FRANK, DAVID J., Ph.D. Aging and Task Representation Updating. (2015)
Directed by Dr. Dayna R. Touron. 96 pp.

Older adults' performance decrements can sometime be traced back to inferior strategic choices compared to their younger counterparts. Additionally, older adults often fail to revise their strategic choices with task experience (Bieman-Copland & Charness, 1994; Brigham & Pressley, 1988; Lovett & Schunn, 1999; Price, Dunlosky, & Hertzog, 2008; Touron & Hertzog, 2004a, 2004b; Touron, Hoyer, & Cerella, 2004). Metacognitive models of strategy selection suggests that beliefs, prior knowledge, goals, and task representation influence strategic decisions (e.g., Winne & Hadwin, 1998). No studies to date have attempted to compare task representation in older and younger adults to determine whether older adults' poor strategic choices might be driven by an impoverished understanding of the tasks they are asked to engage in. In two studies we used a pathfinder methodology to elicit conceptual knowledge about a novel chemistry task. In both studies, more conceptual knowledge was related to superior task performance in both younger and older adults. However, we found no evidence of age-related deficits in task representation, formation, or utilization. Surprisingly, participants' task representation scores did not improve following task practice. However, performance improved over trials, even for items that had to be learned with task practice, suggesting that task representation updating did occur. These findings provide indirect evidence of task representation updating in both younger and older adults. However, no age deficits in the ability to update task representations were found. Exploratory analyses suggest that performance in younger adults was related to motivational issues, whereas performance in older adults was driven by higher levels of processing speed and crystallized intelligence.

AGING AND TASK REPRESENTATION UPDATING

by

David J. Frank

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2015

Approved by

Dayna R. Touron___________
Committee Chair

APPROVAL PAGE

This dissertation written by DAVID J. FRANK has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair	Dayna R. Touron
Committee Members	Michael J Kane
	Peter F. Delaney
	Suzanne Vrshek-Schallhorn

04/27/2015
Date of Acceptance by Committee

04/27/2015
Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Older and younger adults often choose different strategies when performing the same task; this has been shown in a variety of cognitive domains (Bouazzaoui et al., 2010; Dunlosky & Connor, 1997; Frank, Touron, & Hertzog, 2013; Hertzog, Mcguire, & Lineweaver, 2010; Hines, Touron, & Hertzog, 2009; Lemaire, Arnaud, & Lecacheur, 2004; Lemaire & Lecacheur, 2007; Lemaire, 2010; Rawson & Touron, 2009; Rogers & Gilbert, 1997; Starns & Ratcliff, 2010; Touron & Hertzog, 2004a, 2004b). Unfortunately, older adults often choose less effective strategies compared to younger adults. Relative to younger adults, older adults make poorer strategy choices when: choosing memory encoding strategies (Brigham & Pressley, 1988; Dunlosky, Hertzog, & Powell-Moman, 2005; Dunlosky & Hertzog, 2001), deciding which information to study for a test and for how long (Dunlosky & Connor, 1997; Hines et al., 2009), deciding whether to round up or down when performing mental arithmetic (Green, Lemaire, & Dufau, 2007; Lemaire et al., 2004), deciding whether to retrieve information from memory or compute the answer (Frank et al., 2013; Hertzog, Touron, & Hines, 2007; Hertzog & Touron, 2011; Rawson & Touron, 2009; Touron, Hertzog, & Frank, 2012; Touron, Swaim, & Hertzog, 2007; Touron & Hertzog, 2004a, 2004b, 2009; Touron, 2006), and determining the appropriate speed-accuracy trade-off to maximize performance (Starns & Ratcliff, 2010; Strayer & Kramer, 1994). Furthermore, older adults are less likely to update or adjust their strategic choices in response to task stimuli (Lemaire et al., 2004; Lemaire & Lecacheur, 2007; Lemaire, 2010) or poor task performance (Brigham & Pressley, 1988; Price, Hertzog, & Dunlosky, 2008).

The next section briefly describes the role of task understanding in models of strategy selection. The subsequent sections describe how aging impacts each of those processes, what gaps in the literature remain, and how connecting to the literature on mental models, including conceptual knowledge, perspectives, and techniques, can help fill these gaps.

Metacognitive Models of Strategy Selection

Older adults' strategic choices have been shown to be influenced by their beliefs about strategies, and the beliefs in their ability to use strategies (Brigham & Pressley, 1988; Dunlosky et al., 2005; Dunlosky & Hertzog, 2001; Frank et al., 2013; Hertzog et al., 2007; Hertzog & Touron, 2011; Hines et al., 2009; Price et al., 2008; Rawson & Touron, 2009; Touron & Hertzog, 2004b, 2004a). However, models of strategy selection also highlight the importance of having an accurate understanding about the task itself (Bromme, Pieschl, & Stahl, 2009; Dunlosky & Hertzog, 2000; Lovett & Schunn, 1999; Muis, 2007; Winne & Hadwin, 1998). For example, one may choose different study strategies based on whether study time is limited, the test is essay, multiple choice, or fill in the blank, and whether the test is open or closed book. Likewise, initial misconceptions about a task may restrict the set of strategic options one considers, or may result in the misapplication of those strategies (Bromme et al., 2009; Winne & Hadwin, 1998). Theories of strategy selection allow for changes in task representations (and changes in strategy beliefs) via metacognitive monitoring and updating. That is, people can monitor the results of their strategic choices, monitor the structure of the task, and compare these to their mental representations of the task. This process is referred to as metacognitive monitoring (Nelson & Narens, 1990). If accurate monitoring occurs, people may then use this information to update their task representations and strategic beliefs, and change their strategic choices (Bromme et al., 2009; Dunlosky & Hertzog, 2000; Muis, 2007; Winne & Hadwin, 1998). This process is referred to as metacognitive control (Nelson & Narens, 1990).

Aging and Metacognitive Monitoring

On a trial-by-trial basis, older adults generally demonstrate metacognitive monitoring equal to that of younger adults (Dunlosky, Baker, Rawson, & Hertzog, 2006; Dunlosky, Kubat-Silman, & Hertzog, 2003; Dunlosky & Connor, 1997; Hertzog, Dunlosky, & Sinclair, 2010; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Hertzog, Sinclair, & Dunlosky, 2010; Hines, Hertzog, & Touron, 2015; Kuhlmann & Touron, 2011). By contrast, older adults sometimes struggle to use this information when estimating performance across trials. This may result from an aggregation failure. For example, on a learning task, older adults may fail to keep track of what proportion of items was correctly recalled, overall or for a given strategy. For example, Price and colleagues (2008) had participants rate the effectiveness of an interactive imagery strategy (a normatively effective strategy) and rote repetition strategy (a normatively less effective strategy) before and after studying novel word pairs using the two strategies. Immediately after studying each item, participants rated the likelihood that they would recall it on a future test. Younger and older adults initially rated the strategies as equally effective. When rating each item during study, older and younger adults rated the items studied using interactive imagery as being more likely to be remembered than those studied with rote repetition. That is, older and younger adults both showed accurate monitoring at the item level. However, when asked at the end of study to rate the effectiveness of each strategy, only younger adults correctly updated these strategy beliefs (by rating interactive imagery as more effective).

Even when accurate monitoring does occur, people may fail to exert metacognitive control, and continuing using the ineffective strategies despite their updated knowledge. For example, people may realize while studying that a particular strategy isn’t working, or may realize after taking a test that their accuracy was below the desired level. However, they are only exerting metacognitive control if they alter their strategic approach as a result (e.g., altering encoding strategies, study time allocation, or goals). Older adults commonly display impairments in metacognitive control, by not altering their strategies despite accurate performance monitoring (Dunlosky et al., 2006, 2003; Dunlosky & Connor, 1997; Hertzog, Dunlosky, et al., 2010; Hertzog et al., 2002; Hertzog, Sinclair, et al., 2010; Kuhlmann & Touron, 2011). Given older adults' failures to update and utilize new metacognitive information about strategies, it is possible that older adults may also struggle to update or utilize their task representations as well.

While monitoring performance, one may discover that their fundamental task representation is inaccurate (Gaschler & Frensch, 2007; Haider & Frensch, 1996, 1999). For example, one may expect a multiple choice test but be given a free recall test instead. Or one may discover that the instructions they were given were not entirely accurate or that certain task information is irrelevant (Gaschler & Frensch, 2007; Haider & Frensch, 1996, 1999). To the extent that one can identify inconsistencies between their representation of the task and the actual task, they should seek to correct those inconsistencies and use the updated knowledge to alter their strategic approach (Lovett & Schunn, 1999). To date, no study has specifically attempted to measure older adults' task representations. Thus, it is unknown whether older adults enter tasks with accurate task representations or if they update those task representations and utilize them to improve strategic choices. Furthermore, we do not know whether older adults' abilities to form, monitor, and update task representations are equivalent to those of younger adults.

Although studies have not directly compared older and younger adults abilities to form and update task representations, data from eye-tracking and skill acquisition studies provide some support for the hypothesis. Older adults are found to fixate task-irrelevant information to a greater degree than do younger adults (Mitzner, Touron, Rogers, & Hertzog, 2010; Spieler, Mayr, & LaGrone, 2006; Touron et al., 2012), suggesting the possibility that they do not fully understand the task. For example, in a task where participants had to search an array of unrelated word pairs (e.g., ivy-bird, potato-frog) to determine whether a target word pair occurred in the array (Mitzner et al., 2010; Touron et al., 2012). For example, the words “ivy” and “bird” are paired in the array, as are “potato” and “frog.” If the target pair was ivy-bird, the participant would search the array and respond “yes” when they located ivy-bird Because each word occurred only once in the array, a participant could respond to a rearranged pair (e.g., ivy-frog) as soon as they found either of the words in the array (e.g., after locating the pair ivy-bird or potato-frog). However, on early trials, older and younger adults often continued to search the array even after one of the target words (e.g., ivy) had been located. Participants may not have correctly understood that the task possessed a single target, and thus continued to search for additional targets among the distractors. That is, they continued to search for ivy-frog even after locating ivy-bird in the array. After a few trials, younger but not older adults ceased this behavior, terminating their search after visually fixating one of the target words. By contrast, older adults often continued to search the entire array despite this behavior being maladaptive. That is, older adults failed to update their task representation and adjust their strategic process accordingly. The current studies attempt to measure task understanding without reliance on critical errors, to determine whether these persistent task misrepresentations are a common occurrence among older adults.

Limitations of Current Understanding and Methods

Metacognitive ratings (e.g., performance predictions) and questionnaires have been used to examine age differences in beliefs about strategies (Brigham & Pressley, 1988; Dunlosky et al., 2003; Frank et al., 2013; Hertzog, Sinclair, et al., 2010; Price et al., 2008; Touron & Hertzog, 2004a), and critical task errors have suggested specific task misunderstandings (Mitzner et al., 2010; Touron et al., 2012). However, neither of these methods is well suited for measuring overall task representation accuracy.

Metacognitive performance predictions can provide insight into faulty strategic beliefs (e.g., thinking a strategy faster or more accurate than it really is). But performance predictions will not indicate when a participant has faulty beliefs about the structure of the task. For example, performance predictions will capture when a participant is expecting a recognition test but is given a recall test, or when a participant is expecting an immediate test but is given a distractor task between study and recall. By contrast, questionnaires asking participants to reflect on critical aspects of their task representation can provide more target information about task understanding. For example, you can ask directly what kind of test a participant expects, whether something is important to a task (e.g., being fast, being accurate). However, these too have drawbacks. First, questionnaires invariably reflect what the experimenter expects the participant's task representation to contain—they force the researcher's representation space upon the participant (Rouse & Morris, 1986; Rowe & Cooke, 1995). For example, a questionnaire may omit important aspects of the participant's task representation that the experimenter's task representation does not include (Rouse & Morris, 1986; Rowe & Cooke, 1995). This happens when the participant's task representation contains unnecessary or irrelevant information (commissions; Rouse & Morris, 1986; Rowe & Cooke, 1995). Questionnaires can also be reactive if they ask a participant to reflect on an aspect of the task that was not originally part of their task representation (omissions; Rouse & Morris, 1986; Rowe & Cooke, 1995). The reactive effects of questionnaires could be particularly problematic if the questionnaire is administered prior to task completion.[1] The participant may incorporate information from the questionnaire into their task representation,

[1] This criticism can be applied to metacognitive ratings as well. Making performance predictions may cause people to monitor their learning and performance to a greater degree than they ordinarily would. Likewise, asking for performance precisions may cause people to ponder *how* they know that they know something, and may even result in participants considering how manipulated task properties influence their learning.

and then uses that information to make different strategic decisions than they would have made had they not answered the questionnaire.

Lastly, critical task errors may reflect an inaccurate task representation (Gaschler & Frensch, 2007; Haider & Frensch, 1996, 1999; Mitzner et al., 2010; Touron et al., 2012). However these errors may also reflect an inability to utilize one's task representation to make appropriate strategic choices, and cannot identify task misrepresentations when those do not result in obvious task errors. Thus, critical error analyses may under estimate task representation errors.

In contrast to the questionnaires, metacognitive ratings, and critical errors which tap specific aspects of task representations or relevant knowledge (e.g., how strategies impact performance accuracy), the mental model framework utilizes techniques designed to capture a broad measure of conceptual knowledge, and how specific concepts within a participant's conceptual knowledge relate to one another (Goldsmith, Johnson, & Acton, 1991; Novak & Cañas, 2006; Novak, 1990; Rouse & Morris, 1986; Rowe & Cooke, 1995; Staggers & Norica, 1993). If we think of task representations as a form of task-specific conceptual knowledge, these techniques may allow us to examine global task representations.

The Mental Model Framework

While the metacognitive model of strategy choice suggest a clear role for task-relevant knowledge and understanding (Bromme et al., 2009; Lovett & Schunn, 1999; Muis, 2007; Stahl, Pieschl, & Bromme, 2006; Winne & Hadwin, 1998), studies have not measured task representations directly. By contrast, the mental model approach used in the human factors and education literatures elicits more extensive information about people's conceptual knowledge (Goldsmith et al., 1991; Novak & Cañas, 2006; Novak, 1990; Rouse & Morris, 1986; Rowe & Cooke, 1995; Staggers & Norica, 1993).

Within the education and human factors literatures, various terms have been used to describe internal mental representations including: schema, knowledge structure, task representation, situation model, concept map, conceptual framework, and belief (Dorsey, Campbell, Foster, & Miles, 1999; Doyle & Ford, 1998; Rouse & Morris, 1986; Staggers & Norica, 1993). Due to the complexity of internal representations, these terms are particularly difficult to define, resulting in some dispute over the proper definition (for a through treatment, see Doyle & Ford, 1999; Rouse & Morris, 1986). Definitions typically describe and measure mental models as either complex networks of stored knowledge or as fleeting visuospatial representations in working memory. Schnotz and Preuss (1997) differentiate between these two categories using the terms conceptual knowledge and mental models, respectively. They argue that mental models are actually temporary representations *derived from* conceptual knowledge and task cues (Schnotz & Preuss, 1997; see also Seel, 2001; Wilson & Rutherford, 1989). When discussing the relevant literature, I will adopt this terminology, with conceptual knowledge referring to stable long-term memory representations and mental models referring to conceptual knowledge held temporarily within the focus of attention (i.e., in working memory).

Conceptual Knowledge

Conceptual knowledge has been described as a web of interconnected knowledge (Gadgil, Nokes-Malach, & Chi, 2012) or knowledge structure (Dorsey et al., 1999), and includes all levels of knowledge specificity. General schematic knowledge and task-specific representations are both examples of conceptual knowledge. Conceptual knowledge contains not only declarative knowledge, but also the interconnections, and in particular, the causal links between those pieces of declarative knowledge (Gadgil et al., 2012; Van Boven & Thompson, 2003). Research indicates that it is these relationships between pieces of information that predict success on domain-relevant tasks (Goldsmith et al., 1991; Novak & Cañas, 2006; Novak, 1990;

Staggers & Norica, 1993). These findings suggest that declarative knowledge may be a poor indicator of task representation accuracy, whereas the structural relationships between task-relevant concepts may be far more diagnostic of, and critical to, task performance. The next sections describe a method for measuring conceptual knowledge, followed by a discussion of how conceptual knowledge and mental models are formed, utilized, and updated, considering potential age differences at each step.

Measuring Conceptual Knowledge

Concept mapping is perhaps the most popular way to measure conceptual knowledge (Novak & Cañas, 2006). A concept map is a series of nodes (concepts) connected by links representing how those concepts are connected in a participant's knowledge structure (Novak & Cañas, 2006; Novak, 1990). For example, in Figure 1, p. 12, "Bunsen burner," "temperature," and "bacteria" are examples of nodes, whereas "increases" and "decreases" are links.

One way to generate a concept map is to use pair-wise relatedness ratings of domain-relevant terms provided by the researcher (Capelo & Dias, 2009; Clariana & Taricani, 2010; Curtis & Davis, 2003; d'Apoilonia, Charles, & Gary, 2004; Dorsey et al., 1999; Goldsmith et al., 1991; Gomez, Hadfield, & Housner, 1996; Gonzalvo, Cafias, & Bajo, 1994; Johnson, Goldsmith, & Teague, 1994; Kim, 2012; Van Boven & Thompson, 2003) or in rare cases, by the participant (Rowe & Cooke, 1995). This technique has the participant rate the relatedness of all possible pairs of terms (Figure 2, p. 13). Participant pair-wise ratings are then subjected to a pathfinder algorithm. The pathfinder algorithm is a data reduction technique that transforms the matrix of pair-wise relatedness ratings into a concept map. This is done by maintaining strong links between nodes and trimming off weaker links. For example, consider Figure 1 (p. 12). The pathfinder derived ratings might indicate strong relationships between Bunsen burner and temperature and between temperature and bacteria, but a weaker relationship between Bunsen burner and bacteria. Thus, the resulting strong indirect relationship between Bunsen burner and bacteria, through the concept “temperature” would be retained, and the weaker direct relationship would be eliminated.

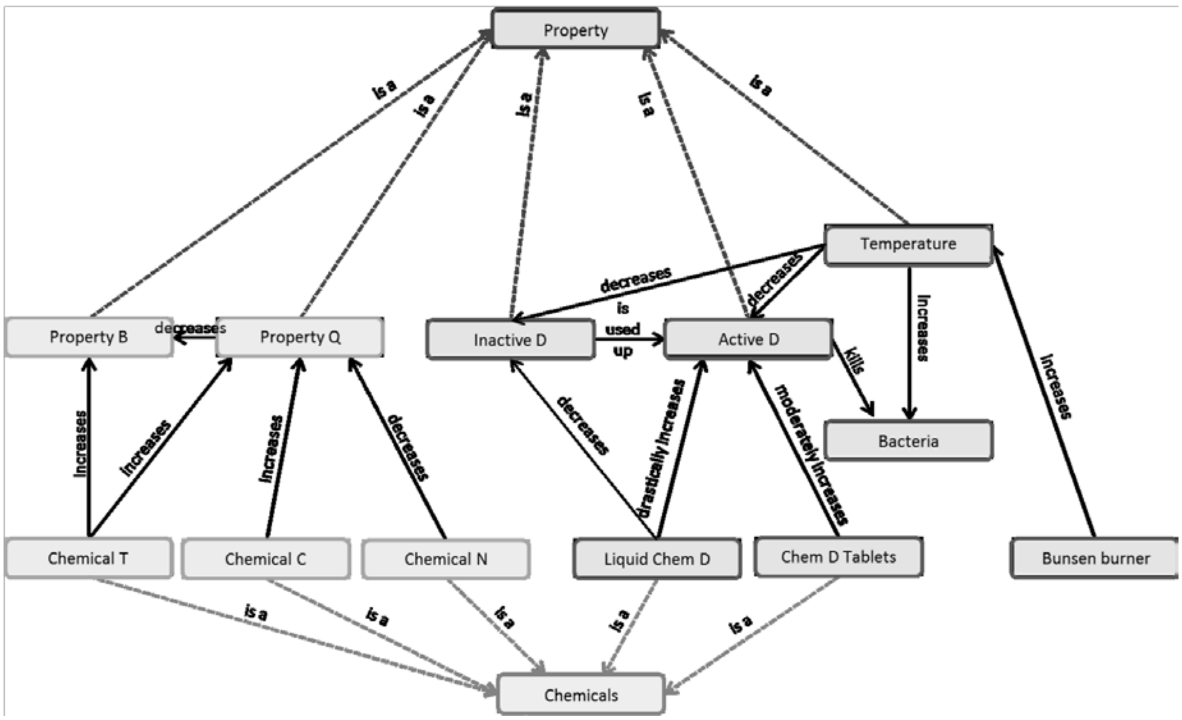


Figure 1.Study 2 Reference Map.

Pathfinder derived concept maps are then compared to a reference map created by an expert or group of experts. The primary advantage of this technique is that the purpose of the task is less transparent, and all relationships are considered, eliminating the bias that occurs when the experimenter selects which relationships to ask about via metacognitive ratings or questionnaires (see Chiu, Chou, & Liu, 2002; Dorsey et al., 1999; Jacobs-lawson & Hershey, 1994; Kim, 2012; Markham, Mintzes, & Jones, 1994; Novak & Cañas, 2006; Poindexter & Clariana, 2006; Rice, Ryan, & Samson, 1998; Rowe & Cooke, 1995; Seel, 2001; Taricani & Clariana, 2006; Wallace & Mintzes, 1990).

