Memory in a Messy Domain: Expertise and Memory for Mental Health Disorder Categories

Jessecae K. Marsh (jesseecae.marsh@yale.edu)
Department of Psychology, Yale University, Box 208205
New Haven, CT 06520 USA

Abstract
Experience within a domain has profound effects on the way a domain is conceptualized. These types of effects, such as experts showing strikingly better memory than novices, have been robustly demonstrated across a wealth of domains. Despite this tradition, the mental health field has had trouble demonstrating benefits for experience. Does this mean that expertise does not develop in mental health care as it does in other fields? To test this question, participants of varying experience in mental health were tested on their memory of patient information. Trainees were shown to have better memory for facts about patients than Novices and Experts. The findings of this study at the same time replicate the intermediate effect of the medical expertise literature while demonstrating findings unique to the domain of mental health.

Keywords: Expertise; memory; real-world categories.

Introduction
Mental health clinicians face a complex real-world domain in which to conduct categorization. The criteria for membership within a mental disorder are explicitly listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM), with membership in a given category met by possessing a certain percentage of the listed criterial features (American Psychiatric Association, 1994). The explicit, probabilistic category structure provided by the DSM contrasts greatly with the theory-like feature networks that have been described for lay people’s representations of everyday categories (e.g., Ahn, 1998). How does experience with the formal structure prescribed for mental health categories affect the way mental health professionals think about members of mental disorder categories? For example, how does experience with mental health categories affect the way clinicians remember information about patients? This paper addresses how experience alters the way clinicians think about mental disorder categories and, specifically, how this affects experienced clinicians’ memory for patient information. To begin, I will review the existing literature on the effects of experience on cognitive abilities within a domain.

A well-known finding within cognitive psychology is that highly experienced people within a domain show a pattern of behaviors that differ from novices in predictable and highly replicable ways. For example, an expert within a field has been shown to be quicker at solving problems in that domain, faster at recognizing meaningful patterns, and perhaps most convincingly demonstrated to have better memory for items related to their domain of expertise (e.g., Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Glaser & Chi, 1988). These benefits have been theorized to arise from experts having a more precise and better-organized representation of their category of expertise than novices (e.g., Lynch, Coley, & Medin, 2000). The robustness of this pattern of cognitive differences, or expertise effects, has been demonstrated in a myriad of varied domains such as chess (deGroot, 1966); musical abilities (Drake & Palmer, 2000); sports (Ward & Williams, 2003); and professions such as software design (Sonnenstag, 1998) and X-ray interpretation (Raufaste, Eyrolle, & Marine, 1998), to name a few. These repeated demonstrations leave little doubt that expertise effects are bound to be found in any given domain.

However, research within the mental health care profession has suggested that expertise may not similarly develop within the mental health domain. One line of research comparing mental health clinicians to laypeople in their ability to conduct professional duties such as therapy, has found no differences over experience. In a classic review of this type of early literature, Durlak (1979) found that of 42 previously published studies, the majority found no difference between laypeople and professionals, and almost all of the remainder found that laypeople actually outperformed professionals in providing therapy. Since this seminal review, similar conclusions of laypeople and professional equivalence in providing therapy have been reported (e.g., Beutler, Machado, & Neufeldt, 1994; Scogin, Bynum, Stephens, & Calhoon, 1990).

Do these findings related to clinical duties indicate that there likewise may be no changes in cognitive abilities as experience is gained within the mental health domain? In general, little work has addressed the cognitive effects of expertise within mental health. In one exception, Brailey, Vasterling, and Franks (2001) tested memory for patient features relevant to the diagnosis of a mental disorder. In another example, participants watched a videotape of a patient they were told had received diagnoses for both of the previously read DSM disorders and made evaluations during the viewing of whether the patient actually met criteria for either diagnosis. When afterwards asked to report what information they could remember that was relevant to the diagnostic process, clinicians showed better memory than graduate students for information about the patient that was diagnostically relevant to the two disorders they were evaluating. However, clinicians were also found to recall more of the DSM diagnostic criteria they read before viewing the patient.

