth*, *19*(3), 279-292. https://doi.org/10.1080/02673843.2014.919599

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*(1), 7-59.

Sanders, R. L., & McKeown, L. (2008). Promoting reflection through action learning in a 3D virtual world. *International Journal of Social Sciences, 2*(1), 50-58. http://doi.org/10.5281/zenodo.1062542

Schaub, H. (2007). The importance of the characteristics of the task to understand team mental models. *CoDesign*, *3*(1), 37–42. http://doi.org/10.1080/15710880601170800

Schaub, H. (2019). *WinFire*. Simulation of Forest Fires (revised version). University of Bamberg.

Schoppek, W., & Fischer, A. (2017). Common process demands of two complex dynamic control tasks: Transfer is mediated by comprehensive strategies. *Frontiers in Psychology, 8.* https://doi.org/10.3389/fpsyg.2017.02145

Shimizu, K., & Hitt, M. A. (2004). Strategic flexibility: Organizational preparedness to reverse ineffective strategic decisions. *Academy of Management Executive, 18*(4), 44-59. https://doi.org/10.5465/AME.2004.15268683

Smith W, Hermida J and Güss CD (2022). "Oh no, the forest is burning!" Cultural differences in the complex problem-solving process only under high uncertainty. *Frontiers in Psychology, 13*:965623. https://doi.org/10.3389/fpsyg.2022.965623

Stadler, M., Fischer, F., & Greiff, S. (2019). Taking a closer look: An exploratory analysis of successful and unsuccessful strategy use in complex problems. *Frontiers in Psychology, 10.* https://doi.org/10.3389/fpsyg.2019.00777

Stenmark, C. K., Antes, A. L., Wang, X., Caughron, J. J., Thiel, C. E., & Mumford, M. D. (2010). Strategies in forecasting outcomes in ethical decision-making: Identifying and analyzing the causes of the problem. *Ethics & Behavior, 20*(2), 110–127. https://doi.org/10.1080/10508421003595935

Sternberg, R. J. (1986). *Intelligence applied? Understanding and increasing your intellectual skills.* Harcourt Brace Jovanovich.

Tversky, A. & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decisions. *Psychological Science, 3*(6), 358-361. https://doi.org/ 10.1111/j.1467-9280.1992.tb00047.x

von Clausewitz, C. (1873). *On war.* Trübner.