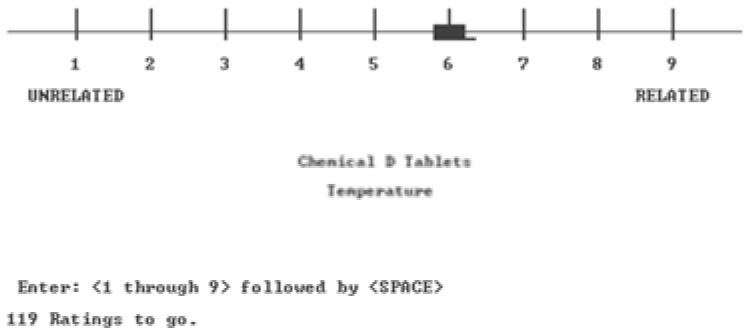


Figure 2. Screenshot of Rate Software used in Study 1.

When concept maps are elicited using pair-wise ratings, accuracy is typically measured as the proportion of overlap between a given concept map and the reference map known as "closeness" or "C" scores (Johnson & Goldsmith, 1994). Thus, C scores range from 0 (no overlap) to 1 (identical concept maps; for a detailed explanation of the pathfinder algorithm and calculations for C, see Goldsmith et al., 1991). C scores from relevant concept maps predict classroom performance in domains such as: history of psychology (Gonzalvo et al., 1994), research methods and statistics (Goldsmith et al., 1991), biology (d'Apoilonia et al., 2004), accounting (Curtis & Davis, 2003), and teaching of elementary mathematics (Gomez et al., 1996). C scores also predict performance outside the classroom in domains such as: accounting job performance (Curtis & Davis, 2003), teaching of elementary mathematics job performance (Gomez et al., 1996), ACT math scores (Johnson et al., 1994), radar warning systems troubleshooting performance (Rowe & Cooke, 1995), and negotiation performance (Van Boven & Thompson, 2003).

Unlike questionnaires which ask about specific relationships deemed important by the researcher, pathfinder-derived concept maps allow a participant to specify any number of relationships among concepts (nodes; Rouse & Morris, 1986; Rowe & Cooke, 1995). That is, pathfinder-derived concept maps do not force the researcher's conceptualization upon the participant to the extent that questionnaires do (Rouse & Morris, 1986; Rowe & Cooke, 1995).

Conceptual Knowledge Formation

According to the mental model perspective, new conceptual knowledge utilizes existing schemas and conceptual knowledge as a starting point, and modifies them accordingly (Schnotz & Preuss, 1997). For example, when presented with a memory task in a laboratory, a participant not only references their beliefs about memory and study strategies, but also uses prior learning experiences to assume—by way of analogy—a certain structure to the task itself (e.g., study proceeded by test; note that these processes need not be intentional or conscious). Task instructions, cues, and experience are then used to modify that general schema into a task-specific representation. Thus, conceptual knowledge formation relies on schematic knowledge, beliefs, monitoring, metacognitive control, and analogical processing.

Because older adults possess greater crystalized knowledge, they should have more task structures and schemas to choose from when seeking a starting model for a new task. Indeed, older adults do demonstrate greater learning when new information is consistent with or can be incorporated into existing knowledge (Castel, 2005, 2007; Stine-Morrow, Miller, & Hertzog, 2006), however these studies focused on incorporating only new pieces of declarative information and not the acquisition of new *structured* conceptual knowledge

Conceptual Knowledge Updating

Metacognitive models of strategy selection suggest that task representations and strategic choices are updated in response to external feedback and participant-identified inaccuracies in task representation (monitoring; Bromme et al., 2009; Lovett & Schunn, 1999; Muis, 2007; Winne & Hadwin, 1998). Likewise, the mental model and conceptual knowledge framework also suggests that conceptual knowledge is updated, but through three specific processes: accretion, tuning, or reorganization (Schnotz & Preuss, 1997). Accretion is when new knowledge is simply added to the existing knowledge and could be thought of as adding a node and link to a concept map (Schnotz & Preuss, 1997). Tuning involves the changing of a single component within the structural knowledge from one value to another, such as changing a link label in a concept map (e.g., correcting the link atoms *are* molecules into atoms *make up* molecules; Schnotz & Preuss, 1997). Because unlearning is involved in tuning, Schnotz and Preuss suggest that it may be more resource demanding than accretion. Lastly, reorganization involves the altering of relationships within conceptual knowledge (Schnotz & Preuss, 1997). This will typically involve far greater changes to the conceptualization and may require even more resources than tuning or accretion, as the very structure of the relationships involved are altered (Schnotz & Preuss, 1997).

Studies on strategy knowledge updating in younger and older adults have looked exclusively at changes akin to accretion and tuning (Bieman-Copland & Charness, 1994; Brigham & Pressley, 1988; Devolder & Pressley, 1992; Matvey, Dunlosky, Shaw, Parks, & Hertzog, 2002; Price, Hertzog, & Dunlosky, 2010). For example, in the study by Price and colleagues (2008) described earlier, younger but not older adults updated their strategic knowledge to reflect the actual effectiveness of two memory encoding strategies. This could be thought of as a tuning of the effectiveness of each strategy. By contrast, it could be argued that the absence of pre-study differences in strategy effectiveness ratings suggests that strategy effectiveness was not part of older or younger adults' conceptual knowledge at all, and thus adding that information should instead involve accretion.

In studies of strategy knowledge updating, both younger and older adults demonstrate difficulty in updating their strategic knowledge, with older adults struggling more than young (Bieman-Copland & Charness, 1994; Brigham & Pressley, 1988; Devolder & Pressley, 1992; Matvey et al., 2002; Price et al., 2010). Given the age differences in updating for what Schnotz and Preuss claim is a less resource demanding type of updating, older adults may show even greater decrements when the underlying structure of their conceptual knowledge needs to be updated. To date no studies have examined this process in older adults. One potential prediction is that older adults may make minor adjustments to their conceptual knowledge (which could be evidenced by increased overlap between their concept maps and those of experts—increased C scores), but may show less updating compared to younger adults. This might be particularly problematic when their concept maps are initially highly inaccurate, and thus greater structural changes are necessary.

Conceptual Knowledge Utilization: Mental Models

In contrast to the conceptual knowledge stored in long-term memory, a mental model is a mental representation of a given system at an exact moment in time (Besnard, Greathead, & Baxter, 2004; Van Boven & Thompson, 2003). Due to the constraints of working memory, people cannot consider the entirety of their conceptual knowledge when solving a problem or making a decision (Besnard et al., 2004; Schnotz & Preuss, 1997; Seel, 2001; Van Boven & Thompson, 2003)[2]. As a result, when a person is faced with a problem, they extract from their conceptual knowledge the necessary elements (and only those necessary elements) to form a mental model of the current situation (Chiou & Anderson, 2010; Kim, 2012; Schnotz & Preuss, 1997; Seel, 2001). This mental model is then used to solve the problem or make a decision (Chiou & Anderson, 2010; Kim, 2012; Schnotz & Preuss, 1997; Seel, 2001). Thus, strategic decisions, like all decisions and problem solving (Johnson-Laird, 1983), should occur at the mental model level.

If age differences occur in conceptual knowledge formation, monitoring, or updating, they will influence the mental models used when making strategic choices (Schnotz & Preuss,

[2] Note that this is also consistent with Anderson's ACT-R model (Anderson, Matessa, & Lebiere, 1997).

1997). Likewise, if age differences occur in the ability to form mental models out of their conceptual knowledge, these differences could influence strategic choices. Unfortunately, measuring mental models is difficult because mental models are fleeting and difficult for the user to describe (Besnard et al., 2004; Van Boven & Thompson, 2003). As a result, the literature on mental models often uses highly controlled situations where mental models can be inferred from behavioral outcomes (such as common errors in reasoning; Besnard, Greathead, & Baxter, 2004; Johnson-Laird, 1983; Van Boven & Thompson, 2003). This approach is not feasible for some tasks. However, a failure to make appropriate strategic choices when conceptual knowledge is accurate would be an indirect indicator of a potential mental model formation or utilization failure.

Because mental models are limited by working memory constraints (Besnard et al., 2004; Schnotz & Preuss, 1997; Seel, 2001; Van Boven & Thompson, 2003), older adults' reduced working memory capacity (Bopp & Verhaeghen, 2007; Hasher & Zacks, 1988) could result in impoverished mental models and thus conceptual knowledge utilization. That is, older adults may struggle to maintain more complex mental models in working memory, forcing older adults to occasionally make decisions based on less information than younger adults. Supporting this view, age differences in mental model utilization appear when multiple mental models are under consideration at a once (Copeland & Radvansky, 2007; Gilinsky & Judd, 1994; Radvansky, Zacks, & Hasher, 1996) and older adults often do consider fewer pieces of information when making decisions (Blanchard-Fields, Hertzog, & Horhota, 2012; Chen & Blanchard-Fields, 2000; Hess, Follett, & McGee, 1998; Klaczynski & Robinson, 2000; but see Hines, Hertzog, & Touron, 2015 for evidence that older adults consider as many cues as younger adults do when making metacongnitive predictions).

However, when a single mental model is necessary, age differences in mental model formation and utilization are rarely reported, with some studies suggesting that older adults rely more on mental models more than do younger adults (Gilbert, Rogers, & Samuelson, 2004; Morrow, Leirer, Altieri, & Fitzsimmons, 1994a; Radvansky, Copeland, Berish, & Dijkstra, 2003; Radvansky, Gerard, Zacks, & Hasher, 1990; Radvansky & Dijkstra, 2007; Radvansky, 1999b; Stine-Morrow, Gagne, Morrow, & DeWall, 2004; Stine-Morrow, Morrow, & Leno, 2002). It is important to note that these studies focused exclusively on mental models of narratives or single sentences. Narrative and text processing are two highly practiced domains for most adults. Additionally, although some narratives require a substantial degree of mental model updating, understanding a complex new task may require a greater degree of updating and manipulation. Thus mental model updating and manipulation in complex novel tasks may tax working memory to the point where age differences in mental model utilization appear.

Current Studies

In two studies we examine whether concept mapping can be used to measure younger and older adults' task representations and whether these task representation influences strategic choices in younger and older adults. In the second study we also examine whether age differences in the ability to update task representation explain age differences in the ability to improve strategic choices following practice.

CHAPTER II

STUDY 1

The goal of Study 1 was to determine whether pathfinder derived concept maps could be used to measure task representations in younger and older adults. To date, no study had used concept mapping to measure *task* representations, nor had any study used concept mapping with a population of older adults. To minimize the influence of prior knowledge and beliefs on task representation formation and utilization, a novel chemistry task with a well-defined task structure (analogous to managing swimming pool chemistry) was created. We were concerned that general age-related decrements in learning (e.g., associative binding deficits; Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008) could produce age differences in underlying declarative knowledge about the task. This was especially concerning given the novelty of the task. If older adults fail to learn the basic terminology of the task, then they will be unable to convey the relatedness of those terms via the pathfinder ratings. To combat this, we used a series of guided learning quizzes and a cumulative criterion test during the task instructions. Although the criterion test required a minimum level of declarative knowledge, previous studies show that such tests are poor predictors of concept map scores (Goldsmith et al., 1991; Novak & Cañas, 2006; Novak, 1990; Staggers & Norica, 1993). Thus, variability in task representations (C scores) were still expected despite requiring a minimum level of declarative knowledge for all participants.

Methods

Participants

Twenty younger adults aged 18-22 (*M* = 18.81) and 31 older adults aged 61-85 (*M* = 70.30) participated in the study. Of the 31 older adults, 11 were excluded for performance reasons (described later). This produced a final older adult sample of 20 participants aged 61-77 (M = 69.49). Younger adults participated for course credit whereas older adults received roughly $10 per hour for participation. Participants were in relatively good health and none reported ever having suffered a major seizure or stroke. All participants had corrected near visual acuity of 20/50 or better. Participant demographics can be found in Table 1.

Table 1

Study 1 Demographic and Performance Data

Demographics	Young	*SD*	Old	*SD*	Age dif *d*
N	20		20		
Age	18.81	1.81	69.49	4.32	
Education	12.51	0.86	15.49	2.48	+1.61**
Medications	0.57	0.90	2.51	1.64	+1.47**
Processing speed	38.62	7.81	27.80	4.89	-1.66**
Vocabulary	13.45	3.12	22.75	7.26	+1.66**
Performance	**Young**	***SD***	**Old**	***SD***	**Age dif *d***
Cumulative test	0.85	0.13	0.75	0.11	-0.83**
C score	0.26	0.09	0.29	0.09	+0.33
Attempts to criterion	1.67	0.91	2.05	0.94	+0.41
Inference test	0.59	0.16	0.56	0.17	-0.18

Note. Young = young adult means; Old = older adult means; Processing speed = number correct on out of 30 on Salthouse's pattern comparison task (1993); Vocabulary = number correct out of 36 on the Advanced Vocabulary Test (Ekstrom, French, & Harman, 1976); Cumulative test = proportion correct on a 15 question, 9-choice, multiple choice cumulative instructions test measuring declarative knowledge of the instructions; C score = proportion of concept map overlap between the participant and reference models ranging from 0 (no overlap) to 1 (perfect overlap); Attempts to reach criterion = number of attempts taken to pass the cumulative instructions test; Inference test = proportion correct on a 19 question, 4-choice, multiple choice test inference test measuring the ability to apply the information from the instructions. *p < .05. **p < .01.

Materials

A novel chemistry task was created for this study. The task structure is perfectly analogous to the daily management of a swimming pool. However, the task is disguised as a novel chemistry task in which the participant maintains a chemical solution in a beaker rather than a pool (Appendix B). All of the terms normally associated with a swimming pool were disguised by replacing the term with a single letter (Appendix C).

The swimming pool task was chosen for a number of reasons. First, the task is difficult and complex, but clearly manageable (people *can* learn to manage a swimming pool). Second, the relationships between properties of a swimming pool vary in complexity. For example, it is rather straightforward that adding chlorine tablets and liquid chlorine will boost chlorine levels. By contrast, managing pH is more complex. First, the pH of a swimming pool cannot be stabilized without an optimal level of alkalinity. Thus, you can only raise the pH to the desired level after first adjusting the alkalinity. Second, in order lower alkalinity, you must lower pH to a very acidic level which is not recommended for swimming, and then raise the pH back into the ideal range after the alkalinity level is in the desired range. Thus, some aspects of the task (like chlorine management) allow for a fairly rudimentary understanding, whereas others (like pH and alkalinity management) require an integrated understanding of how various chemical properties interact. For the purposes of Study 1, participants only learned how to do the chemical task. They did not actually perform the task.

Pathfinder ratings were collected using the Rate software available from Interlink, Inc., Las Cruces, NM. The Rate software first shows all pairs of words that will be used for the task. Two words are then shown in the center of the screen and the participant indicates via keypress how related they think those terms are (Figure 2, p. 13).

Additionally, two tests were created. The first was a 15 question, 9-choice multiple choice test cumulative instructions test, designed to capture knowledge for explicitly stated relationships from the instructional text (declarative knowledge; Appendix D). The second was a 19 question, 4-choice multiple choice inference test relevant to the novel chemistry task, designed to capture the ability to apply the knowledge about the chemistry task (Appendices D and E).

Procedures

Participants first completed a demographics questionnaire followed by a near visual acuity test, pattern comparison, and vocabulary tests. Data from these measures are presented in Table 1 (p. 19-20).

For the learning phase, participants were told that they would play the role of chemists and had to manage a solution in a beaker. Participants then read instructions about how the novel chemistry task would work and were asked to learn the basic information about the chemicals (e.g., learn that Liquid D and Chemical D Tablets were used to increase Active D in the solution). To guide and aid the participants' learning they were given a series of "open book" mini-quizzes throughout the learning phase. That is, participants were free to switch between the quiz and chemical task instructions while taking the mini-quizzes. If a participant did not get 100% accuracy on one of these quizzes, they were asked to restudy the text and retry the mini-quiz. Participants were given feedback on which items were correct and incorrect, but were told to scan the text for the correct answers to the items they missed. The mini-quiz items were taken directly from the cumulative instructions test.

After obtaining 100% accuracy on a mini-quiz, the participant was permitted to continue on to the next part of the instructional text. In total there were 15 mini-quiz questions, seven relating to the properties analogous to Chlorine and eight relating to the properties analogous to pH and alkalinity. After passing all mini-quizzes they completed a cumulative test containing the same 15 items as the mini-quizzes. In order to pass the cumulative test, the participant had to correctly answer at minimum four of the seven "chlorine questions" and four of the eight "pH and alkalinity questions." If a participant did not meet this minimum criterion they were taken back to the start of the learning phase and asked to restudy the instructional text and their answers to the mini quizzes.[3] Participants were informed that this phase would take up a substantial portion of the experiment and that most people had to reread the information several times.

Next, participants completed the pathfinder ratings for 19 relevant terms (Appendix F), followed by the inference test (Table 1, p. 19-20), and a post-task questionnaire (Table 2, p. 23-25). The inference test contained 22, four-choice multiple choice responses. This test was designed to test a participant's ability to use the information they gained from the instructional test in ways that would be similar to performing the actual chemistry task. The post-task questionnaire asked them to make a series of 1-5 ratings regarding their task experience, and indicate any terms they felt were omitted from the ratings task.[4] Additionally the post-task questionnaire asked, "What did this task make you think of?" and "Have you ever owned or managed a swimming pool before?" No participant indicated that the task reminded them of managing a swimming pool.[5]

---

[3] Initial piloting did not include the mini-quizzes and included minimal feedback on criterion test performance. Under these initial conditions most participants did not reach the criterion and all voiced considerable frustration with the task. Adding the mini-quizzes and increasing the degree to feedback improved criterion test performance and alleviated frustration resulting in improved compliance.

[4] No participant suggested any additional terms for the ratings task.

[5] Participants having owned or managed a swimming pool did not perform differently on any measure relative to those not having owned or managed a pool. When Table 1 age differences are examined using Age X Ownership ANOVAs instead of *t*-tests, the pattern of significance is unchanged with no main effects of ownership and no Age X Ownership interactions (all *F*s < 1).

Table 2

Post-task Questionnaire Data

Study 1

Phase		Young	*SD*	Old	*SD*	Age dif *d*
Instructions	Learning difficulty	3.67	1.02	4.60	0.75	+1.04**
	Reading difficulty	2.57	1.12	2.30	1.13	-0.24
Ratings	Difficulty	3.76	1.04	3.95	1.24	+0.17
	Tediousness	4.14	0.79	4.20	1.20	+0.06
Inference test	Difficulty	3.90	0.94	4.35	0.93	+0.48

Study 2

Phase		Young	*SD*	Old	*SD*	Age dif *d*
Instructions	Difficulty	6.89	1.88	7.73	1.34	+0.51*
	Tediousness	7.26	2.39	7.03	2.03	-0.10
	Motivation	3.03	2.37	7.43	1.91	+2.04**
	Effort	5.46	2.11	7.70	1.39	+1.25**
Cumulative test 1	Difficulty	6.46	2.78	8.07	0.94	+0.78**
	Tediousness	6.66	2.55	6.80	1.94	+0.06
	Motivation	4.20	2.47	7.87	1.61	+1.80**
	Effort	5.57	2.40	7.80	1.58	+1.10**
Cumulative test 2	Difficulty	5.63	2.31	7.90	1.42	+1.18**
	Tediousness	6.00	2.59	6.80	2.12	+0.34
	Motivation	4.51	2.64	7.60	1.98	+1.32**
	Effort	5.51	2.45	7.50	1.72	+0.94**
Ratings 1	Difficulty	6.60	2.61	7.90	1.35	+0.63*
	Tediousness	7.51	2.13	7.07	2.10	-0.21
	Motivation	3.94	2.76	7.83	1.60	+1.72**
	Effort	5.29	2.16	7.77	1.59	+1.31**
Ratings 2	Difficulty	5.66	2.65	7.77	1.43	+0.99**
	Tediousness	7.00	2.29	6.90	2.20	-0.04
	Motivation	4.03	2.54	7.73	1.78	+1.69**
	Effort	5.43	2.56	7.80	1.61	+1.11**
Chemistry Task	Difficulty	7.03	2.29	7.83	1.32	+0.43
	Tediousness	8.00	1.75	6.70	2.22	-0.65*
	Motivation	3.76	3.01	7.83	1.70	+1.67**
	Effort	5.46	2.96	7.77	1.59	+0.97**
Learning Missing information	Difficulty	6.51	2.55	7.47	1.48	+0.46**
	Tediousness	7.20	2.41	6.60	2.13	-0.26
	Motivation	4.31	2.69	7.73	1.70	+1.52**
	Effort	5.40	2.65	7.70	1.56	+1.06**
	Identify missing	5.71	2.46	6.63	1.96	+0.41
	Mental list	.59	-	.60	-	-

Recall missing information	4.83	2.22	5.80	2.34	+0.43
Prior chemistry knowledge	3.77	2.33	3.53	2.54	-0.10
Task understanding	3.80	2.35	4.47	2.43	+0.28

Note. Study 1 self-ratings of learning difficulty, difficulty reading the text, difficulty making relatedness ratings, tediousness of making the relatedness ratings, and difficulty of the inference tests are on 1 (not at all) to 5 (very much) Likert scales; prior knowledge of chemistry were self-rated from 1 (very poor) to 5 (Excellent); attempts to criterion indicates the number of attempts it took a participant to pass the learning criterion test. Study 2 self-ratings of difficulty, tediousness, motivation, effort, and ability to identify and recall which information was missing from the task instructions, are on 1 (not at all/very little/very easy) to 9 (very/as much as possible) Likert scales; Instructions = difficulty, tediousness, motivation, effort while trying to learn the information from the instructional text; Cumulative test 1 = difficulty, tediousness, motivation, effort on the pre-task cumulative instructions test; Cumulative test 2 = difficulty, tediousness, motivation, effort on the post-task cumulative instructions test; Ratings 1 = difficulty, tediousness, motivation, effort on the pre-task pathfinder ratings; Ratings 2 = difficulty, tediousness, motivation, effort on the post-task pathfinder ratings; Chemistry task = difficulty, tediousness, motivation, effort for performing the novel chemistry task; Learning missing information = difficulty, tediousness, motivation, effort for learning the missing information via performing the novel chemistry task. Prior chemistry knowledge, pre- and post-task chemical task understanding on 1 (very poor) to 9 (excellent) Likert scales; whether participants made a mental list of the missing information from the task is the percentage of participants indicating "yes." $^{*}p < .05$. $^{**}p < .01$.