This paper demonstrates that expertise effects in the mental health domain are as robust as in any other domain. A well-recognized finding within cognitive psychology is that experts show striking and replicable benefits in their memory for information related to their domain of expertise (e.g., Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Glaser & Chi, 1988). These benefits have been theorized to arise from experts having a more precise and better-organized representation of their category of expertise than novices (e.g., Lynch, Coley, & Medin, 2000). The robustness of this pattern of cognitive differences, or expertise effects, has been demonstrated in a myriad of varied domains such as chess (deGroot, 1966); musical abilities (Drake & Palmer, 2000); sports (Ward & Williams, 2003); and professions such as software design (Sonnenstag, 1998) and X-ray interpretation (Raufaste, Eyrolle, & Marine, 1998), to name a few. These repeated demonstrations leave little doubt that expertise effects are bound to be found in any given domain.

However, research within the mental health care profession has suggested that expertise may not similarly develop within the mental health domain. One line of research comparing mental health clinicians to laypeople in their ability to conduct professional duties such as therapy, has found no differences over experience. In a classic review of this type of early literature, Durlak (1979) found that of 42 previously published studies, the majority found no difference between laypeople and professionals, and almost all of the remainder found that laypeople actually outperformed professionals in providing therapy. Since this seminal review, similar conclusions of laypeople and professional equivalence in providing therapy have been reported (e.g., Beutler, Machado, & Neufeldt, 1994; Scogin, Bynum, Stephens, & Calhoon, 1990).

Do these findings related to clinical duties indicate that there likewise may be no changes in cognitive abilities as experience is gained within the mental health domain? In general, little work has addressed the cognitive effects of expertise within mental health. In one exception, Brailey, Vasterling, and Franks (2001) tested memory for patient features relevant to the diagnosis of a mental disorder. In another example, participants watched a videotape of a patient they were told had received diagnoses for both of the previously read DSM disorders and made evaluations during the viewing of whether the patient actually met criteria for either diagnosis. When afterwards asked to report what information they could remember that was relevant to the diagnostic process, clinicians showed better memory than graduate students for information about the patient that was diagnostically relevant to the two disorders they were evaluating. However, clinicians were also found to recall more of the DSM diagnostic criteria they read before viewing the patient.
than graduate students. Thus, it is unclear whether clinicians actually were better at remembering diagnostic patient information or were just better at remembering what were the relevant diagnostic criteria they should reference in that guided recall task. In addition, clinicians did not differ from graduate students in their ability to recall relevant patient information in separate memory tasks that did not specifically ask for diagnostic information. Therefore, it is still unclear how clinicians’ general memory for patients may differ from less experienced groups.

In order to consider the question of expertise within mental health more carefully we can use as a comparison point the extensive expertise findings of a profession that has close parallels to mental health: medicine. When looking at expert memory performance within the medical expertise literature, an interesting twist occurs. In line with previous expertise research, memory for patient descriptions was better in final year medical school students than in laypeople. However, surprisingly these intermediately trained students also recalled more patient information than experts (e.g., Claassen & Bosshuizen, 1985; Patel, Groen, & Frederiksen, 1986; Schmidt & Bosshuizen, 1993). This inverted U-shaped trend found within the medical expertise literature has been referred to as the intermediate effect, referencing intermediates’ superior memory performance.

The discussed literature suggests several possible findings for memory ability over experience in the mental health care profession. The general expertise literature suggests that as experience in mental health increases, memory for patient information should likewise increase. The medical expertise literature suggests that intermediate trainees should show the best memory performance, with experts recalling less than trainees. Finally, previous work in non-cognitive facets of mental health suggest expertise may not generally develop, with Brailey, et al.’s (2001) results suggesting that some memory-related benefits may arise for diagnostically-relevant information alone. The following experiment tests the influence of expertise on general memory within mental health. Participant groups varying in experience read the influence of expertise on general memory within mental health. The following experiment tests some memory-related benefits may arise for diagnostically-relevant information alone. The following experiment tests the influence of expertise on general memory within mental health. Participant groups varying in experience read

Participants

Novices. (N=21) Yale University undergraduates were used as Novices. Participants were screened to have no previous experience working in mental health care. The average age for novice participants was 19. No student had taken Abnormal Psychology and a majority (N=12) had not taken a psychology course beyond Introduction to Psychology. Novices completed the experiment either as part of an introductory course experimental requirement or for pay at a rate of $10 per hour.