Results

Eleven older adults required more than five attempts to pass the cumulative test at the end of the learning phase. In each case the experimenter had noted that the participant had switched to a strategy of rapidly advancing through the instructional text and guessing on the cumulative test until they reached the criterion. For this reason these eleven participants were replaced. After removing these participants, the age difference in the number of attempts required to reach criterion was not significant, $t(38) = 1.30$, $p = .202$.

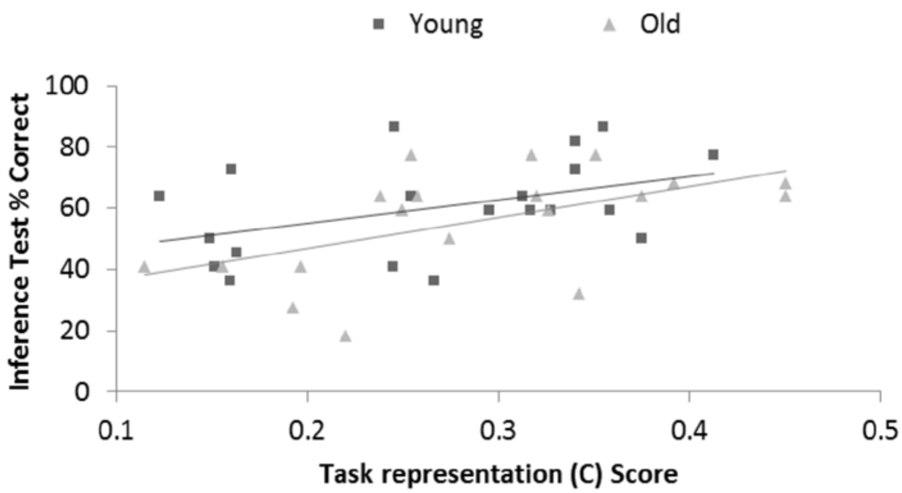


Figure 3. Relationship between C Scores and Inference Test Scores (Percent Correct).

Neither C scores, nor inference test scores, differed with age (Table 1, p. 19-20; Figure 3). However these measures were positively correlated overall, $r(38) = .52$, $p = .001$, and within younger, $r(18) = .54$, and older adults, $r(18) = .54$, (Figure 3, p. 25). The eleven older adults whom were excluded from analysis all had C scores similar to the lowest performing participants that were retained for analysis and inference test scores near chance. Therefore, we conclude that these participants' C scores accurately reflected their failure to learn any aspect of the task, but were likely driven by associative deficits or low motivation, which are not of interest to the current study.

Discussion

Despite all participants reaching a minimum level of declarative knowledge (as measured by the cumulative instructions test), participants demonstrated varied levels of task representation in both C scores and inference test scores. Furthermore, high C scores were associated with higher inference test performance. This is consistent with the mental model and conceptual knowledge frameworks which claim that the underlying structure of knowledge and the interrelations among ideas are more critical for knowledge utilization than is declarative knowledge. These findings also suggest that task representations can be measured using pathfinder-derived concept maps. Most importantly these results held for both younger and older adults, suggesting that pathfinder-derived concept maps can be used to measure conceptual knowledge in older adults. When older adults possessed accurate task representations they scored as highly on the inference test as did younger adults. Thus, we found no evidence for a utilization deficit among older adults. This is consistent with a general failure to find mental model utilization deficits in older adults in other domains, namely reading comprehension and narrative processing (Gilbert et al., 2004; Morrow et al., 1994a; Radvansky et al., 2003, 1990; Radvansky & Dijkstra, 2007; Stine-Morrow et al., 2004, 2002).

The difficulty of learning novel task relationships was much greater than anticipated. This was particularly true for older adults, 11 of whom were not able to learn the basic surface relationships. The time course for the learning phase of the study was also much longer than expected with older adults taking roughly an hour to pass the criterion test—twice as long as younger adults. These issues were taken in consideration when designing Study 2.

CHAPTER III

STUDY 2

Study 2 measured task representations both before and after completing the novel chemistry task to test: (1) whether initial task representations influence initial strategic choices in older and younger adults, (2) for the presence of age-related differences in task representation updating, (3) whether age-related differences in task representation updating account for older adults' continued use of suboptimal strategies.

Methods

Participants

Thirty-five younger (aged 18-25) and 30 older adults (aged 60-75) were tested.[6] Younger adults participated for course credit and older adults received roughly $10 per hour for participation. Participants were in relatively good health and none reported ever having suffered a major seizure or stroke. All participants had corrected near visual acuity of 20/50 or better. Participant demographics can be found in Table 3.

Materials

Pretests, learning, and pathfinder rating materials. The same pretests used in Study 1 were used for Study 2. Pilot studies revealed that the learning quizzes and cumulative instructions test criterion used in Study 1 produced near ceiling performance on the novel chemistry task for many young adults. Therefore the novel chemistry task instructions were presented without any guided learning quizzes. The cumulative instructions test was retained, but the criterion for

[6] The researchers noted that adults over 75 were subjectively more likely express difficulty and frustration in Study 1. Thus the age range was reduced to 60-75 for Study 2.

passing this test was eliminated (participants continued regardless of test performance) and three items were removed from the quiz because they tested information that was not specifically relevant to managing the chemistry task (e.g., whether a property being too high caused corrosion or toxic fumes). Additionally, to ensure that all participants would enter the novel chemistry task with less than perfect task representations, some information was removed from the novel chemistry task instructions. Participants were explicitly informed that the information was incomplete and told that part of their job during the task was to determine what the missing relationships were. The missing information included which chemical properties were affected by temperature (Active D and Inactive D evaporate more quickly at high temperatures) and how chemicals C, T, and N would impact Properties B and Q (C increases Property Q, T increases Properties B and Q, and N decreases Property Q; see Appendix G). While understanding the effects of temperature could aid in the management of Active D and Inactive D, these properties could be reasonably managed without this knowledge. By contrast, effectively managing Properties B and Q is impossible without first learning the missing information.

Table 3

Study 2 Demographic and Performance Data

Demographics	Young	*SD*	Old	*SD*	Age difference *d*
N	35		30		
Age	18.54	0.85	70.90	3.35	
Education	12.40	0.69	16.45	1.82	+2.94**
Medications	0.54	0.82	2.13	1.98	+1.05**
Processing speed	38.89	7.46	27.76	5.34	-1.72**
Vocabulary	15.60	4.36	24.70	6.01	+1.73**
Performance	**Young**	***SD***	**Old**	***SD***	**Age difference *d***
Cumulative test 1	.42	.25	.47	.23	+0.21
C score 1	.25	.09	.30	.12	+0.47
Chemistry task	-2.59	25.88	3.01	20.69	+0.23
Properties B & Q	3.84	21.08	-2.59	21.31	-0.30
C score 2	.25	.12	.29	.07	+0.41
Cumulative test 2	.45	.31	.42	.25	-0.11

Note. Young = young adult means; Old = older adult means; Processing speed = number correct on out of 30 on Salthouse's pattern comparison task (1993); Vocabulary = number correct out of 36 on the Advanced Vocabulary Test (Ekstrom et al., 1976); Cumulative test 1 = proportion correct on a 15 question, 9-choice, multiple choice, pre-task cumulative instructions test measuring declarative knowledge of the instructions; C score 1 = pre-task proportion of concept map overlap between the participant and reference models ranging from 0 (no overlap) to 1 (perfect overlap); Chemistry task = Corrected performance on the chemistry task, where the mean performance is zero and scores above zero are above average and scores below zero are below average; Properties B & Q = Corrected performance scores on the chemistry task when restricting the data to only performance on Properties B and Q, the sub-set of the chemistry task that requires structural updating for high performance; C score 2 = post-task proportion of concept map overlap ; Cumulative test 2 = proportion correct on a 15 question, 9-choice, multiple choice, post-task cumulative instructions test measuring declarative knowledge of the instructions. *p < .05. **p < .01.

Novel chemistry task. Participants in Study 2 not only learned about, but also performed the novel chemistry task. The task was programmed in E-Prime 2, and featured five solution properties which need to be maintained within an ideal range. The task trials were presented as "days" to the participant. For each trial the participant received a report indicating the current level of each solution property, and the ideal range for each property. The participant then had the option to add various doses of chemicals or adjust the setting of the Bunsen burner. The goal of the task was to use these options to keep each solution property as close as possible to the center of its ideal range. After selecting the dose for each chemical additive, the participant clicked a submit button. The display then showed how the chemical additives changed the solution properties. The participant then clicked on an "advance day" button which started the next trial. Solution properties then changed overnight in accordance with how solution properties were described in the instructions (e.g., temperature and Active D [free chlorine] decrease over night). At the end of 7 days the participant received a new chemical solution. Each of these chemical solutions had a different problem starting state, meaning different solution properties would be outside the ideal range. Thus, the each participant had to "fix" different problems with the solution. Poor strategic choices (adding the wrong chemicals) could create further problems as well, whereas proper strategic choices would eliminate problems and bring the solution properties into their ideal ranges. However, even with ideal strategic choice, some problems take multiple days to correct.

There were 18 different problem starting states, each of which occurred once during the first half of the chemistry task and once during the second half of the chemistry task. The second occurrence of each problem state was altered slightly as to not be identical to the first occurrence, but required the same strategic choices. Thus there were 252 "days" in all, which constituted 36 different "trials" each with its own starting state. The order of the starting states within each half of the chemistry task was randomized. Participants were informed that they would receive a new solution every seven days. In addition to displaying the total number of 252 days they had completed, the display also counted down the number of days until a new solution would be given (Figure 4).

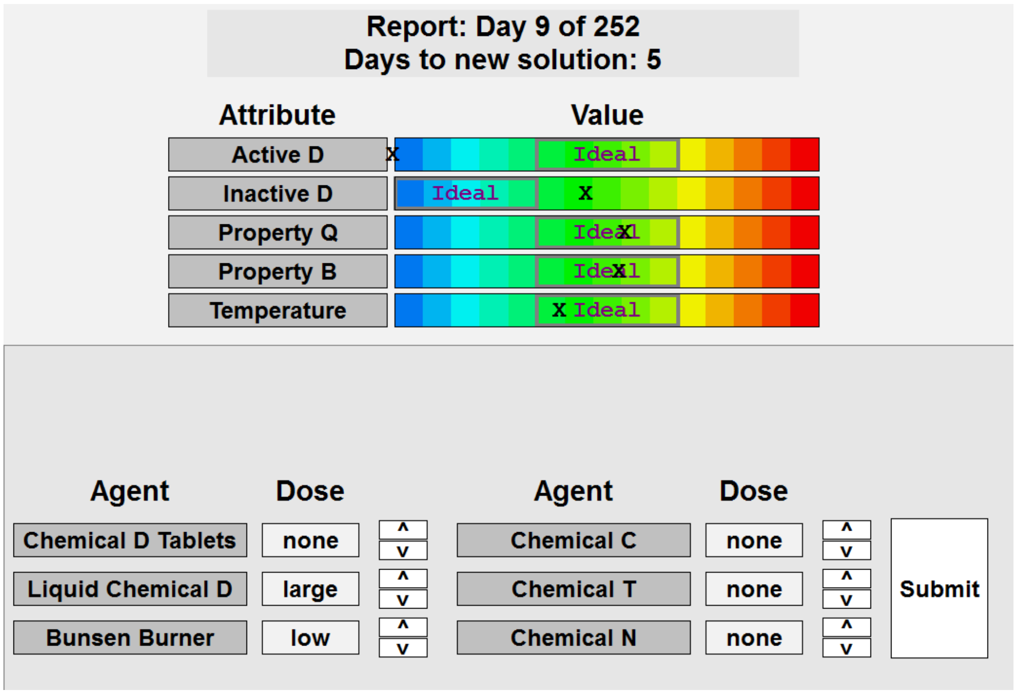


Figure 4. Chemistry Task Interface.

Scoring. Values for each solution property ranged from 0 to 438. The ideal range for each solution property was between 146 and 292, with the exception of Inactive D which is ideally kept at zero, and thus has an ideal range from zero to 146 (see Table 4 for an example, p. 32). First, solution property values were averaged across the seven days for each trial. Next, for each property a (absolute value) deviation score was computed from the mid-point of the ideal range, referred to as the “ideal value.” Again, the exception is Inactive D, where ideal value is zero. Thus, for each property, the deviation scores range 0 to 219, except for Inactive D, which ranges from 0 to 438. However, deviation scores for Inactive D were rarely greater than 219. In all cases higher deviation corresponded to worse performance. For example, in Table 4 (p. 32)

Property B is at 286. This value is within the ideal range ($146 < 286 < 292$), but is 67 points above the ideal value. Thus, the deviation score for Property B is 67, indicating good but not excellent management of Property B. By contrast, Active D is far above the ideal value and outside the ideal range, earning a very high (poor) deviation score.

Note that participants were asked to keep the properties as close to the midpoint of the ideal range as possible. Thus values closer to the midpoint earn lower (better) deviation scores than those within the ideal range but farther from the midpoint. When properties fell outside the ideal range the effects on the solution became more robust. Thus it was best to keep the properties in the ideal range as much as possible, but also as close to the midpoint as possible. Again Inactive D was the one exception to this rule with values at zero earning the most points.

Because some solution states required multiple chemical adjustments across days to be brought back into the ideal range whereas others require only a single adjustment, some solution states necessitated higher overall deviation scores for the trial. To correct for this difference in solution state difficulty, the mean performance for each property in each solution starting state (collapsing across participants) was subtracted from the corresponding deviation score to produce a corrected score. The corrected score was then multiplied by -1 so that positive corrected scores now indicated above average performance and negative corrected scores indicated below average performance. Thus, in Table 4, the participant performed close to average on Properties B and Q, but was far below average on her management of Property B, but above average at managing Inactive D and Temperature, producing a corrected total score just below average for the trial. The total scores were not visible to the participant, but instead the participant could deduce the effectiveness of their strategic decisions based on whether the chemical properties returned to their ideal ranges.

Table 4

Chemistry Task Scoring

	Property level	Ideal value	Deviatio n score	Solution difficulty	Corrected Score
Property B	286.4	219	67.4	65.6	-1.8
Property Q	162	219	56.7	76.6	19.9
Active D	438	219	219.6	27.4	-192.1
Inactive D	12	0	12.1	55	42.9
Temperature	205	219	13.8	94.6	80.8
Trial score			**73.9**	**63.9**	**-10.0**

Note. Ideal value indicates the midpoint of the ideal range which scores are computed from, except Inactive D where 0 is ideal value. Property level indicates the current level of a given property. Deviation indicates the absolute value of the number of points above or below the ideal value each property level is.

Procedures

Participants first completed the demographics questionnaire followed by cognitive pre-tests. Participants next read the instructions for the novel chemistry task at their own pace and were given the option to re-read the instructions prior to taking the cumulative test. As in Study 1, the initial group of participants (20 young, 16 old) was informed that this phase would take up a substantial portion of the experiment and that most people had to reread the information several times. However, of these participants, only half of the older adults completed the study in the designated time period (3 hours for the first 12 participants, extended to 4 hours for the next 4 participants). We were concerned that the aforementioned instruction differentially affected older and younger adults' study times. After removing these statements from the instructions no older adult participant failed to complete the task within four hours. Younger adults' task completion times were not affected by the removal of these instructions, $t(33) = 0.17$, $p = .854$.

Next participants completed the cumulative instructions quiz, but no performance criterion was used. We also removed three questions from the cumulative instructions test that asked about peripheral details (which solution properties produced corrosion, calcium deposits, and toxic fumes when they were outside their ideal range) that were not necessary for performing the chemistry task. Participants then completed the pathfinder ratings task. For Study 2, the pathfinder ratings task was reprogrammed in E-prime 2. This new version of the task presented four item pairs at once and answers were logged after they clicked a submit button. This was done to reduce completion times and to allow participants the option to change responses and prevent miskeys from contaminating the data. As with the cumulative instructions quiz, several items from the pathfinder ratings task that were not directly relevant to novel chemistry task (e.g., whether a property being too high caused corrosion or toxic fumes), were omitted. Thus Study 2 used only 14 items, reducing the total number of ratings to made, from 170 to 136. Participants were prompted half-way through the pathfinder ratings and offered a break.

Following the pre-task pathfinder ratings, participants completed the novel chemistry task, and then completed the ratings task again. Finally, participants completed the instructions quiz a second time, followed by a post-task questionnaire in which participants rated how difficult and tedious each phase of the experiment was, how motivated they were to perform well, and how much effort they put into performing well on a 1-9 Likert scales (Table 2, p. 23-25).

As in Study 2, participants were asked if there were any additional terms they felt should have been included in the pathfinder ratings task, and what the task made them think of.[7] Lastly, participants were asked whether they had ever owned or managed a swimming pool before.

[7] Four participants (one younger and three older adults) listed items that were already in the ratings task. An additional young adult listed, "Rate of Growth" and an additional older adult wrote, "in working with each other or against each other."

Results

Task Representations

As in Study 1, task representations were measured using C scores. A 2 (Age: young, old) X 2 (Time: pre-task, post-task) ANOVA with time as a within subjects factor and age as a between subjects factor was used to examine age differences in task representation and changes in task representations. The main effects of age, $F(1, 63) = 3.57$, $p = .063$, time $F(1, 63) = 0.45$, $p = .505$, and the Age X Time interaction, $F(1, 63) = 0.00$, $p = .957$, were not significant. As in Study 1, older and younger adults had equally accurate task initial representations. Contrary to predictions, neither age group showed strong evidence of task representation updating. Only half of either age group (young = 17, old = 15) showed positive changes in task representation scores, with the remainder showing negative changes (Figure 5, p 35-36). Even for those showing positive changes in C scores, most of those changes were small. Only four younger and three older adults showed changes in C score of .10 of greater.

Novel Task Performance

Corrected chemistry task performance scores were analyzed using SAS Proc MIXED (Littell, Milliken, Stroup, & Wolfinger, 2000) examining the influences of age, initial task representation, changes in task representation (pre, post) and trial as predictors of chemistry task performance.[8] Trials were entered as trials 0 through 35 rather than 1 through 36 so that the intercept indicates Trial 1 performance.

---

[8] Prior to entering trial effects into statistical models, the effects of trial and potential interactions with trial (Age X Trial, Age X Change X Trial) were examined using both visual inspection of individual data plots and trend analyses. The linear trend for trial accounted for 52% of the variance with the remaining variance contained primarily beyond the 10th order, suggesting that the remaining variance was primarily idiosyncratic rather than meaningful changes in the data. Trend analyses for the Age X Trial interaction and the Age X Trial X C score change interaction, indicated that only 10-20% of the variance respectively could be accounted for in the first three orders of trial, suggesting no interaction between age and trial (linear or otherwise).

Changes in C scores

.20	
.19	4
.17	<u>8</u>
.15	<u>4</u>56
.14	<u>1</u>
.13	5
.12	
.11	
.10	
.09	<u>48</u>9
.08	3
.07	<u>2</u>
.06	<u>0</u>0<u>0</u>
.05	<u>2</u>46
.04	06
.03	2<u>4</u>
.02	<u>5</u>88
.01	01<u>1</u>2<u>357</u>
.00	
-.00	<u>1</u>
-.01	012<u>6</u>
-.02	<u>5</u>9
-.03	0<u>9</u>
-.04	2<u>8</u>
-.05	44
-.06	<u>11</u>55
-.07	<u>9</u>
-.08	<u>7</u>
-.09	34
-.10	3
-.11	<u>4</u>
-.12	<u>5</u>7
-.13	<u>6</u>
e	8
-.15	3<u>6</u>
-.16	<u>1</u>3
-.17	<u>2</u>
-.18	
-.19	
-.20	2

Figure 5. Stem and Leaf Plot for Changes in C Scores (C Score 1 – C Score 2). Underlined numbers indicate older adults.

Model 1: Changes in performance with practice. The first model tested for age differences in overall performance and performance changes with practice (Table 5, p. 37-38). A main effect of trial indicates that participants' performance on the chemistry task improved with practice. However, the failure to obtain a main effect of age or an Age X Trial interaction indicates that older and younger adults performed similarly on chemistry task and showed similar improvements across trials (Figure 6, p. 39). If anything, the non-significant Age X Trial trend indicates that older adults may have improved slightly more over trials compared to younger adults. This suggests that older adult's strategic choices, which directly impact performance scores, were as good as those of young adults and improved over trials at comparable, if not greater rate. That is, older adults did not demonstrate suboptimal strategic choices or failures to alter their strategic choices with practice.

Table 5

Model Results for Chemistry Task Performance

Model 1

	Estimate	SE	df	t-value	p-value
Intercept	-5.65	1.80	63	-3.14	.003
Age	2.53	2.64	63	0.96	.343
Trial	0.17	0.09	2273	1.98	.048
Age X Trial	0.18	0.13	2273	1.35	.176

Model 2

	Estimate	SE	df	t-value	p-value
Intercept	-45.49	2.95	63	-15.44	< .001
Age	12.58	4.20	63	2.99	.004
Initial C score	157.66	9.65	2271	16.33	< .001
Trial	0.17	0.08	2271	2.15	.032
Age X Initial C score	-59.33	12.50	2271	-4.75	< .001
Age X Trial	0.18	0.12	2271	1.47	.141

Model 3

	Estimate	SE	df	t-value	p-value
Intercept	-50.07	2.90	63	-17.26	< .001
Age	13.75	4.86	63	2.83	.006
Initial C score	178.71	9.59	2267	18.64	< .001
C score change	100.54	17.18	2267	5.85	< .001
Trial	0.17	0.08	2267	2.20	.028
Age X Initial C score	-68.70	15.23	2267	-4.51	< .001
Age X C score change	-86.11	27.62	2267	-3.12	.002
Age X Trial	0.18	0.12	2267	1.53	.127
C score change X Trial	-0.06	0.84	2267	-0.07	.943
Age X Trial X C score change	0.32	1.24	2267	0.26	.798

Model 4

	Estimate	SE	df	t-value	p-value
Intercept	-37.71	6.46	59	-5.83	< .001
Age	-44.25	9.11	59	-4.86	< .001
Initial C score	170.15	10.18	2267	16.72	< .001
C score change	102.59	17.04	2267	6.02	< .001
Vocabulary	0.14	0.20	59	0.71	.483
Processing speed	-0.32	0.12	59	-2.77	.008
Trial	0.17	0.08	2267	2.24	.025
Age X Initial C score	-75.87	15.58	2267	-4.87	< .001
Age X C score change	-87.04	27.26	2267	-3.19	.001
Age X Vocabulary	0.64	0.26	59	2.51	.015
Age X Processing speed	1.41	0.20	59	6.95	< .001
Age X Trial	0.18	0.11	2267	1.56	.120
C score change X Trial	-0.06	0.82	2267	-0.07	.942
Age X Trial X C score change	0.32	1.22	2267	0.26	.795

Note. Models of chemistry task performance. Performance scores are corrected for trial difficulty. Performance scores of zero indicate average performance, with positive values being above average and negative values being below average. Younger adults serve as the reference group. Thus positive age effects indicate better performance or steeper slopes for older adults relative to young. Main effects indicate the beta estimate for young adults with the interaction coefficient indicating how beta estimate changes for older adults.

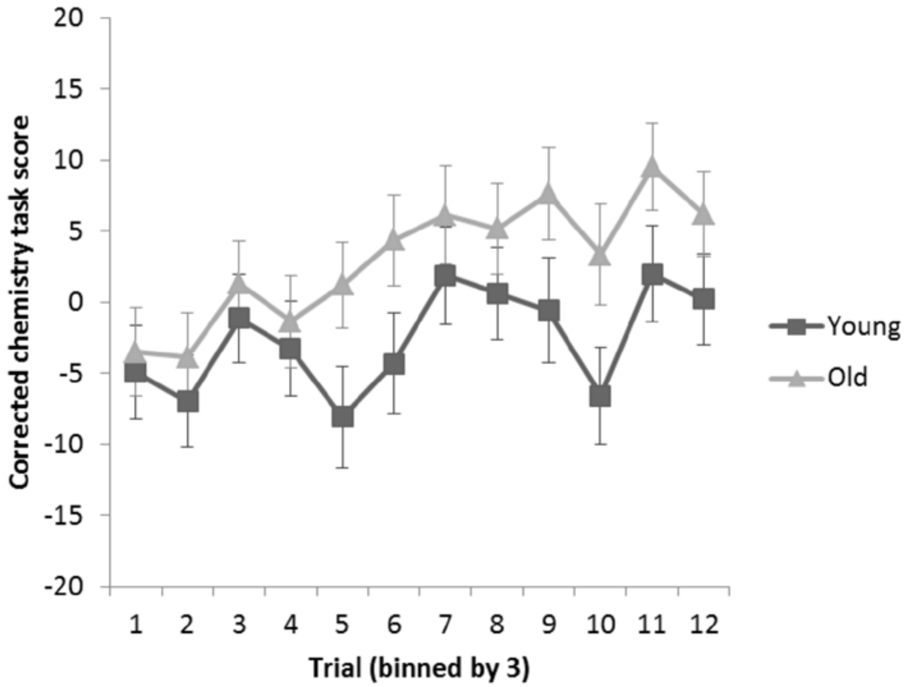


Figure 6. Chemistry Task Performance over Trials by Age Group. Data averaged over bins of 3 trials. Note that data were binned for visual purposes, but unbinned data were analyzed.

Model 2: Performance as a function of initial task representation. Despite not finding an age difference in performance or task representation updating, we continued with the a priori analysis plan to determine whether task representations influenced strategic choices for younger and older adults. Model 2 added initial (pre-task) C scores and the interaction with age as predictors of chemistry task performance (Table 5, p. 37-38). A main effect of pre-task C scores confirmed that more accurate initial task representations produced better chemistry performance (Figure 7, p. 40). When including pre-task C scores in the model, we obtain a significant age effect, with older adults outperforming younger adults when C scores are held constant. However, these main effects were qualified by a significant Age X Initial C score interaction. This effect indicated that the slope of the initial C score-performance relationship was steeper for young adults than for older adults. The main effect of trial continued to be significant, and the Age X Trial interaction remained non-significant.

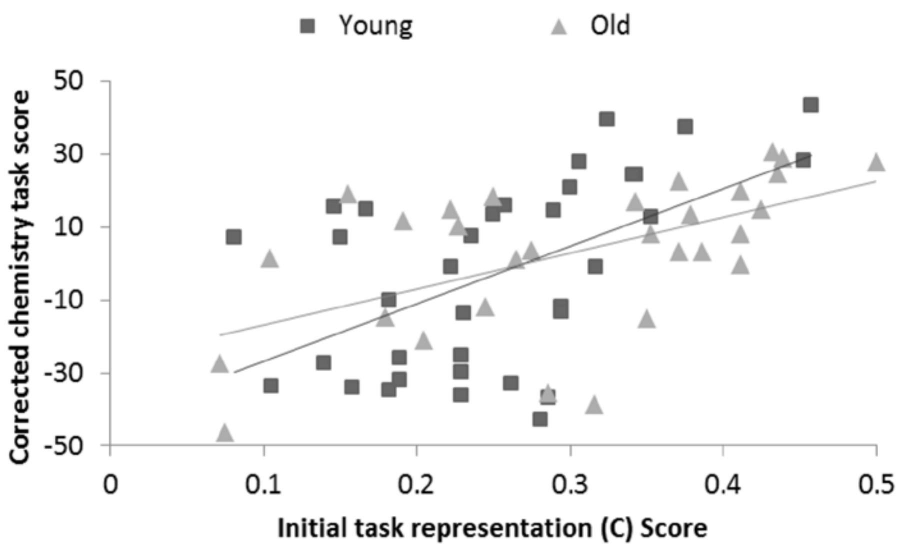


Figure 7. Relationship between Initial C Scores and Corrected Chemistry Task Score.

To illustrate the combined main effect of age and significant Age X Initial C score interaction, consider the examples in Table 6 (p. 41). Although the main effect of age indicates higher performance in older adults, this is largely undermined by the stronger Age X Initial C score interaction. Older adults' shallower slope for initial C scores results in higher performance estimates (relative to younger adults) when task representations are poor, but lower performance estimates (relative to younger adults) when initial C scores moderate to high. For example, an older adult with an initial C score of .15 will typically perform somewhat better (chemistry task score of -18.16) compared to a younger adult with a similarly poor task representation (chemistry task score of -21.84). Note that these estimates are for the first trial (0) of the task, and thus both age groups can be expected to attain higher scores following practice. However, an older adult with a highly accurate task representation (initial C score of .45) will perform slightly lower (chemistry task score of 11.34) compared to a younger adult with a similar initial C score (chemistry task score of 25.46). Therefore, younger adults with accurate task representations appear to be slightly better at utilizing them than their older adult counterparts. However, older adults appear to be performing slightly better than young when task representations are poor. This could be the result of motivational factors or C scores perhaps being poorer measures of older adults' task representations. Each of these is considered in the discussion and supplemental analyses.

Table 6

Model 2 Chemistry Task Score Examples

Initial C score (C1)	Young	Old
.15	-21.84	-18.16
.30	1.81	-3.41
.45	25.46	11.34

Note. Model-derived estimates for chemistry task scores for older and younger adults at three different initial C scores when Trial = 0 (first trial).

Model 3: Changes in performance as a function of task representation updating. Although no general improvement in C scores were found, some participants did show improvements in their task representations, whereas others showed little change or even decrements following practice. These latter decrements could reflect forgetting or confusion, whereas the improvements could reflect actual improvements and may influence task performance. Thus we continued with the a priori plan to examine the importance of task representation updating via a mixed model which added changes in C scores and the interactions of C score Change, Trial, and Age into the model (Table 5, p. 37-38).

The main effects of initial C scores and the Age X Initial C score interaction remained significant, with higher initial C scores predicting better overall task performance, and the relationship between the two being steeper for young adults. A significant positive main effect of Changes in C scores was qualified by a significant negative Age X C score change interaction. The main effect indicates that improvements in task representation scores for pre- to post-task were associated with higher performance for young adults. However, the high negative regression coefficient (-89.11) for the interaction with age indicates that older adults showed only modest gains when C scores improved, with a regression weight of 14.43 (100.54-86.11) compared to younger adults 100.54.

Theoretically, C scores are thought to reflect deep structural knowledge. Thus, any changes in C scores should reflect the addition of new knowledge or the reorganization of existing knowledge. If deep structural task representation updating is responsible for improvements in the task over trials, then one would expect a C score change X Trial interaction, with changes in performance over trials being greater for those participants showing evidence of deep structural updating. However, continued significance of the main effect of trial after accounting for C score change, and the non-significant Change X Trial, and Age X Change X Trial interactions indicate that the changes in task performance over trials seen in Model 1 are not entirely explainable by the structural updating of task-related knowledge, as it is measured by changes in C scores. It may be the case that deep structural knowledge updating occurs relatively early in the task (if at all), and thus is reflected in performance throughout the task. By contrast, the small increases in performance over trials throughout the task may indicate tuning of relevant knowledge, and not the addition or reorganization of knowledge. That is, one may tune their knowledge of *how much* a chemical agent influences a solution property without learning new structural relationships between those chemical agents and solution properties.

The minimal effect of changes in C scores on older adults' performance also calls into question the validity of pre- and post-task changes in C scores as a measure of knowledge updating in older adults. We comment on this further in the supplemental analyses and results section.

Supplemental Analyses

To follow up on the age equivalence in C scores and chemistry task performance, we examined the potential roles of task motivation, processing speed, and crystalized knowledge as explanations for older adults' high level of performance. We also reanalyzed the chemistry task data after isolating performance to two properties in the chemistry task that *required* new learning to manage (because the chemistry task instructions omitted critical information about these properties). If knowledge updating is as rare as changes in C scores suggest, then we should see minimal improvement over trials on these specific properties. Additionally, if performance also improves on these select properties, then they provide a better test for potential age differences in knowledge updating.

Motivation and effort. As noted earlier, ratings for motivation and effort were higher for older adults compared to younger adults at all stages of the task (Table 2, p. 23-25). Motivation and effort during the chemistry task, and motivation and effort for learning the missing information from the instructions were highly correlated with one another (rs = .23—.95) and were thus combined to form a single measure. This combined motivation/effort measure correlated with chemistry task performance, r = .45, p = .002 (r_{young} = .51, p = .002; r_{old} = .40, p = .028; Figure 8, p. 44). The scatterplot in Figure 8 shows that a number of highly motivated older adults still performed rather poorly. Thus, although some older adults may compensate for age related changes via enhanced motivation and effort, this explanation appears to be inadequate for a number of older adults. Ultimately a restriction of range issue makes the relationship between motivation and performance in older adults difficult to assess. Because motivation correlated strongly with subjective understanding of the chemistry task (rated 1 to 9; r_{young} = .52, p = .001; r_{old} = .50, p = .004), it is possible that subjective success increased motivation, or that subjective success was used as a basis for inferring motivation (both were rated post-task). Likewise subjective task performance correlated strongly in actual performance (r_{young} = .43, p = .009; r_{old} = .51, p = .004), indicating that participants had a somewhat accurate sense of their performance and understanding during the task.

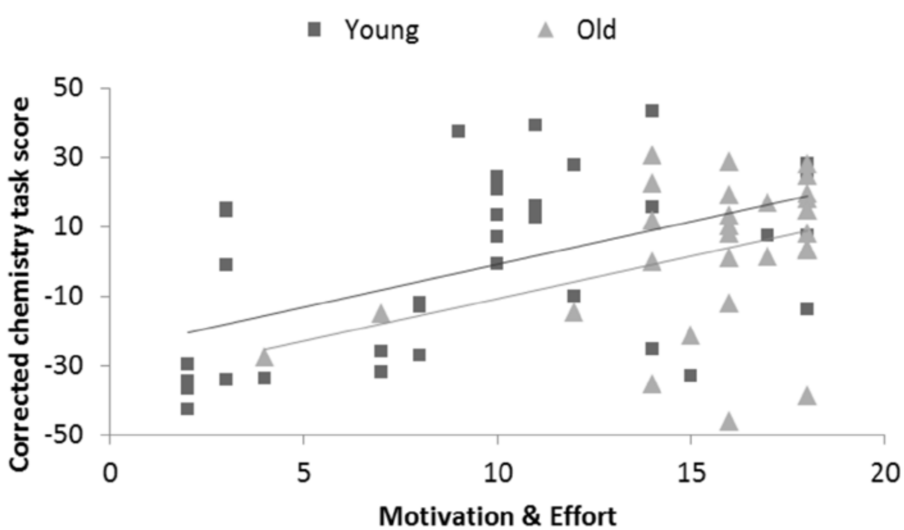


Figure 8. Relationship between Motivation and Chemistry Task Performance in Younger and Older Adults. Motivation & Effort = sum of self-reported motivation to do well on the chemistry task (1-9 scale) and effort on the chemistry task (1-9 scale). Higher values indicate greater motivation and effort.

Vocabulary knowledge and processing speed. Older adults also had higher vocabulary scores. In addition to being a loose measure of crystalized intelligence, vocabulary scores often correlate highly with reading comprehension scores (Jeon & Yamashita, 2014). Here, vocabulary scores correlated positively with chemistry task performance, particularly in older adults, but not necessarily for younger adults (Table 7, p.45). Similar effects were found for processing speed, again in older, but not younger adults. These correlations suggest that older adults with adequate processing speed and superior vocabulary knowledge may use these abilities to overcome age-related declines in other cognitive abilities.

Table 7

Correlations with Vocabulary and Processing Speed

	Chemistry Task	Properties B & Q	C score 1	C score 2	Cumulative Quiz 1	Cumulative Quiz 2
			Young			
Vocabulary	.27	.28	*.30*	**.34**	*.32*	*.32*
Processing speed	-.18	-.19	-.25	-.05	**-.38**	-.11
			Old			
Vocabulary	*.35*	**.45**	.17	.16	*.32*	*.32*
Processing speed	**.41**	*.31*	.19	.06	-.04	.13

Note. Processing speed = number correct on out of 30 on Salthouse's pattern comparison task (1993); Vocabulary = number correct out of 36 on the Advanced Vocabulary Test (Ekstrom et al., 1976); Chemistry task = Corrected performance on the chemistry task; Properties B & Q = Corrected performance scores on the chemistry task when restricting the data to only performance on Properties B and Q; C score 1 = pre-task proportion of concept map overlap between the participant and reference models ranging from 0 (no overlap) to 1 (perfect overlap); C score 2 = post-task proportion of concept map overlap; Cumulative test 2 = proportion correct on out of post-task cumulative instructions test; Cumulative test 1 = proportion correct on out of 12 items on the pre-task cumulative instructions test. Bold values indicate significance at the $p < .05$ level. Italicized values indicate trending significance, $p < .10$.

To test whether vocabulary or processing speed could account for variance above and beyond that already explained by C scores, we entered these factors into a fourth model predicting chemistry task scores (Table 5, 37-38). The main effect of vocabulary score was not significant, but an Age X Vocabulary interaction reveals that vocabulary performance was related to higher chemistry task performance in older but not younger adults. Processing speed was surprisingly a significant negative predictor for younger adults, but an Age X Processing speed interaction shows that higher processing speed was a positive predictor for older adults. These results support the hypothesis that older adults with spared processing speed and/or greater crystalized intelligence may rely on these capacities to maintain high performance on novel and cognitively demanding tasks (Masunaga & Horn, 2001; Morrow et al., 2003, 2009).

Although self-rated prior knowledge of chemistry did not differ with age, it correlated positively with chemistry task performance in both younger, $r = .37$, $p = .027$, and older adults $r = .43$, $p = .010$. However, this measure was taken after participants performed the novel chemistry task, and thus may be contaminated by perceptions of task performance. Indeed, chemistry knowledge ratings and subjective task understanding correlated strongly in both younger ($r = .57$, $p < .001$) and older adults ($r = .77$, $p < .001$).

Vocabulary knowledge and processing speed in Study 1. Given the predictive effects of vocabulary knowledge and processing speed in for chemistry task performance, we reanalyzed Study 1 using multiple regression with C scores, vocabulary, processing speed, age, and all possible interactions with age as predictors of inference test performance (Table 8). As can be seen in Table 8, neither vocabulary ability nor processing speed predicted inference task performance after accounting for C scores. Additionally the interactions with age were not significant. Thus, crystalized knowledge and processing speed do not appear to be generally beneficial for utilizing task representations, but instead appear to be uniquely important for performing novel tasks. However, it is important to keep in mind the lower power in Study 1 compared to Study 2.

Table 8

Study 1 Inference Test Performance

	Estimate	SE	df	t-value	p-value
Intercept	0.56	0.22	32	2.60	.001
Age	-0.34	0.30	32	-1.13	.265
C score	1.23	0.39	32	3.17	.003
Vocabulary	-0.01	0.01	32	-1.30	.202
Processing speed	-0.00	0.00	32	-0.70	.487
Age X C score	-0.26	0.52	32	-0.51	.617
Age X Vocabulary	0.02	0.01	32	1.45	.149
Age X Processing speed	0.00	0.01	32	0.31	.762

Note. Study 1 proportion correct on the inference test. Younger adults serve as the reference group. Thus positive age effects indicate better performance or steeper slopes for older adults relative to young. Main effects indicate the beta estimate for young adults with the interaction coefficient indicating how beta estimate changes for older adults.

Evidence for deep structural updating. In addition to the null effects for changes in C scores, performance on the chemistry instructions quiz for the impoverished information remained near floor, with younger and older adults on average correctly answering only 0.58 and 0.61 (out of 3 questions), respectively.

But do these findings really indicate that all improvements in chemistry task scores were the result of tuning rather than structural updating of knowledge? To investigate this possibility, we isolated performance on the chemistry task for just the management of Properties B and Q. The impoverished instructions indicated that low levels of Property Q would lower Property B, and that Chemicals C, T, and N would impact Properties B and Q. However, the instructions did not tell participants *how* Chemicals C, T, and N would impact Properties B and Q. Without this specific information, it is impossible to systematically adjust Properties B and Q. Additionally, Properties B and Q did not interact with any other chemical agents or chemical properties.

Therefore, any changes in the ability to manage Properties B and Q, *necessitate* some degree of structural knowledge updating. Thus we re-ran Models 1 through 4 on this subset of chemical properties (Table 9, p. 49-51).

Model 5: Changes in performance managing Properties B and Q with practice. A main effect of trial indicates that participants do show indirect evidence of structural knowledge updating (Figure 9, p. 49).[9] A main effect of age indicates that older adults performed worse on these particular items relative to younger adults. The Age X Trial interaction was not significant, suggesting that although older adults performed worse on average, they improved at a similar rate compared to younger adults. This may indicate that older adults generally struggled somewhat to update their task representations relative to young adults, but like younger adults continued to improve their performance via tuning when they were able to update. For example, older adults may be learning how one of the chemicals affects Property B or Q, but without understanding how the other chemicals function will achieve only modest scores on these properties. Note that this interpretation is inconsistent with the argument that adding new structural knowledge is less resource consuming than tuning existing structural knowledge (Schnotz & Preuss, 1997).

[9] As with overall chemistry task performance, trend analyses indicated that performance changes for Properties B and Q were primarily linear, with the linear trend accounting for over 50% of the variance and the remaining variance accounted for beyond 4^{th} order.