Trainees. (N=20) Clinical psychology graduate students who had completed at least two years of graduate school were recruited as an intermediate experience (Trainee) participant group. The average age for Trainee participants was 27. Trainees had on average 3.6 years of graduate training, with a range of 3 to 5 years, and had seen patients for an average of 2.3 years. Every Trainee had been responsible for the individual treatment of patients and for diagnosing patients in a clinical setting. Trainees were compensated with a flat fee of $30.

Experts. (N=17) Professional mental health clinicians were recruited who matched the following two criteria: 1) the clinician had been licensed for at least 10 years as either a licensed clinical social worker (LCSW) (N=6), a psychiatrist (N=2), or a psychologist (N=9); and 2) the clinician specialized in treating either Substance Use Disorders OR Eating Disorders (without in-depth experience with the other type of disorder). The average age for the Expert group was 52. Experts reported on average 26 years of experience in seeing patients, with a range of 19 to 38 years. Experts were paid at a rate of $85 per hour.

Materials

Three different cases depicting mental health patients were used. Each case was presented as a transcription of an interview between a patient and a therapist. Two of the cases, termed the Coherent cases, were transcriptions from clinical training tapes of therapists interviewing actual patients with a mental disorder diagnosis. One Coherent case (referred to as the Coherent-ED case) depicted a patient who was diagnosed with the Eating Disorder Anorexia Nervosa. The other coherent case (the Coherent-SU case) described a patient who had been diagnosed with the Substance Use Disorder Alcohol Dependence. In short the Coherent cases presented one disorder within each Expert’s specific area of expertise and one case outside of that area. The third story (referred to as the Jumble case) described a patient for whom no coherent categorization could be made. This case had the same structure and length as the Coherent cases. Each symptom in the Jumble case was taken from a different DSM-IV disorder. This story served as an additional control that will not be discussed here.

A fourth story was created to test participants’ baseline free recall performance. This Baseline case used the same script format as the other three cases but portrayed an individual interviewing for a job. The Baseline case did not include any discussion of mental disorder information.

Procedure

Participants completed the following ten tasks in the described order. The focus of this paper is the results for the

1 Due to the small number of participants in each licensure group, the results are collapsed across this factor. It is an interesting question for future research to determine how specific licensure training affects recall for patients.
memory tasks, so only those tasks will be described thoroughly.

1) Reading of Patient Cases. Participants first read the three hypothetical patient cases (Coherent-ED, Coherent-SU, and Jumble cases). Participants were told they would be asked to provide a diagnosis and treatment plan for each case. This instruction was included to help guide the participants reading of the stories and to ensure that the patient descriptions were read thoroughly. The order in which the patient cases were read was counterbalanced across participants.

2) Case Diagnosis. After reading the cases, participants provided a diagnosis for each of the three cases.

3) Treatment Plan Generation. Participants were asked to describe any type of treatment that might aid each patient with their problems.

4) Inference Generation. Participants were asked to freely generate features of each case that were not mentioned in the case but they believed would be true of the described patients. An example was given to clarify: “For example, if a person had said that he/she is very anxious in social situations you could infer that he/she often blushes.”

5) Reading of Baseline Case.

6) Baseline Memory Recall. Participants completed a free recall task for the Baseline case where they listed as many features of the story as they could.

7) Free Recall Memory Test. Participants were asked to list as much information as they could about each of the three patient cases from the beginning of the experiment.