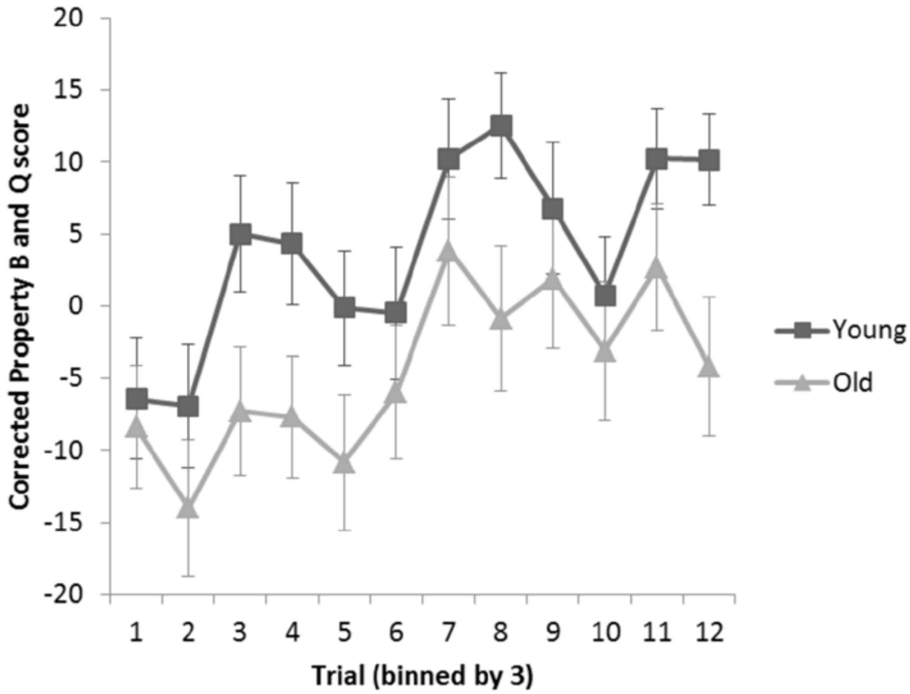


Figure 9. Chemistry Task Performance for Properties B and Q over Trials by Age Group. Data averaged over bins of 3 trials. Note that data were binned for visual purposes, but unbinned data were analyzed.

Table 9

Model Results for Chemistry Task Performance on Properties B and Q

Model 5

	Estimate	SE	Df	t-value	p-value
Intercept	-3.88	2.36	63	-1.64	.105
Age	-6.97	3.48	63	-2.00	.050
Trial	0.44	0.12	2273	3.80	< .001
Age X Trial	-0.07	0.17	2273	-0.45	.651

Model 6

	Estimate	SE	Df	t-value	p-value
Intercept	-32.57	4.09	63	-7.96	< .001
Age	-5.54	5.84	63	-0.95	.346
Initial C score	113.50	13.41	2271	8.47	< .001
Trial	0.44	0.11	2271	3.91	< .001
Age X Initial C score	-23.54	17.36	2271	-1.36	.175
Age X Trial	-0.08	0.17	2271	-0.47	.642

Model 7

	Estimate	SE	df	t-value	p-value
Intercept	-34.60	4.12	63	-8.08	< .001
Age	-21.93	14.14	63	-1.55	.133
Initial C score	123.79	13.60	2267	9.10	< .001
C score change	78.20	24.37	2267	3.21	.001
Trial	0.43	0.11	2267	3.81	< .001
Age X Initial C score	-0.04	21.61	2267	-0.00	.999
Age X C score change	-9.59	39.18	2267	-0.24	.807
C score change X Trial	-1.69	1.19	2267	-1.42	.157
Age X Trial	-0.07	0.17	2267	-0.43	.665
Age X Trial X C score change	0.90	1.76	2267	0.51	.609

Model 8

	Estimate	SE	df	t-value	p-value
Intercept	-27.87	9.24	59	-3.02	.004
Age	-65.88	13.01	59	-5.06	< .001
Initial C score	111.04	14.54	2267	7.64	< .001
C score change	76.90	24.34	2267	3.16	.002
Vocabulary	0.45	0.29	59	1.54	.128
Processing speed	-0.27	0.16	59	-1.64	.107
Trial	0.43	0.11	2267	3.85	< .001
Age X Initial C score	-6.87	22.27	2267	-0.31	.758
Age X C score change	-11.97	38.96	2267	-0.31	.759
Age X Vocabulary	0.78	0.36	59	2.14	.036
Age X Processing speed	1.01	0.29	59	3.49	< .001
Age X Trial	-0.07	0.16	2267	-0.44	.662
C score change X Trial	-1.69	1.18	2267	-1.43	.152
Age X Trial X C score change	0.90	1.74	2267	0.52	.605

Note. Models of chemistry task performance for Properties B and Q (only). Performance scores are corrected for trial difficulty. Performance scores of zero indicate average performance, with positive values being above average and negative values being below average. Younger adults serve as the reference group. Thus positive age effects indicate better performance or steeper slopes for older adults relative to young. Main effects indicate the beta estimate for young adults with the interaction coefficient indicating how beta estimate changes for older adults.

Model 6: Performance on Properties B and Q as a function of initial task representations. We next added initial C scores and the interactions with age and trial to the model (Table 9, p. 49-51). Somewhat surprisingly, initial C scores were predictive of performance on Properties B and Q. This is despite the fact that managing these properties requires information not present prior to the initial pathfinder ratings task. Although participants lack specific knowledge of how Chemicals C, T, and N influence Properties B and Q, initial C scores may still contain useful information that supports learning and management of Properties B and Q—if the participant can learn update the missing information in their task understanding.

For example, understanding that Chemicals C, T, and N will influence only Properties B and Q, as opposed to thinking they will influence other Properties, is valuable for limiting the number of options that need to be experimented with early in the task. Likewise, knowing how the other properties (Active D, Inactive D, and Temperature) function reduces the need to monitor and update one's knowledge of those properties. This may free up resources to monitor Properties B and Q specifically. Lastly, knowing that Properties B and Q interact with one another should allow for better management of the two *after* one has learned which chemicals influence each property.

Model 7: Changes in performance on Properties B and Q as a function of task representation updating. We next added changes in C scores and interactions with age and trial (Table 9, p. 49-51). Unlike in Model 3 for overall chemistry task performance (Table 5, p. 37-38), changes in C scores were similarly positive predictors for both age groups (main effect, but no interaction with age). Again, changes in C scores did not interact with trial for improvements in performance on Properties B and Q, casting doubt on the hypothesis that changes in C scores serve as a useful measure for changes in deep structural updating, as reflected in improved task performance. We comment further on this finding in the discussion section.

Model 8: The role of processing speed and vocabulary on performance managing Properties B and Q. As with overall chemistry task performance, we examined the potential role of processing speed and vocabulary scores on performance (Table 9, p. 49-51). As with overall task performance (Table 5, p. 37-38), both processing speed and vocabulary ability predicted higher performance in older adults (significant interactions with age) but in this case neither was predictive for younger adults (no main effect of processing speed or vocabulary). This again supports the hypothesis that older adults with spared cognitive resources (higher processing speed) are better able to perform this novel task, in this case, even when performance necessitates updating one's task understanding. Likewise, these results suggest that greater vocabulary ability (whether this reflects crystalized knowledge, general language ability, or both) also serves as a protective factor for older adults, and may explain why many older adults were able to perform at or near the level as their younger counterparts.

Discussion

The finding that older adults form equally accurate initial task representations was not surprising given the results of Study 1. This occurred even after the learning criterion was removed. These findings are consistent with prior research on narrative processing which suggests that older adults show no impairment in the ability to form mental representations of text passages (Gilbert et al., 2004; Morrow, Leirer, Altieri, & Fitzsimmons, 1994b; Radvansky et al., 2003, 1990; Radvansky & Dijkstra, 2007; Radvansky, 1999a; Stine-Morrow et al., 2004, 2002). Likewise, older adults showed a similar relationship between initial task representations (C scores) and task performance, indicating little or no deficit in the ability to utilize structural knowledge to make strategic task decisions.

What was unexpected was the age equivalence in chemistry task performance throughout the task and continued age equivalence in C scores at the end of the task. Whereas we predicted age-related declines in performance and substantial task representation updating for young but not old, we found age equivalence in performance and no initial evidence of task representation updating in either age group (although there is individual variability suggesting that some participants may have updated whereas others may have forgotten information, or mis-updated). These results suggested that the deep structural updating thought to be measured by changes in C scores was generally absent, with improvements in task performance possibly being driven by tuning (modifying the strength of existing relationships), rather than structural updating. Additional support for this argument comes from the failure to find any evidence of a C score change X Trial interaction. This null effect suggests that changes in task understanding were not responsible for the observed linear changes in task performance over trials.

However, changes in C scores, though strongly related to overall performance in younger adults were only weakly related to performance in older adults. This calls into question the validity of the post-task C score measure, particularly for older adults. It is possible that older adults (and possibly some young adults) do update their structural knowledge, but that this new knowledge is simply not captured on the second set of pathfinder ratings. One explanation is that older adults may learn the spatial mapping of the chemical additives in the chemistry task and learn which property each influences. However, if they do not encode the names of these chemical additives, they will be unable to demonstrate this new knowledge during the ratings task—which requires that the names of terms be tied knowledge about those concepts. For example, a participant may learn that the chemical in top left raises Properties Q, the chemical in the lower left lowers Property Q, and the middle left chemical raises both Properties B and Q. However, if the participant only learns these spatial-to-effect mappings, instead of the actual chemical names, then they will still be unable to accurately indicate which chemicals are related to which properties on the subsequent ratings task. To consider this possibility, a ratings task with location information might be used.

The mental model and conceptual knowledge frameworks argue that deep structural knowledge is more important than declarative knowledge (Curtis & Davis, 2003). Proponents of the concept mapping approach also argue that concept mapping is better at tapping this important deep structural knowledge than are explicit tests of declarative knowledge (Goldsmith et al., 1991; Novak & Cañas, 2006; Novak, 1990; Staggers & Norica, 1993). Thus, it may be that older adults did not incorporate the new information into their existing knowledge structure, but instead learned explicit relationships as declarative knowledge only. However, performance on the post-task quiz, which explicitly asked participants to indicate which chemicals increased and which decreased Properties B and Q, was near floor, arguing against this interpretation. However both these findings can be explained by older adults simply failing to learn the explicit labels necessary to accurately complete either the post-task ratings or cumulative test.

The strongest support that participants did indeed update their structural knowledge and not merely tune existing knowledge comes in the form of performance changes on Properties B and Q specifically. Although the instructions contained important information regarding how these properties interacted with each other, it did not contain information on how each chemical would impact these properties. Specifically it only indicated that Chemicals T, C, and N would influence the properties but that participants would have to learn via completing the task which chemicals raised and lowered which properties. Thus, any improvement in managing Properties B and Q requires updating the structural relationships between concepts. Although older adults generally did worse when managing these specific properties, there was no age difference in the rate of improvement across trials, suggesting that older adults were generally as effective as young at updating their task understanding for these properties and chemicals. This is despite the lack of change in older adults pre- to post-task C scores, again suggesting that post-task C scores may fail to capture new structural knowledge in older adults. Showing the interface during the ratings task might allow older and younger adults to better convey their new knowledge if they encoded only the spatial locations and effects of, but not the names, of the different task elements.

Whether through equivalent rates of tuning, accretion (adding new knowledge) or reorganization, older adults demonstrated improvements in performance comparable to those of younger adults. This is particularly interesting given the supposed difficulty of both tuning and reorganizing existing knowledge (Schnotz & Preuss, 1997). Given this age equivalence, as well as the age equivalence in initial C scores, we investigated potential explanations for why older adults were able to achieve performance levels equivalent to those of younger adults. Older adults generally indicated being more motivated to perform well at each stage of the task and having put more effort into the task. However these ratings were assessed post-task and may have been influenced by subjective task performance. By contrast, processing speed and vocabulary were assessed with commonly used psychometric tests prior to chemistry task performance. Both processing speed and vocabulary scores predicted better performance for older but not younger adults on overall chemistry task performance and on Properties B and Q specifically. These findings suggest that spared cognitive functioning (processing speed) was important for learning and performing the novel chemistry task for older adults. Additionally, vocabulary scores were significant predictors of older adults' performance even after accounting for processing speed. This suggests that older adults with lesser cognitive abilities may have been able to use their generally greater crystalized intelligence to compensate for age-related declines in processing speed. By contrast, vocabulary scores and processing speed did not predict superior performance in younger adults, who may not need to rely as heavily on crystalized knowledge, and whom likely possess adequate cognitive abilities (i.e. processing speed) to perform the task. Performance in younger adults was instead related to low motivation and effort. But again, these ratings were collected following chemistry task performance and may be contaminated by subject performance on the chemistry task.

Although consistent with prior research (Masunaga & Horn, 2001; Morrow et al., 2003, 2009), the relationship between vocabulary scores and chemistry task performance in older adults was surprising. The novelty of the chemistry task should have minimized the likelihood of prior knowledge of any sort influencing task performance. Alternatively, vocabulary knowledge may have related to task performance via its relation to reading comprehension (Jeon & Yamashita, 2014). If this is the case, older adults with higher vocabulary scores may be better able to extract information from the instructions, which would aid chemistry task performance.

CHAPTER IV

GENERAL DISCUSSION

Task representations are theorized to guide strategic choices on a variety of tasks (Bromme et al., 2009; Lovett & Schunn, 1999; Muis, 2007; Winne & Hadwin, 1998). However no study to our knowledge had examined age differences in task representations, despite findings of age differences in strategic choices. In two studies we demonstrate that pathfinder-derived concept maps (C scores) can be used to measure younger and older adults task representations for deep structural relationships learned via text. Likewise, we found strong relationships between C scores and performance on an inference test (Study 1) and a novel chemistry task (Study 2), consistent with the contention that task representations are an important determinant of strategic choices. However, a general lack of improvement in C scores from pre- to post-task calls into question the utility of concept mapping for measuring task representation updating—particularly for older adults.

Although we found some evidence of task representation updating via performance changes, we did not find evidence for age decrements in updating. A number of factors may explain this age equivalence. Older adults with better cognitive abilities (i.e. better processing speed) may have had the adequate resources necessary for simultaneously monitoring and updating while performing the novel chemistry task. Other older adults may have relied on superior linguistic skills, as well as any prior knowledge they were able to relate to the material to buffer against age-related changes. Likewise, because the learning phase was self-paced, older adults may have been better able to form accurate initial task representations which may require less updating of the already learned information and thus allow more cognitive resources to be allocated to learning the new information.

Additionally, because the chemistry task itself was self-paced, it may have allowed older adults to compensate for age-related declines by taking more time to make decisions and learn the information via the task. Therefore, these results may not generalize to knowledge updating on tasks involving time limits or instructions emphasizing speeded performance. Under speeded conditions older adults may indeed show deficits in knowledge updating. Why then, does higher processing speed predicted higher performance in older adults on a self-paced task? The answers to this question come in two forms. First, when processing speed is inadequate for a given cognitive process, it can have downstream effects on other cognitive processes that occur simultaneously or later on, which also rely on the inadequately completed aforementioned process (Salthouse, 1996). Second, age-related changes in processing speed often coincide with demyelination and a general compromising of cognitive capacity (Salthouse, 1996). Thus, superior processing speed itself may not be driving better performance in older adults, but rather, superior processing speed may be a marker for generally spared cognitive abilities. Thus, some other ability which also remains intact in our older adults with greater processing speed, for example working memory, may actually be responsible for our older adults' spared utilization of task representations.

Because vocabulary and processing speed correlated more strongly with performance on the novel chemistry task than with initial C scores, post-task C scores, pre-task cumulative test scores, or post-task cumulative test scores (Table 7, p. 45), this would suggest that these abilities were responsible for the utilization, not formation or updating, of task representations. Likewise, the reanalysis of Study 1 data found that neither processing speed nor vocabulary scores predicted performance on the novel inference test, after accounting for task representation. Taken together, these findings suggest that performing rather than merely learning about a novel task draws on these abilities in older adults. Future research should include latent variable measures of working memory, processing speed, and crystalized intelligence to consider these questions.

Lastly, the age equivalence in performance in both Studies 1 and 2 suggest no age-related decrement in the ability to utilize structural knowledge. According to the mental model perspective these results would suggest age-related sparing in the ability to form and manipulate mental models to solve problems or make inferences (Johnson-Laird, 2001; Schnotz & Preuss, 1997). These results add to existing literature on age-related sparing of mental representation formation and utilization (Masunaga & Horn, 2001; Morrow et al., 2003, 2009).

Better methods for measuring changes in task representations need to be developed. A general method for measuring changes in task representations (such as concept mapping) may not be the most effective approach. By contrast, idiosyncratic methods depending on the task may be more effective. For example Harada, Mori, and Taniue (2010) examined older adults' use of an electronic diet-support system. They found that different methods of instruction had an impact on the interface navigation patterns of older adults, with some instructions producing fewer navigation errors (entering the wrong menu for a given task) than others. Thus, cessation of such task errors could be a valuable indictor of changes in task representation. Likewise, some strategies necessitate a change in task representation. For example, Haider and Frensch (1996) developed an alphabet verification task where participants were told to check an alphabetical string for errors. Contrary to the instructions, all alphabetical errors were contained only in the final three characters of the string. Thus participants could shift to a more efficient strategy of processing only the critical three characters at the end of the string, but only if they corrected their initial faulty representation of the task. Likewise, Kieras and Bovair (1983) gave participants a series of procedures to follow to make a computerized spaceship fire a laser. However, the procedures involved a number for step peripheral to the laser system, which were thus unnecessary. A number of participants identified these unnecessary steps and switched to a more efficient procedure. Presumably, this occurred because these participants were able to form an accurate mental representation of the spaceship and update their task representation accordingly (Kieras & Bovair, 1983). However, neither of these studies compared younger and older adults' abilities to update their task representations and switch strategies accordingly. The downside to this methodology is that if age differences are found, it will remain unclear whether these constitute a failure to update knowledge, or a metacognitive control failure to utilize the new knowledge (hence the current studies attempt to measure task representations directly).

In addition to better measurement of task representation updating, future research should aim to determine under what situations younger and older adults are more likely to engage in task representation updating, and under what circumstances age decrements in updating are most likely to occur. For example, older adults may be more likely to update their task representations in familiar domains, where working memory constraints may be reduced. Likewise, explicitly informing older adults to consider the task design and their task approach, throughout a task, or during mandatory task breaks may allow older adults to better consider their approach when relived of the demands of simultaneously performing the task.

Conclusions

Although the current studies found only minimal evidence for age deficits in task representation updating, they provide several important findings. Concept mapping appears to be a valuable tool for assessing initial task representations but may be limited in the ability to capture changes in those representations—particularly among older adults. The general age equivalence in task representations and performance also contributes to the larger literature on mental representations of text where older adults demonstrate equivalent performance to younger adults. Lastly, the findings that processing speed, superior crystalized intelligence, and high motivation/effort may be potential explanations for age-related sparing in mental representations are novel and warrant future research.

REFERENCES

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology*, *86*(2), 303–311. doi:10.1037//0022-0663.86.2.303

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT–R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*, 439–462.

Besnard, D., Greathead, D., & Baxter, G. (2004). When mental models go wrong: co-occurrences in dynamic, critical systems. *International Journal of Human-Computer Studies*, *60*(1), 117–128. doi:10.1016/j.ijhcs.2003.09.001

Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging*, *9*(2), 287–302. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8054177

Blanchard-Fields, F., Hertzog, C., & Horhota, M. (2012). Violate my beliefs? Then you're to blame! Belief content as an explanation for causal attribution biases. *Psychology and Aging*, *27*(2), 324–337. doi:10.1037/a0024423

Bouazzaoui, B., Isingrini, M., Fay, S., Angel, L., Vanneste, S., Clarys, D., & Taconnat, L. (2010). Aging and self-reported internal and external memory strategy uses: The role of executive functioning. *Acta Psychologica*, *135*(1), 59–66. doi:10.1016/j.actpsy.2010.05.007

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging*, *3*(3), 249–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3268266

Bromme, R., Pieschl, S., & Stahl, E. (2009). Epistemological beliefs are standards for adaptive learning: A functional theory about epistemological beliefs and metacognition. *Metacognition and Learning*, *5*(1), 7–26. doi:10.1007/s11409-009-9053-5

Capelo, C., & Dias, J. F. (2009). A feedback learning and mental models perspective on strategic decision making. *Educational Technology Research and Development*, *57*(5), 629–644. doi:10.1007/s11423-009-9123-z

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: the role of schematic support. *Psychology and Aging*, *20*(4), 718–21. doi:10.1037/0882-7974.20.4.718

Castel, A. D. (2007). Aging and memory for numerical information: the role of specificity and expertise in associative memory. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *62*(3), P194–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17507589

Chen, Y., & Blanchard-Fields, F. (2000). Unwanted thought: age differences in the correction of social judgements. *Psychology and Aging*, *15*(3), 475–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11014710

Chiou, G.-L., & Anderson, O. R. (2010). A study of undergraduate physics students' understanding of heat conduction based on mental model theory and an ontology-process analysis. *Science Education*, *94*(5), 825–854. doi:10.1002/sce.20385

Chiu, M.-H., Chou, C.-C., & Liu, C.-J. (2002). Dynamic processes of conceptual change: Analysis of constructing mental models of chemical equilibrium. *Journal of Research in Science Teaching*, *39*(8), 688–712. doi:10.1002/tea.10041

Clariana, R. B., & Taricani, E. M. (2010). The consequence of increasing the number of terms used to score open-ended concept maps. *International Journal of Instructional Media*, *37*(2), 163–174.

Copeland, D. E., & Radvansky, G. A. (2007). Aging and integrating spatial mental models. *Psychology and Aging*, *22*(3), 569–79. doi:10.1037/0882-7974.22.3.569

Curtis, M. B., & Davis, M. a. (2003). Assessing knowledge structure in accounting education: An application of Pathfinder Associative Networks. *Journal of Accounting Education*, *21*(3), 185–195. doi:10.1016/S0748-5751(03)00024-1

d'Apoilonia, S. T., Charles, E. S., & Gary, M. (2004). Acquisition of complex systemic thinking: Mental models of evolution*. *Educational Research and Evaluation*, *10*(4-6), 499–521.