8) Recognition Memory Test. Participants were given a separate 25-item Old-New recognition task for each of the three patient cases. Items remembered as being in the original story were labeled by participants as Old. Items not explicitly stated by the patient in the original case were labeled as New. The 25 items for the task were composed of 10 Old items representing facts explicitly stated in the original cases; 10 New items that were associated with the correct diagnosis of the patient, but were not described by the patient in the case; and 5 participant-tailored Inference Lures that were composed of randomly selected items from each participant’s responses from the Inference Generation task for the matching case. The correct answer to these items was New.

9) Question Generation. Participants were asked to list questions they would have liked to ask the patient.

10) Background Questionnaire. Participants finished the experiment by filling out a background questionnaire that probed their experiences working in the mental health field and asked for personal demographic information. Participants rated their familiarity on a scale of 0 (not familiar at all) to 10 (extremely familiar) for the two disorder categories present in the Coherent cases (specifically Anorexia Nervosa and Alcohol Dependence).

For a given task, participants made responses for the three patient cases side-by-side on the same screen in the order that the cases were read. The name of each patient along with the order in which the corresponding story was read was provided as a cue above the appropriate response area (e.g., “Story 3 [the story of Chris]”). This presentation format was used to minimize the possibility of confusing responses across stories.

Participants completed the experiment either in a computerized, in-person experiment format (N=26) or as an online survey presentation (N=32). Participants completed the in-person format either by coming into the laboratory or by having the experiment brought to their office or home. Participants in the online format completed the experiment at a computer of their choice, be it in their office or home. Each task that required free listing of responses asked participants to separate their responses by providing distinct responses on separate lines of an Excel response sheet or an online survey, as appropriate, with unlimited space provided for each set of responses. The diagnosis and treatment plan tasks were presented on one screen of the experiment together. All subsequent tasks were presented on separate screens of the experiment.

Results

Coding

Two independent coders analyzed the free recall responses to determine what type of information each response represented. The coders first developed coding guides that outlined the individual pieces of information stated in the original cases. The coders used the coding guides to divvy participants’ responses into the following four categories.

Facts: responses that corresponded to statements made by the patients in the original cases. Items identified as Facts were additionally labeled with a code for which Fact from the original story they represented. A Fact coding was given both for items where only the gist of the fact was recalled, as was well as for items that were completely accurate representations of a fact mentioned in a case.

Extrapolations: responses that were plausibly true of a case but were not explicitly mentioned in the case. For example, in the Coherent-SU case the response “sees beer as different from liquor, not as damaging (part of denial)” was coded as an Extrapolation because the patient did not state any of this information about himself but it was plausibly inferred about the patient.

Intrusions: responses that were recalled under the wrong case heading or were blatantly incorrect.

Evaluations: responses that discussed the therapist or the therapeutic process. For example, comments about the style of the interviewer (e.g., "interviewer prompted too often") or about the type of interaction going on between the interviewer and interviewee were coded as Evaluations.

Across the four stories the coders agreed on the parsing of items into one of the four response categories on 88% of responses (overall Cohen’s Kappa = .605). The coders

---

2 Since the Jumble case did not represent a patient from a single diagnostic category, these items for the Jumble case were features that are common to many mental health disorders and were not diagnostic of a given category (e.g., “lost interest in hobbies”).
agreed on which specific fact unit a Fact represented 91% of the time overall (Kappa = .860). Disagreements in coding were settled through discussion between the coders.

Intrusions and Evaluations were very infrequent, together equaling .008% of the total responses. Since both the Intrusion and Evaluation categories do not represent actual facts recalled about the stories, they were excluded from further analyses. Therefore, the total number of responses a participant recalled was calculated by summing for a given participant the number of generated responses that were labeled as 1) Facts and 2) Extrapolations.