Devolder, P. A., & Pressley, M. (1992). Causal attributions and strategy use in relation to memory performance differences in younger and older adults. *Applied Cognitive psychology1*, *6*, 629–642.

Dorsey, D. W., Campbell, G. E., Foster, L. L., & Miles, D. E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance*, *12*(1), 31–57.

Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review*, *14*(1), 3–29. doi:10.1002/(SICI)1099-1727(199821)14:1<3::AID-SDR140>3.0.CO;2-K

Dunlosky, J., Baker, J. M. C., Rawson, K. a, & Hertzog, C. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging*, *21*(2), 390–400. doi:10.1037/0882-7974.21.2.390

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory and Cognition*, *25*(5), 691–700.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, *15*(3), 462–474. doi:10.1037//0882-7974.15J.4

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: the relative utility of concurrent versus retrospective reports. *Memory & Cognition*, *29*(2), 247–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11352207

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*(2), 389–400. doi:10.1037/0012-1649.41.2.389

Dunlosky, J., Kubat-Silman, A. K., & Hertzog, C. (2003). Effects of aging on the magnitude and accuracy of quality-of-encoding judgments. *The American Journal of Psychology*, *116*(3), 431–54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14503394

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: US: Education Testing Service.

Frank, D. J., Touron, D. R., & Hertzog, C. (2013). Age Differences in Strategy Shift: Retrieval Avoidance or General Shift Reluctance? *Psychology and Aging*, *28*(3), 778–788. doi:10.1037/a0030473

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*(1), 47–61. doi:10.1016/j.learninstruc.2011.06.002

Gaschler, R., & Frensch, P. a. (2007). Is information reduction an item-specific or an item-general process? *International Journal of Psychology*, *42*(4), 218–228. doi:10.1080/00207590701396526

Gilbert, D. K., Rogers, W. A., & Samuelson, M. E. (2004). Long-term retention of a spatial mental model for younger and older adults. *Esperimental Aging Research*, *30*, 217–224.

Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the life span. *Psychology and Aging*, *9*(3), 356–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7999321

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, *83*(1), 88–96. doi:10.1037//0022-0663.83.1.88

Gomez, R. L., Hadfield, O. D., & Housner, L. D. (1996). Conceptual maps and simulated teaching episodes as indicators of competence in teaching elementary mathematics. *Journal of Educational Psychology*, *88*(3), 572–585. doi:10.1037//0022-0663.88.3.572

Gonzalvo, P., Cafias, J. J., & Bajo, M. (1994). Structural representations in knowledge acquisition. *Journal of Educational Psychology*, *86*(4), 601–616.

Green, H. J., Lemaire, P., & Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta Psychologica*, *125*(3), 257–78. doi:10.1016/j.actpsy.2006.08.001

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*(3), 304–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8660787

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 172–190. doi:10.1037//0278-7393.25.1.172

Harada, E. T., Mori, K., & Taniue, N. (2010). Cognitive aging and the usability of IT-based equipment: Learning is the key. *Japanese Psychological Research*, *52*(3), 227–243. doi:10.1111/j.1468-5884.2010.00440.x

Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, *38*(6), 771–84. doi:10.3758/MC.38.6.771

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging*, *17*(2), 209–225. doi:10.1037//0882-7974.17.2.209

Hertzog, C., Mcguire, C. L., & Lineweaver, T. T. (2010). Aging, attributions, perceived control, and strategy use in a free recall task aging, attributions, perceived control, and strategy use in a free recall task*. *Aging, Neuroscience, and Cognition*, *5*(2), 85–106.

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology*, *46*(4), 939–48. doi:10.1037/a0019812

Hertzog, C., & Touron, D. R. (2011). Age differences in memory retrieval shift: Governed by feeling-of-knowing? *Psychology and Aging*, *26*(3), 647–60. doi:10.1037/a0021875

Hertzog, C., Touron, D. R., & Hines, J. C. (2007). Does a time-monitoring deficit influence older adults' delayed retrieval shift during skill acquisition? *Psychology and Aging*, *22*(3), 607–24. doi:10.1037/0882-7974.22.3.607

Hess, T. M., Follett, K. J., & McGee, K. A. (1998). Aging and impression formation: the impact of processing skills and goals. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *53*(3), 175–187. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9602833

Hines, J. C., Hertzog, C., & Touron, D. R. (2015). Younger and older adults weigh multiple cues in a similar manner to generate judgments of learning. *Aging, Neuropsychology, and Cognition*, (April), 1–19. doi:10.1080/13825585.2015.1028884

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, *24*(2), 462–75. doi:10.1037/a0014417

Jacobs-lawson, J. M., & Hershey, D. A. (1994). Concept maps as an assessment tool in psychology courses. *Methods and Techniques*, *29*(1), 2002.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. doi:10.1111/lang.12034

Johnson, P. J., Goldsmith, T. E., & Teague, K. W. (1994). Locus of the predictive advantage in Pathfinder-based representations of classroom knowledge. *Journal of Educational Psychology*, *86*(4), 617–626. doi:10.1037//0022-0663.86.4.617

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*(10), 434–442. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11707382

Kieras, D. E., & Bovair, S. (1983). The role of a mental model in learning to operate a device*. *Cognitive Science*, *273*(1964), 255–273.

Kim, M. K. (2012). Cross-validation study of methods and technologies to assess mental models in a complex problem solving situation. *Computers in Human Behavior*, *28*(2), 703–717. doi:10.1016/j.chb.2011.11.018

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning biases. *Psychology and Aging*, *15*(3), 400–416. doi:10.1037//0882-7TO.15.3.400

Kuhlmann, B. G., & Touron, D. R. (2011). Older adults' use of metacognitive knowledge in source monitoring: spared monitoring but impaired control. *Psychology and Aging*, *26*(1), 143–9. doi:10.1037/a0021055

Lemaire, P. (2010). Cognitive Strategy Variations During Aging. *Current Directions in Psychological Science*, *19*(6), 363–369. doi:10.1177/0963721410390354

Lemaire, P., Arnaud, L., & Lecacheur, M. (2004). Adults' age-related differences in adaptivity of strategy choices: Evidence from computational estimation. *Psychology and Aging*, *19*(3), 467–81. doi:10.1037/0882-7974.19.3.467

Lemaire, P., & Lecacheur, M. (2007). Aging and numerosity estimation. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *62*(6), P305–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18079414

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2000). *SAS system for mixed models* (4th ed.). Cary, NC: SAS Institute.

Lovett, M. C., & Schunn, C. D. (1999). Task representations, strategy variability, and base-rate neglect. *Journal of Experimental Psychology: General*, *128*(2), 107–130. doi:10.1037//0096-3445.128.2.107

Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). The concept map as a research and evaluation tool: Further evidence of validity. *Journal of Research in Science Teaching*, *31*(1), 91–101. doi:10.1002/tea.3660310109

Masunaga, H., & Horn, J. (2001). Expertise and age-related changes in components of intelligence. *Psychology and Aging*, *16*(2), 293–311. doi:10.1037//0882-7974.16.2.293

Matvey, G., Dunlosky, J., Shaw, R. J., Parks, C., & Hertzog, C. (2002). Age-related equivalence and deficit in knowledge updating of cue effectiveness. *Psychology and Aging*, *17*(4), 589–597. doi:10.1037//0882-7974.17.4.589

Mitzner, T. L., Touron, D. R., Rogers, W. A., & Hertzog, C. (2010). Checking it twice: Age-related differences in double checking during visual search. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting* (pp. 1326–1330). Santa Monica, CA.

Morrow, D. G., Leirer, V., Altieri, P., & Fitzsimmons, C. (1994a). Age differences in creating spatial models from narratives. *Language and Cognitive Processes*, *9*(2), 203–220.

Morrow, D. G., Leirer, V., Altieri, P., & Fitzsimmons, C. (1994b). When expertise reduces age differences in performance. *Psychology and Aging*, *9*(I), 134–148.

Morrow, D. G., Miller, L. M. S., Heather, E., Magnor, C., Fischer, U. M., & Kokayeff, N. K. (2009). Expertise and age differences in pilot decision making. *Aging, Neuropsychology, and Cognition*, *16*, 33–55.

Morrow, D. G., Ridolfo, H. E., Menard, W. E., Sanborn, A., Stine-morrow, E. A. L., Magnor, C., … Bryant, D. (2003). Environmental support promotes expertise-based mitigation of age differences on pilot communication tasks. *Psychology and Aging*, *18*(2), 268–284. doi:10.1037/0882-7974.18.2.268

Muis, K. R. (2007). The Role of Epistemic Beliefs in Self-Regulated Learning. *Educational Psychologist*, *42*(3), 173–190. doi:10.1080/00461520701416306

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. doi:10.1037//0278-7393.26.5.1170

Nelson, T., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. *The Psychology of Learning and Motivation: Advances in Research and Theory*, *26*, 125–173.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, *27*(10), 937–949. doi:10.1002/tea.3660271003

Novak, J. D., & Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool⋆. *Information Visualization*, *5*(3), 175–184. doi:10.1057/palgrave.ivs.9500126

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychology and Aging*, *23*(1), 104–18. doi:10.1037/0882-7974.23.1.104

Poindexter, M. T., & Clariana, R. B. (2006). The influence of relational and proposition-specifc processing on structural knowledge and traditional learning outcomes. *International Journal of Instructional Media*, *33*(2), 177–187.

Price, J., Hertzog, C., & Dunlosky, J. (2008). Age-related differences in strategy knowledge updating: Blocked testing produces greater improvements in metacognitive accuracy for younger than older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *15*(5), 601–26. doi:10.1080/13825580801956225

Price, J., Hertzog, C., & Dunlosky, J. (2010). *Self-regulated learning in younger and older adults: Does aging affect metacognitive control? Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition* (Vol. 17). doi:10.1080/13825580903287941

Radvansky, G. A. (1999a). Aging, Memory, and Comprehension. *Current Directions in Psychological Science*, *8*(2), 49–53. doi:10.1111/1467-8721.00012

Radvansky, G. A. (1999b). Memory retrieval and suppression: the inhibition of situation models. *Journal of Experimental Psychology. General*, *128*(4), 563–79. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10650586

Radvansky, G. A., Copeland, D. E., Berish, D. E., & Dijkstra, K. (2003). Aging and situation model updating. *Aging, Neuropsychology, and Cognition*, *10*(2), 158–166.

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review*, *14*(6), 1027–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18229472

Radvansky, G. A., Gerard, L. D., Zacks, R. T., & Hasher, L. (1990). Younger and older adults' use of mental models as representations for text materials. *Psychology and Aging*, *5*(2), 209–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2378686

Radvansky, G. A., Zacks, R. T., & Hasher, L. (1996). Fact retrieval in younger and older adults: the role of mental models. *Psychology and Aging*, *11*(2), 258–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8795054

Rawson, K. A., & Touron, D. R. (2009). Age differences and similarities in the shift from computation to retrieval during reading comprehension. *Psychology and Aging*, *24*(2), 423–37. doi:10.1037/a0016044

Rice, D. C., Ryan, J. M., & Samson, S. M. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? *Journal of Research in Science Teaching*, *35*(10), 1103–1127. doi:10.1002/(SICI)1098-2736(199812)35:10<1103::AID-TEA4>3.0.CO;2-P

Rogers, W. A., & Gilbert, D. T. (1997). Do performance strategies mediate age-related differences in associative learning? *Psychology and Aging*, *12*(4), 620–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9416631

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*(3), 349–363. doi:10.1037//0033-2909.100.3.349

Rowe, A. L., & Cooke, N. J. (1995). Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quarterly*, *6*(3), 243–255. doi:10.1002/hrdq.3920060303

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8759042

Schnotz, W., & Preuss, A. (1997). Task-dependent construction of mental models as a basis for conceptual change. *European Journal of Psychology of Education*, *12*(2), 185–211.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: "Like a bridge over troubled water." *Instructional Science*, *29*, 403–427.

Spieler, D. H., Mayr, U., & LaGrone, S. (2006). Outsourcing cognitive control to the environment: adult age differences in the use of task cues. *Psychonomic Bulletin & Review*, *13*(5), 787–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17328374

Staggers, N., & Norica, A. F. (1993). Mental models: Concepts for human-computer interaction research. *International Journal of Man-Machine Studies*, *38*, 587–605.

Stahl, E., Pieschl, S., & Bromme, R. (2006). Task complexity, epistemological beliefs and metacognitive calibration: An exploratory study. *Journal of Educational Computing Research*, *35*(4), 319–338. doi:10.2190/1266-0413-387K-7J51

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, *25*(2), 377–90. doi:10.1037/a0018022

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition*, *32*(2), 696–710. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18323066

Stine-Morrow, E. A. L., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin*, *132*(4), 582–606. doi:10.1037/0033-2909.132.4.582

Stine-Morrow, E. A. L., Morrow, D. G., & Leno, R. (2002). Aging and the representation of spatial situations in narrative understanding. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *57*(4), P291–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12084780

Strayer, D. L., & Kramer, A. F. (1994). Aging and skill acquisition: learning-performance distinctions. *Psychology and Aging*, *9*(4), 589–605. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7893430

Taricani, E. M., & Clariana, R. B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research and Development*, *54*(1), 65–82. doi:10.1007/s11423-006-6497-z

Touron, D. R. (2006). Are item-level strategy shifts abrupt and collective? Age differences in cognitive skill acquisition. *Psychonomic Bulletin & Review*, *13*(5), 781–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17328373

Touron, D. R., & Hertzog, C. (2004a). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging*, *19*(3), 452–66. doi:10.1037/0882-7974.19.3.452

Touron, D. R., & Hertzog, C. (2004b). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition*, *32*(2), 298–310. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15190721

Touron, D. R., & Hertzog, C. (2009). Age differences in strategic behavior during a computation-based skill acquisition task. *Psychology and Aging*, *24*(3), 574–85. doi:10.1037/a0015966

Touron, D. R., Hertzog, C., & Frank, D. (2012). Eye movements and strategy shift in skill acquisition: Adult age differences. *Journal of Gerontology, Psychological Sciences*, *66B*(2), 151–159. doi:10.1093/geronb/gbq076.

Touron, D. R., Swaim, E. T., & Hertzog, C. (2007). Moderation of Older Adults ' Retrieval Reluctance Through Task Instructions an ... *Journal of Gerontology, Psychological Sciences*, *62B*(3), 149–155.

Van Boven, L., & Thompson, L. (2003). A look into the mind of the negotiator: Mental models in negotiation. *Group Processes & Intergroup Relations*, *6*(4), 387–404. doi:10.1177/13684302030064005

Wallace, J. D., & Mintzes, J. J. (1990). The concept map as a research tool: Exploring conceptual change in biology. *Journal of Research in Science Teaching*, *27*(10), 1033–1052. doi:10.1002/tea.3660271010

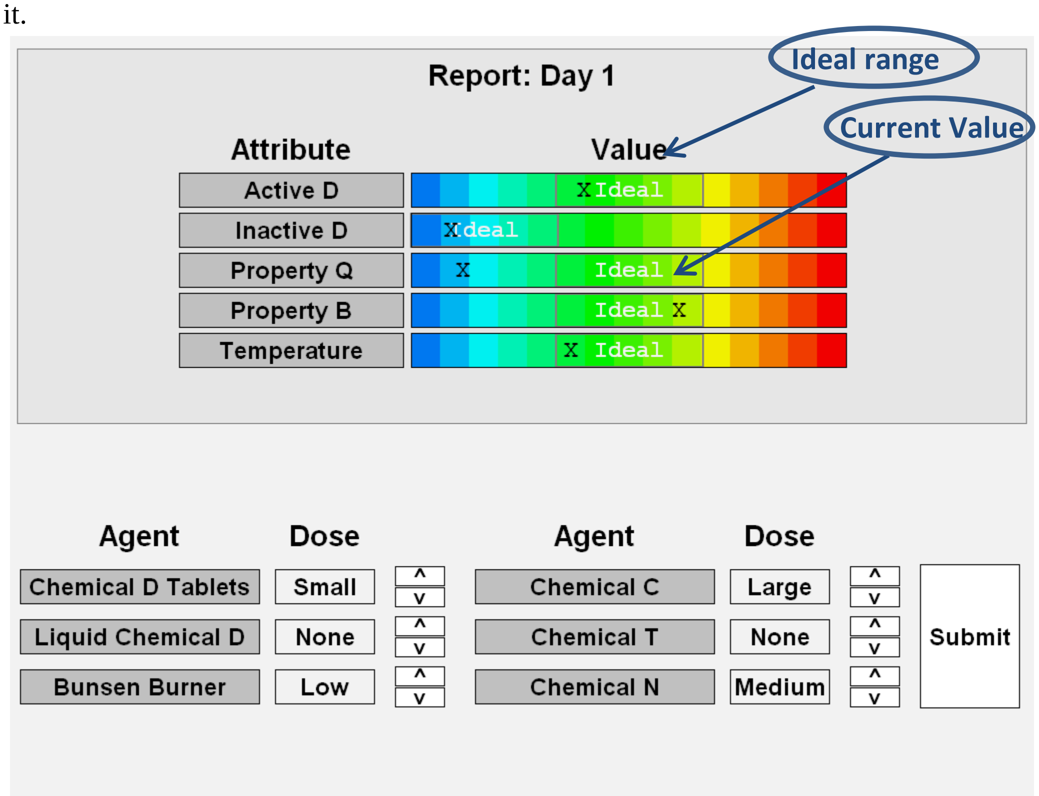
Winne, P. H., & Hadwin, A. (1998). Studying as self-regulated learning. In J. Dunlosky, D. J. Hacker, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaumm Associates.

APPENDIX A

STUDY 1 NOVEL CHEMISTRY TASK INSTRUCTIONS

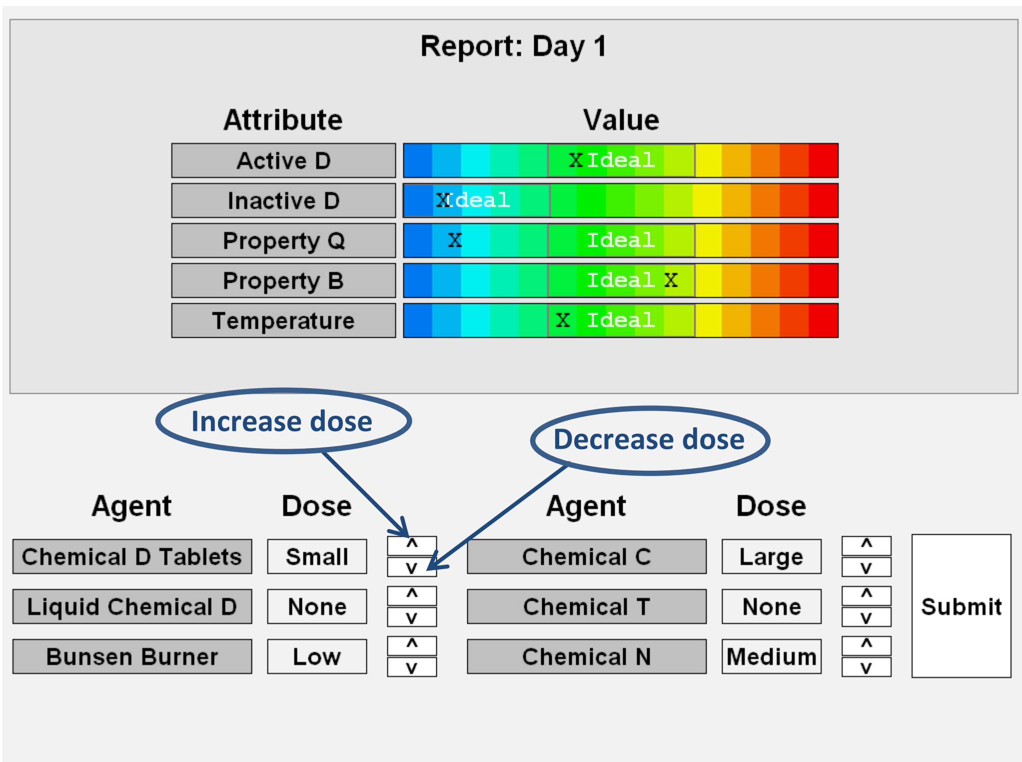
Instructions

In this study you will play the role of chemist. Your job is to keep various attributes of a solution within specified ideal ranges over a series of hypothetical days. For each "day" in the task you will receive a report containing the current levels of each chemical attribute indicated by an "X." The report is color coded and the ideal range marked by a gray box with the word "ideal" inside it.



(Press the RIGHT ARROW to continue)

When any attribute gets too high or too low, you will need to decide what to do to return that attribute to its optimal level. These actions may include exposing the solution to heat or adding various chemicals to the solution. Each chemical can be added to the solution in one of three doses: small, medium, or large and the Bunsen burner can be turned off or set to either a low or high level. You can click on the up and down arrows next to a chemical to change the size of the dose for that chemical. Changes can only be made to the solution once per day. After deciding what to do for a given day, you will click the "Submit" button to advance to the next day of the task.