**Free Recall Performance**

The first question of interest is for the cases depicting the patients diagnosed with actual disorders (i.e., the Coherent patient cases), how does experts’ recall differ for patients inside versus outside their area of expertise? The general expertise literature found that greater experience promotes greater recall, thereby predicting that Experts should recall more for the patient in their area of expertise than the out of area patient. On the other hand, the medical expertise literature has found the reverse trend with Experts recalling less inside their area of expertise than outside their expertise area (consistent with an inverted U-shaped trend for expertise memory performance discussed earlier). To test these alternative predictions, Experts’ free recall performance for the Coherent case in their area of expertise (In-area case) and the Coherent case outside their area of expertise (Out-area case) were compared through a paired t-test. Mental health Experts recalled significantly more for the In-area (M=13.6) case than the Out-area case (M=12.1), t(16)=2.24, p < .04 (See Figure 1). This finding is a departure from the predictions of the medical expertise literature, and instead is in line with the general expertise literature.

In order to compare Experts’ performance to the other participants groups, total Coherent case free recall scores were calculated by averaging across performance for the In-area and Out-area cases. For Novices and Trainees, performance for the two Coherent cases was also averaged together to create one composite Coherent case recall score. The Coherent recall scores were then separated into their two components, Facts and Extrapolations. Each of these components is analyzed separately as follows.

The number of items coded as Facts was compared across groups by a one-way ANCOVA, with Expertise as a between subjects factor and baseline memory performance as a covariate. A significant main effect of Expertise was found, F(2, 54)= 4.44, p < .02 (see Figure 2). Specific independent t-tests compared the number of responses coded as Facts across the participant groups. Trainees recalled significantly more items coded as Facts (M=14.5) than Novices (M=10.4; t(39)=2.49, p < .02) consistent with the idea that expertise improves memory for category members. However, Trainees also recalled significantly more Facts than Experts (M=9.12), t(35)=4.13, p < .001. Novices and Experts did not differ, p > .4. These results are consistent with the findings of the medical expertise literature showing superior performance by Trainees, instead of supporting the general expertise trend of superior Expert performance as seen in the Expert In- versus Out-area comparison.

---

3 The differences in In-area versus Out-area recall performance could reflect better memory for a familiar than non-familiar category as opposed to something particular to expertise. Novice and Trainee’s familiarity ratings collected at the end of the experiment were used to identify a disorder of greater familiarity for these participants. Paired t-tests found no significant difference in the number of items recalled in the less and more familiar disorders, p’s > .3, ruling out this possibility.

4 A paired t-test found no difference in recall between the two Coherent cases, p > .17, justifying these composite scores.
The same comparisons were done for the number of recall items coded as Extrapolations. The main effect of Expertise was significant with baseline memory performance as a covariate, $F(2, 54)=3.60, p < .04$. Figure 3 shows the mean number of responses across the participant groups. Novices and Trainees did not differ in the number of responses coded as Extrapolations (Novices: $M=2.12$, Trainees: $M=2.08$; $p > .9$). Experts produced the most Extrapolations ($M=3.71$), recalling significantly more than Trainees, $t(35)=2.04, p < .05$. Experts recalled only marginally more Extrapolations than Novices ($p=.057$).

![Figure 3: Recall for Extrapolations](image)

What is being represented in these Extrapolations? One explanation comes from the medical expertise literature. Here, it has been proposed that experienced physicians do not report every individual feature of a patient they remember, but rather group those features into fewer higher order clusters and generate only those clusters when asked to freely recall. Applied to our experiment, Experts may take the individual patient features of “often sad”, “loss of pleasure in daily activities”, and “loss of appetite” and recall them as the single cluster of “has symptoms of depression”. This pattern of behavior could account for Experts recalling fewer Facts, but more items coded as Extrapolations (i.e., many individual Facts are combined and recalled as a few Extrapolations).

Do Experts remember the individual Facts used to create an Extrapolation? The free recall task alone cannot answer this question. However, a cued recognition task allows for the probing of individual Facts from the cases to test whether Experts remember the parts they may have used to build Extrapolations. Participants completed a cued recognition task for each patient case to test their memory for individual Facts.