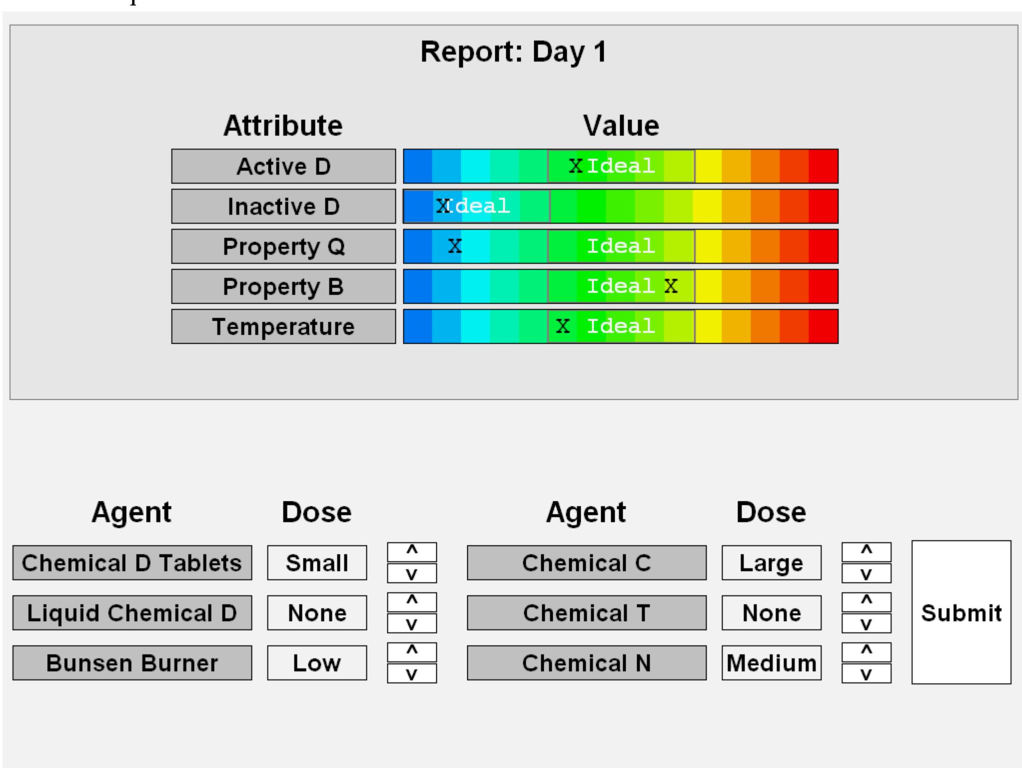


(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Next, you will read information regarding the attributes of the solution that need to be monitored and how they can be adjusted up or down.
You will have to pass a quiz at the end of the instructions to continue to the next part of the experiment. Most people take a few tries to learn all the information, so don't be discouraged if it takes you a few tries as well.

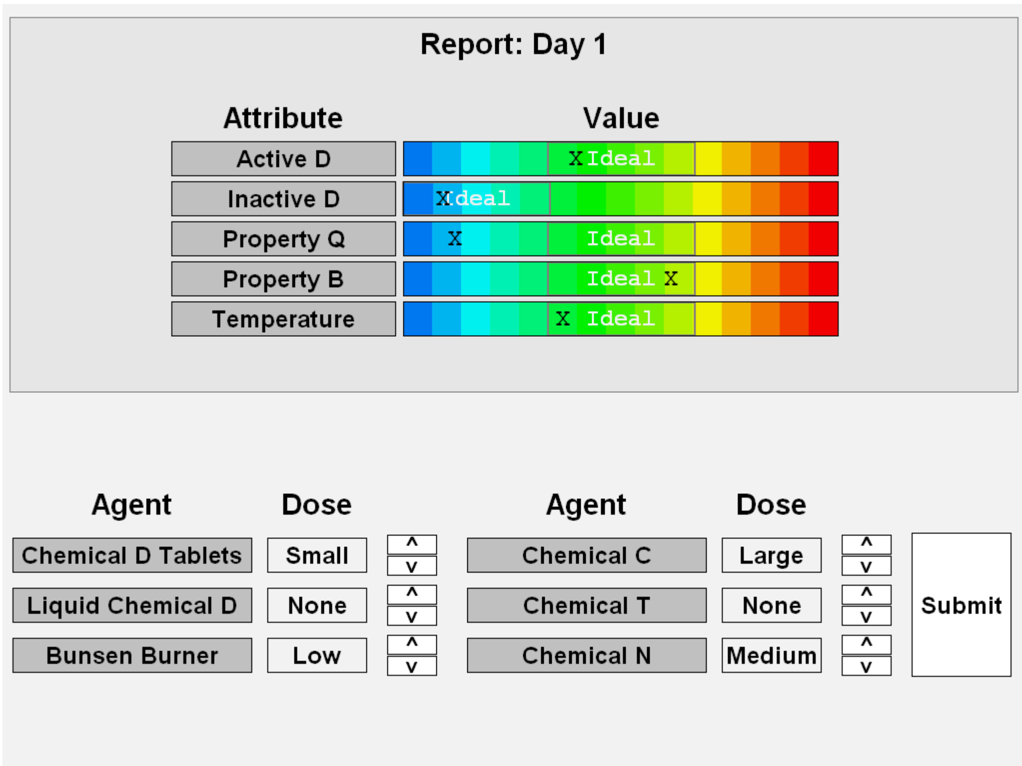
(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Active D
Bacteria will grow in the solution if the temperature or chemical composition of the solution is not ideal. To combat this, a constant level of disinfectant must be maintained in the solution. Active D shows the level of disinfectant available to keep the solution sanitary. It is important that you do not allow Active D to get too low, or you run the risk of bacterial growth in the solution (explained on a later screen).
Active D is converted into Inactive D (explained on a later screen) as it breaks down and kills bacteria in the solution. As a result, the level of Active D decreases a little each time bacteria enter the solution. The more bacteria enter the solution, the more drastic the decrease in the level of Active D and the more drastic the increase in Inactive D. Active D will also gradually evaporate from the solution. High temperatures will increase the rate of evaporation. Therefore, you must regularly replenish Active D to prevent the growth of bacteria.
Active D can be maintained by adding a small dose of Chemical D Tablets to the solution daily. Active D can also be increased moderately by adding a medium or large dose of Chemical D Tablets to the solution. Active D can be increased drastically by adding a small, medium, or large dose of Liquid Chemical D to the solution.



(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Inactive D
As Active D kills bacteria it is converted into Inactive D. Inactive D cannot help break down and kill new bacteria. As bacteria enter the solution, the level of Active D decreases and the level of Inactive D increases, as Active D is converted into Inactive D. High levels of Inactive D produce toxic fumes. As a result Inactive D should be kept as low as possible (note the lower ideal range). As with Active D, Inactive D will gradually evaporate from the solution, and high temperatures will increase the rate of evaporation.
Inactive D will normally remain low if you maintain an appropriate Active D level.



(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Bacterial growth

Only small amounts of bacteria are likely to enter the solution at a given time. Therefore, when Active D is in the ideal range, all bacteria that enter the solution are killed before they can multiply. As a result Active D decreases slightly and Inactive D increases slightly when bacteria enter the solution. Inactive D then evaporates from the solution within a day or two, bringing the Inactive D level back down to the ideal range. However, you may need to add additional Chemical D tablets if Active D gets too low.

High temperatures will facilitate the growth of bacteria. When the level of Active D is low or the temperature is too high, Active D is unable to kill all the bacteria before they begin to multiply. This results in bacterial growth in the solution.

Once bacteria begin to grow in the solution, Chemical D Tablets alone will not increase the Active D level of the solution. Instead the Active D released by the tablets will be instantly converted into Inactive D as it kills some, but not all of the bacteria. This rapidly results in very low levels of Active D and very high levels of Inactive D.

Adding a large dose of Liquid Chemical D to the solution will provide enough Active D to kill all of the bacteria, ending the bacterial growth in the solution.

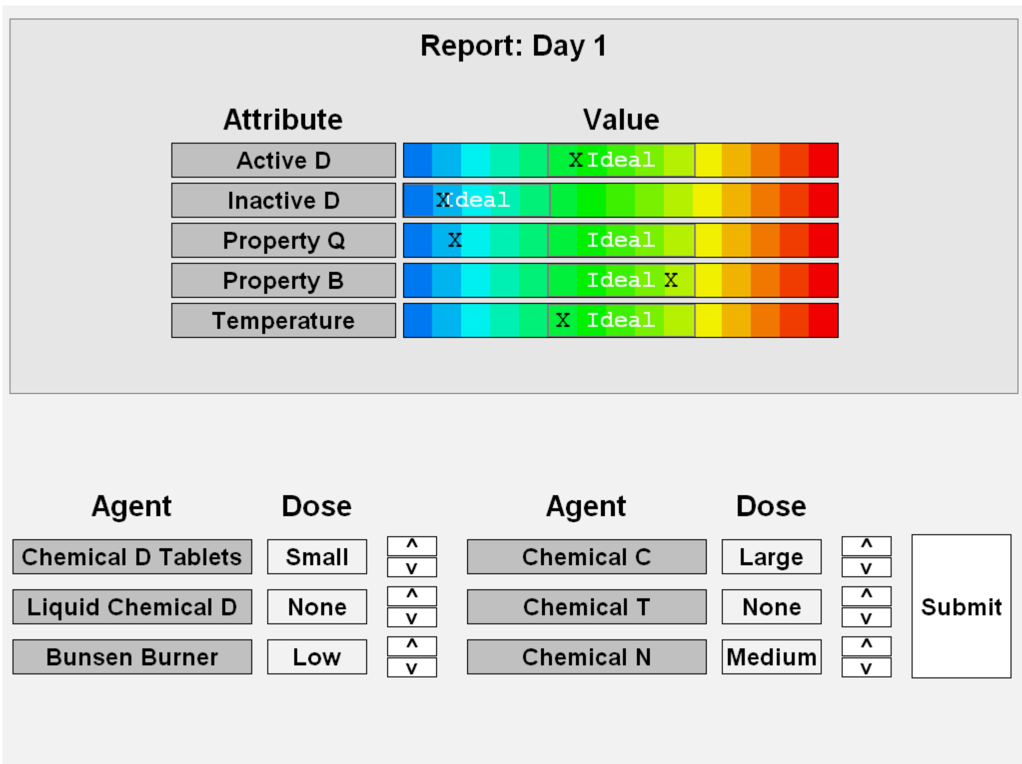
Therefore, low Active D levels and high Inactive D levels indicate bacterial growth in the solution, which should be treated by adding Liquid Chemical D to the solution.

Because Inactive D naturally evaporates from the solution, the Inactive D level will gradually decrease after the bacterial growth has ended. Temporarily increasing the solution temperature can speed the evaporation of Inactive D.

(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Property Q

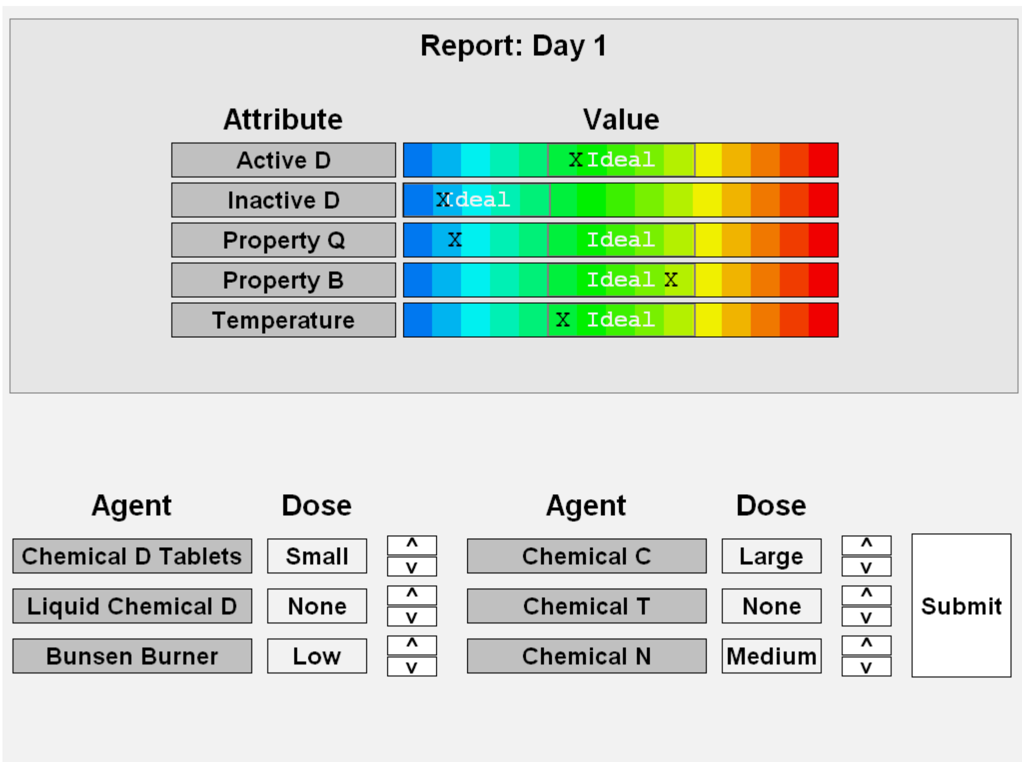
Property Q must be kept within the ideal range to avoid damage to the solution container. Low Property Q will cause corrosion of the solution container. High Property Q will cause calcium deposits to form on the solution container. Property Q can be increased by adding either Chemical C or Chemical T to the solution. Property Q can be decreased by adding Chemical N to the solution. How much Property Q changes with each dose of Chemical C, T, or N will depend on the level of Property B (explained on the next screen).



(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Property B

Property B is a chemical buffer for Property Q. A chemical buffer controls how easily another chemical attribute changes.
When the level of Property B is ideal, Property Q levels will remain relatively stable and a medium dose of Chemical C, T, or N will produce a moderate change in Property Q.
At low levels of Property B, Property Q fluctuates drastically and even a small dose of Chemical C, T, or N will produce a drastic change in Property Q. That is, there is nothing to buffer the changes in Property Q.
At high levels of Property B, a large dose of Chemical C, T, or N is needed to produce a moderate increase or moderate decrease in Property Q. Like Property Q, high levels of Property B will also cause calcium deposits to form on the solution container.

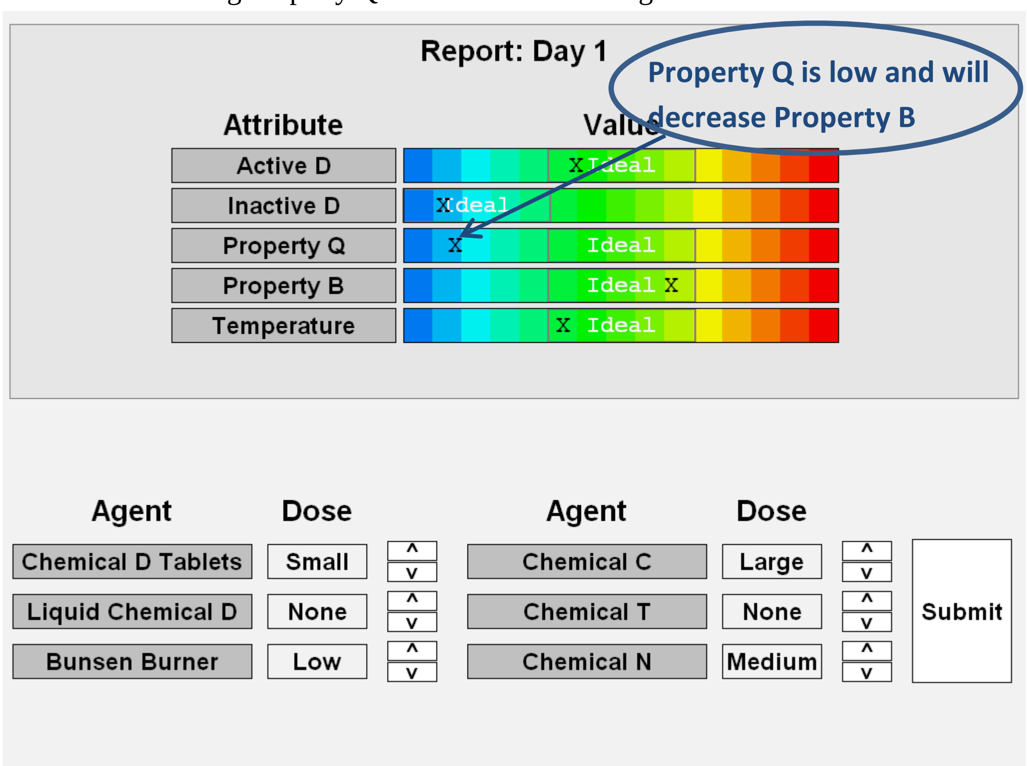


(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Adjusting Properties Q and B

When Property B and Q are both outside the ideal range, you should always adjust Property B before attempting to adjust Property Q on the following day. This is because Property Q will be difficult to keep in the ideal range without the proper level of Property B present to buffer changes in Property Q.

You can increase Property B slightly, moderately, or drastically, using a comparable (small, medium, or large) dose of Chemical T. Note that Chemical T increases both Property Q and B. Therefore, if you increase Property B with Chemical T, Property Q will also increase. By contrast, Property Q can instead be increased with Chemical C without altering Property B. Decreasing Property B is more complex. You can only decrease Property B by first decreasing Property Q to a very low level using Chemical N (see image below). When Property Q is low, Property B will begin to decrease. When Property B reaches the ideal range, you will then add Chemical C to bring Property Q back into its ideal range.



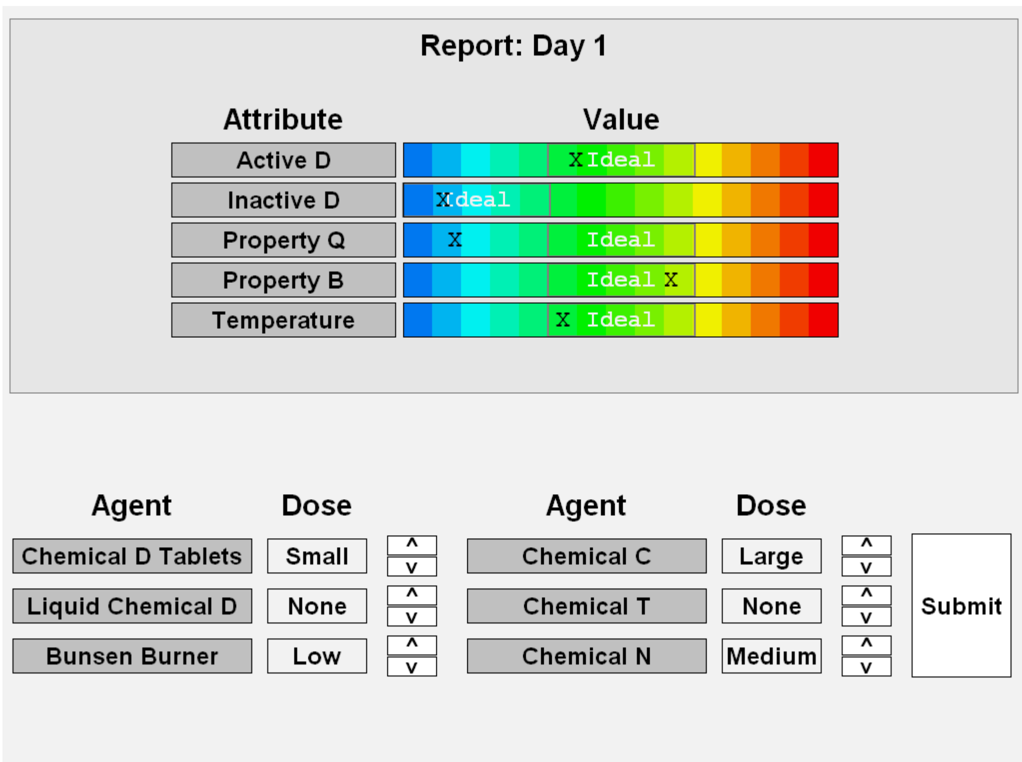
(Press the LEFT ARROW to go back)
(Press the RIGHT ARROW to continue)

Temperature

To maintain stable levels of Active D, the temperature of the solution should be higher than the temperature in the lab.

Because the solution will cool gradually over time, temperature can be increased by turning up the Bunsen burner. Turning the Bunsen burner off will allow the temperature of the solution to decrease. Turning the Bunsen burner on low or high will increase the temperature moderately or drastically (respectively).

High temperatures can cause Active and Inactive D to evaporate from the solution more rapidly. High temperatures also facilitate the growth of bacteria in the solution. At high temperatures, bacteria may begin to multiply before Active D kills them. Temperature has no impact on Property Q or Property B and does not alter the effects of the chemicals added to the solution (Chemical D tablets, Liquid Chemical D, Chemicals C, T, and N).



(Press the LEFT ARROW to go back)
(Press the ENTER when you are done reading)

APPENDIX B

SWIMMING POOL TERMS AND THEIR NOVEL CHEMISTRY TASK ANALOGS

Swimming pool	Novel Chemistry task
Free chlorine	Active D
Combined chlorine	Inactive D
pH	Property Q
Alkalinity	Property B
Temperature	Temperature
Chlorine tablets	Chemical D tablets
Liquid chlorine	Liquid chemical D
Pool heater	Bunsen burner
Chemical C	Borax
Chemical T	Sodium bicarbonate
Chemical N	Muriatic acid

APPENDIX C

STUDY 1 DECLARATIVE KNOWLEDGE QUIZ ITEMS

This ONE chemical can be used to maintain or moderately increase Active D.

This ONE property needs to be maintained in the solution to kill bacteria.

This ONE chemical can be used to drastically increase Active D.

This ONE property will cause toxic fumes if it gets too high.

This ONE property increases when bacteria contaminate the solution.

This ONE chemical is added to eliminate bacteria AFTER they begin to grow in the solution.

These TWO properties will evaporate from the solution at high temperatures.

Either of these TWO chemicals can be used to increase Property Q.

This ONE chemical can be used to decrease Property Q.

This ONE property will cause corrosion to form on the solution container if it gets too low.

These TWO properties will cause calcium deposits to form on the solution container if either gets too high.

This ONE property acts as a chemical buffer for Property Q.

This ONE chemical can be used to increase Property B.

This ONE property must be lowered to decrease Property B.

This ONE property should always be adjusted before adjusting Property Q.

Response options

Property B

Chemical C

Chemical D Tablets

Inactive D

Active D

Liquid chemical D

Chemical N

Property Q

Chemical T

APPENDIX D

STUDY 1 INFERENCE TEST ITEMS

Question	Options			
Which chemical has the opposite effects as Chemical T?	A) Chemical N	B) Chemical C	C) Liquid Chemical D	D) Chemical D tablets
Your solution is being used in an industrial facility where temperatures can be in excess of 200 degrees. What should you recommend to the facility employee in charge of purchasing chemicals?	A) Purchase large quantities of Chemical C and T	B) Purchase large quantities of Chemical N	C) Purchase large quantities of Chemical D Tablets and Liquid Chemical D	D) Purchase industrial sized heaters to heat the solution
When Properties B and Q are low, what should be done to increase Property Q?	A) Add liquid Chemical D	B) Add Chemical C	C) Add Chemical T	D) Add Chemical N
When Property B is high and Property Q is low, what should be done to the solution?	A) Wait until Property B is lower then add Chemical C	B) Add Chemical C and then add Chemical T	C) Add Chemical C then add Chemical N	D) Add Chemical T then add Chemical N
When Active D is low and Inactive D is high, what should be done to the solution?	A) Add Chemical D tablets	B) Add Liquid Chemical D	C) Turn on the Bunsen burner	D) Add Chemical T
A fellow chemist is having trouble adjusting property Q in his solution. What might be the cause of his problem?	A) Property B is too high or too low	B) Active D is too low	C) Inactive D is too high	D) The solution temperature is too low
If Inactive D and Active D are both extremely high, in addition to adding Liquid Chemical D, you should ____.	A) Add Chemical C	B) Add Chemical D Tablets	C) Add Chemical N	D) Turn on the Bunsen burner

Although it is not ideal to do so, you are asked to keep the solution at a lower temperature than usual. As a result of this procedure, you can expect to ____?	A) Use less Chemcial N and more Chemcial T	B) Experience more bacterial contamination	C) Use less Liquid Chemical D and fewer Chemical D Tablets	D) Use more Liquid Chemical D and more Chemical D Tablets
If Active D is normal and Inactive D is zero, what should be done to the solution?	A) Add Liquid Chemical D	B) Add Chemical D Tablets	C) Turn on the Bunsen Burner	D) Nothing
Another chemist designs a cover for the solution that helps prevent heat loss and evaporation. If you use this cover you could expect to also ______.	A) Use more Chemical D tablets	B) Use more Chemical N	C) Use fewer Chemical D tablets	D) Use less Chemical N
If the solution were kept in an environment where high levels of Chemical N were present and likely to contaminate the solution, then you would likely ____.	A) Frequently use the Bunsen burner	B) Frequently add Chemical T or C to the solution	C) Frequently add Chemical N to the solution	D) Frequently add Liquid Chemical D
Active D and Inactive D levels dropped drastically from Day 1 to Day 2. Which of the following is a possible cause of this effect?	A) Someone added too many Chemical D tablets to the solution	B) Someone added too much Liquid Chemical D to the solution	C) Someone added too much Chemical N to the solution	D) Someone left the Bunsen burner on
Yesterday someone added Liquid Chemical D to the solution and turned the Bunsen burner on high. What was the likely state of the solution that prompted this action?	A) Property Q was high	B) Property B was low	C) Inactive D was low and temperature was low	D) Active and Inactive D were high
Someone spilled Chemical C into your solution. What course of action should you take to counteract these effects?	A) Add Liquid Chemical D	B) Add Chemical N	C) Add Chemical T	D) Turn on the Bunsen burner

Someone accidentally added too much Liquid Chemical D to her solution. What should she do to counteract the effect?	A) Add Chemical D Tablets	B) Add Chemical N	C) Add Chemical C	D) Turn on the Bunsen burner
Active D is low and Inactive D is high. What is the likely cause of this situation?	A) The temperature is too high	B) Chemical D tablets were added to the solution	C) Property B is too low	D) Bacteria is contaminating the solution
Calcium deposits have begun to form on the solution container. What is the likely cause of this situation?	A) Property Q is too high	B) Property B is too high	C) Both property Q and B are too high	D) Any of the above could cause this situation
The solution container has begun to corrode. What is the likely cause of this situation?	A) Property Q is too low	B) Property B is too low	C) Both Property Q and B are too low	D) Any of the above could cause this situation
Bacteria have contaminated the solution. What action should you take?	A) Add Chemical D tablets	B) Add Liquid Chemical D	C) Add Chemical N	D) Add Chemical C
Property Q has been mostly stable over the last few days, drifting upwards only slightly. What can we assume about the solution?	A) There is no bacteria in the solution	B) The temperature is within the optimal range	C) Property B is in the optimal range	D) Active D is in the optimal range
Property Q was low so you added Chemical C. Then Property Q became extremely high so you added Chemical N. Now Property Q is extremely low. What should you do to resolve this situation?	A) Add Chemical C	B) Add Chemical N	C) Add Chemical T	D) Turn on the Bunsen burner
The lab has run out of Liquid Chemical D and Chemical D tablets. What could you do to prevent the growth of bacteria?	A) Keep the temperature on the low side of the optimal range	B) Keep Property Q on the high side of the optimal range	C) Keep Property B on the low side of the optimal range	D) Keep Inactive D on the high ends of the optimal range

APPENDIX E

STUDY 1 PATHFINDER TERMS

Active D

Inactive D

Property B

Chemical C

Chemical T

Property Q

Chemical N

Liquid Chemical D

Chemical D Tablets

Bunsen burner

Temperature

Evaporate

Bacteria

Disinfectant

Calcium deposits

Corrosion

Buffer

Chemical

Property

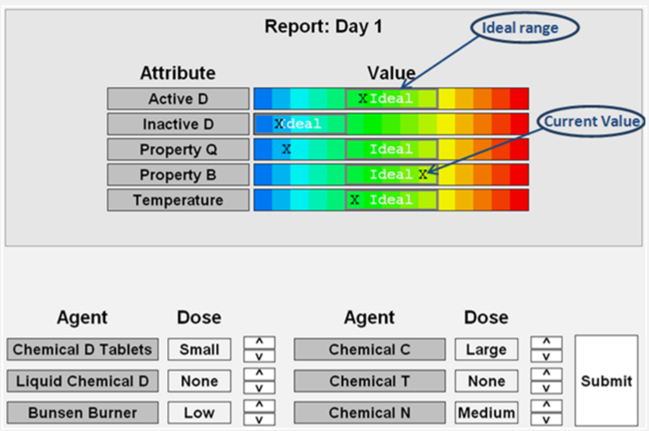
APPENDIX F

STUDY 2 NOVEL CHEMISTRY TASK INSTRUCTIONS

Instructions

In this study you will play the role of chemist. Your job is to keep various attributes of a solution within specified ideal ranges over a series of hypothetical days. For each "day" in the task you will receive a report containing the current levels of each attribute indicated by an "X." The report is color coded and the ideal range marked by a gray box with the word "ideal" inside it.

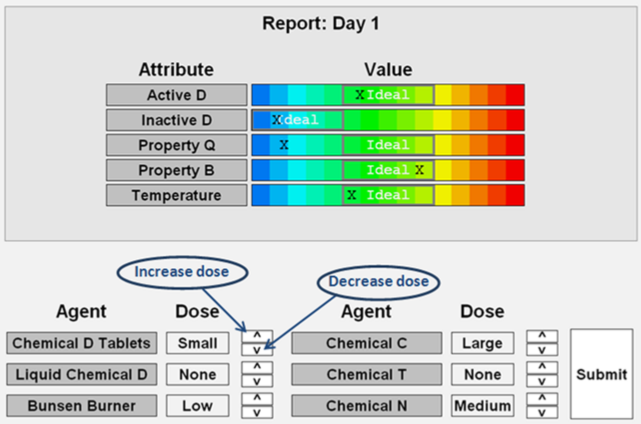
NOTE: Below is an example of the interface. We show you a similar example on each page. You do not need to click anything on the interface during this learning phase of today's experiment.



When any attribute gets too high or too low, you will need to decide what to do to return that attribute to its ideal level. These actions may include exposing the solution to heat or adding various chemical agents to the solution. Each chemical can be added to the solution in various doses: small, medium, or large and the Bunsen burner can be turned off or set to either a low or high level. You can click on the up and down arrows next to a chemical agent (or the Bunsen burner) to change the size of the dose for that chemical (or adjust the Bunsen burner setting).

Changes can only be made to the solution once per day. After deciding what to do for a given day, you will click the "Submit" button to apply those changes and see the results. Click the button again to advance to the next day of the task.

You should try to keep each attribute as close to the center of the ideal range as possible.



Next, you will read information regarding the attributes of the solution that need to be monitored.

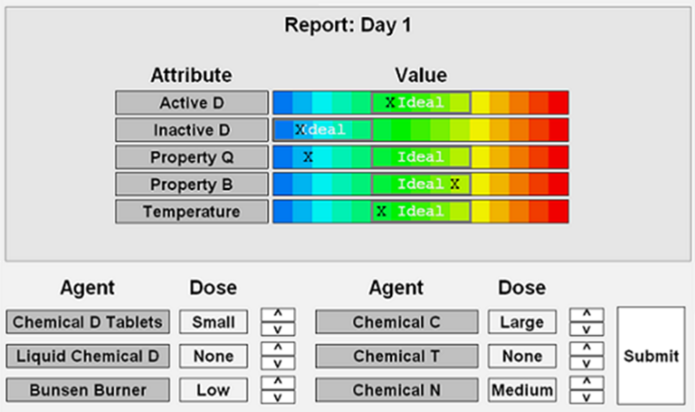
In some cases the instructions will explain how an attribute can be adjusted up or down. In other cases you will receive less complete instructions. When the instructions are incomplete, you will have to figure out how the various chemicals influence the solution when you later perform the chemistry task.

There will be a quiz at the end of the instructions, so please read carefully.

Active D

Bacteria will grow in the solution if the composition of the solution is not ideal. To combat this, a constant level of disinfectant must be maintained in the solution.

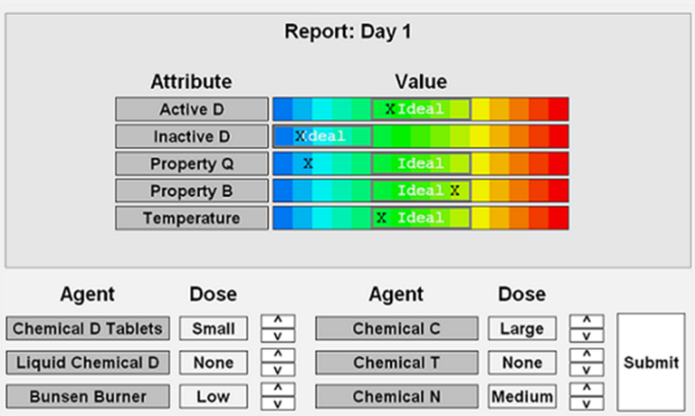
Active D shows the level of disinfectant available to keep the solution sanitary. It is important that you do not allow Active D to get too low, or you run the risk of bacterial growth in the solution (explained on a later screen).



Active D (continued)

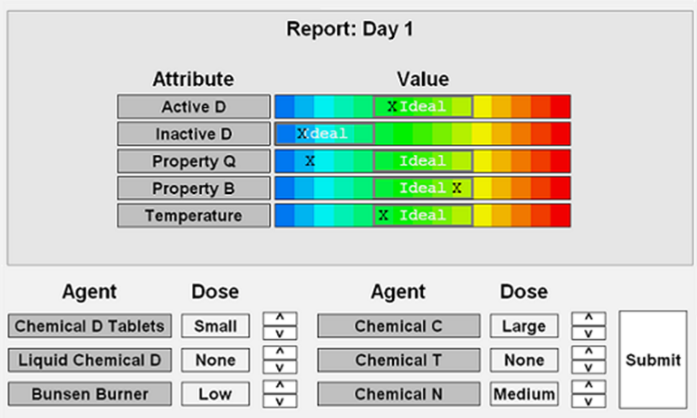
Active D is converted into Inactive D (explained on a later screen) as it breaks down and kills bacteria in the solution. As a result, the level of Active D decreases a little each time bacteria enter the solution. The more bacteria enter the solution, the more drastic the decrease in the level of Active D and the more drastic the increase in Inactive D. Therefore, you must regularly replenish Active D to prevent the growth of bacteria.

Active D can be maintained by adding a small dose of Chemical D Tablets to the solution regularly. Active D can also be increased moderately by adding a medium or large dose of Chemical D Tablets to the solution. Active D can be increased drastically by adding a small or large dose of Liquid Chemical D to the solution.



Inactive D

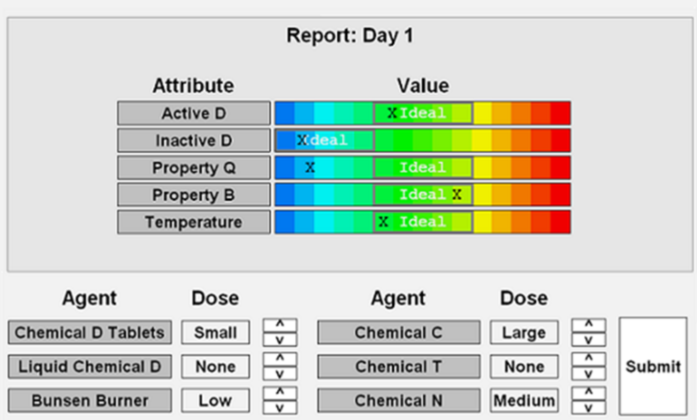
As Active D kills bacteria it is converted into Inactive D. Inactive D cannot help break down and kill new bacteria. As bacteria enter the solution, the level of Active D decreases and the level of Inactive D increases, as Active D is converted into Inactive D. High levels of Inactive D produce toxic fumes. As a result Inactive D should be kept as low as possible (note the lower ideal range). You will have to figure out how to lower Inactive D during the task.



Bacterial growth

Only small amounts of bacteria are likely to enter the solution at a given time. Therefore, when Active D is in the ideal range, all bacteria that enter the solution are killed before they can multiply. As a result Active D decreases slightly and Inactive D increases slightly when bacteria enter the solution. Inactive D then evaporates from the solution within a day or two, bringing the Inactive D level back down to the ideal range. However, you may need to add additional Chemical D tablets if Active D gets too low.

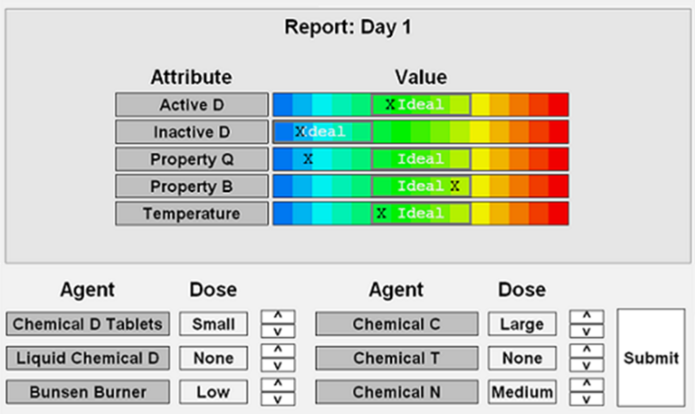
When the level of Active D is low, Active D is unable to kill all the bacteria before they begin to multiply. This results in bacterial growth in the solution.



Bacterial Growth (continued)

Once bacteria begin to grow in the solution, Chemical D Tablets alone will not increase the Active D level of the solution. Instead the Active D released by the tablets will be instantly converted into Inactive D as it kills some, but not all of the bacteria. This rapidly results in very low levels of Active D and very high levels of Inactive D.

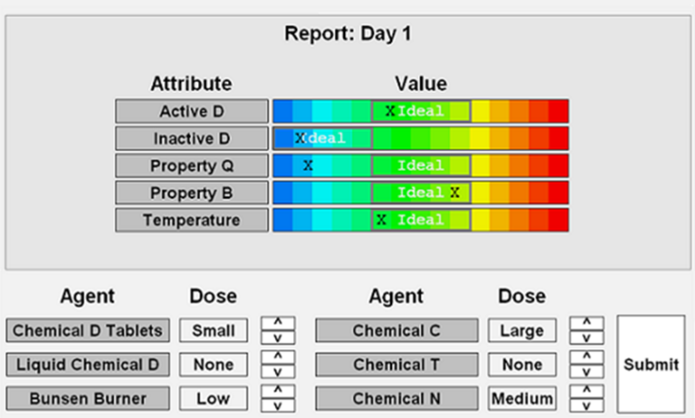
Adding a large dose of Liquid Chemical D to the solution will provide enough Active D to kill all of the bacteria, ending the bacterial growth in the solution. Therefore, low Active D levels and high Inactive D levels indicate bacterial growth in the solution, which should be treated by adding Liquid Chemical D to the solution.



Property Q

Property Q must be kept within the ideal range to avoid damage to the solution container. Low Property Q will cause corrosion of the solution container. High Property Q will cause calcium deposits to form on the solution container. Property Q can be changed using Chemicals T, C, and N. How much Property Q changes with each dose of Chemical C, T, or N will depend on the level of Property B (explained on the next screen).

It is up to you to figure out how Chemicals T, C, and N affect Property Q.

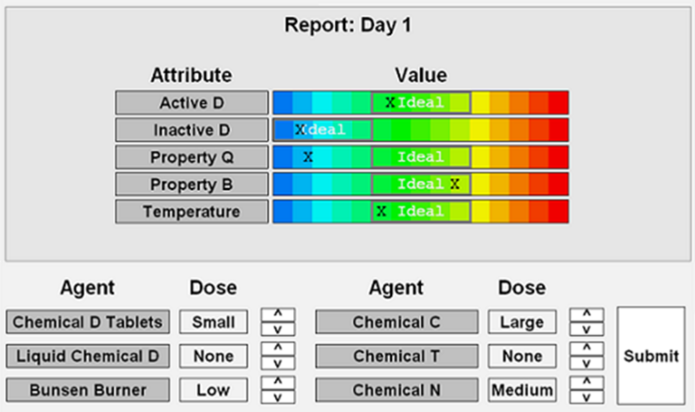


Property B

Property B is a chemical buffer for Property Q. A chemical buffer controls how easily another chemical attribute changes. When the level of Property B is ideal, Property Q levels will remain relatively stable and a medium dose of Chemical C, T, or N will produce a moderate change in Property Q.

At low levels of Property B, Property Q fluctuates drastically and even a small dose of Chemical C, T, or N will produce a drastic change in Property Q. That is, there is nothing to buffer the changes in Property Q.

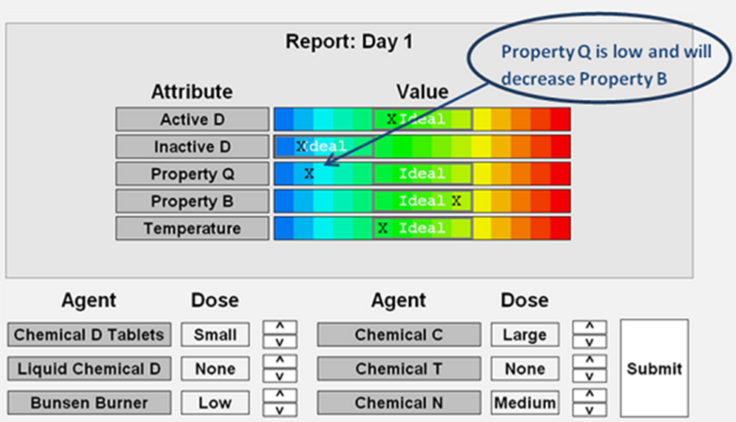
At high levels of Property B, a large dose of Chemical C, T, or N is needed to produce a moderate increase or moderate decrease in Property Q. Like Property Q, high levels of Property B will also cause calcium deposits to form on the solution container.



Adjusting Properties Q and B

Some chemical agent(s) will directly increase Property B. You will have to figure out which chemical(s) increase Property B.

Decreasing Property B is more complex. You can only decrease Property B by first decreasing Property Q to a very low level (see image below). When Property Q is low, Property B will begin to decrease. When Property B reaches the ideal range, you will then add chemicals to bring Property Q back into its ideal range.



Temperature

The solution must be kept at a higher than the temperature in the lab.

Because the solution will cool gradually over time, temperature can be increased by turning on the Bunsen burner. Turning the Bunsen burner off will allow the temperature of the solution to decrease. Turning the Bunsen burner on low or high will increase the temperature moderately or drastically (respectively).

Higher than "ideal" temperatures can affect some attributes of the solution. Thus applying heat to the solution may be necessary at times. You will have to figure out how temperature changes the solution.