**Recognition Performance**

Participants completed a separate Old-New Recognition task for each of the three patient cases. Performance was measured by calculating Corrected Recognition (CR) scores for each participant in each story. CR is a ratio of the proportion of responses correctly identified as old to the proportion of responses that were falsely identified as old (i.e., the ratio of hits to false alarms) and ranges between a lowest score of 0 and a highest score of 1. The averages of the CR scores for the Coherent cases were submitted to a one-way ANOVA with Expertise as a between-subjects variable. A significant main effect of Expertise was found, $F(2, 55)=4.42, p < .02$. To explore this finding, the different participant groups were compared via independent t-tests. Trainees showed better CR performance than Novices (Trainees: $M=.740$; Novices: $M=.618$, $t(39)=3.26$, $p < .005$). Trainees CR performance was marginally different from Experts ($M=.652$), $p=.057$. Novices did not differ from Experts, $p > .4$. CR performance seemed to demonstrate another form of the intermediate effect found in the medical expertise literature.

The differences between participant groups were further explored by comparing performance across groups for the disorder participants were more familiar with (In-area) and separately for the disorder participants were less familiar with (Out-area; see footnote 3). A one-way ANOVA was conducted for the In-area CR scores with Expertise as a factor. The main effect was not significant, $p > .3$. The same ANOVA conducted over Out-area scores found a significant main effect of Expertise, $F(2, 49)=4.39, p < .02$. Novices ($M=.565$) and Experts ($M=.619$) did not differ on their Out-area CR performance, $p > .3$. Trainees ($M=.737$) showed significantly better Out-area CR performance than Novices, $t(33)=3.21, p < .005$ and marginally different performance than Experts, $p = .068$. These findings indicate that Experts did not differ in recognition memory from other groups for their In-area disorder, but tended to differ in the ability to correctly recognize facts for the Out-area disorder.

**General Discussion**

The results of this study provide an interesting demonstration of expertise effects within the mental health profession. These expertise effects are complex in nature. The intermediate effect so broadly shown in the medical expertise literature was seen here in the overall pattern of free recall responding for Facts and for cued recognition performance. However, when experts’ memory was compared inside versus outside their specific area of expertise, experts were found to freely recall significantly more about a patient diagnosed in their area of expertise than a patient diagnosed outside their area of expertise. This finding is strictly in opposition to similar comparisons made within the medical expertise literature (e.g., Rikers, Schmidt, & Boshuizen, 2002).

What structure do mental disorder categories take in the clinician’s mind to result in the pattern of results seen in this experiment? As discussed previously, the general expertise

---

5 Specific t-tests found that none of the three groups differed from one another on CR for their In-area disorder, all $p$’s > .15.
literature posits that experts have more elaborate knowledge and conceptualizations for a category inside than outside their area of expertise. This elaboration would explain what the differences between Trainees and Experts. If Experts do gain sophistication and detail in conceptualizations of their expertise area, then why would these conceptualizations not be more elaborate than Trainees’ and provide for better memory for Facts for Experts? Further research is required to answer these questions and better delineate how clinicians think of mental disorder categories.

The results presented here have interesting implications for the clinical setting. Imagine a patient that possesses features falling in two mental disorder categories, such as Anorexia and Alcohol Dependence. It is possible that an expert in Eating Disorders may remember more information related to the diagnosis of Anorexia and forget the details pertinent to the other disorder. However, a Substance Use Expert may remember the opposite information about the same patient; that is, she may remember information relating to the diagnosis of Alcohol Dependence and forget Anorexia-related symptoms. Such a bias in memory for patient information could have profound effects for clinical care. A startlingly large percentage of patients (~28%) meet criteria for two or more disorder categories (Kessler et al., 2005). Such statistics underscore the importance of understanding how expertise for one disorder may affect the diagnostic and therapeutic process in mental health care.

Acknowledgments
Support for this research was provided in part by Yale University Psychology Department graduate research funds awarded to Jessecah K. Marsh.

References