

## Memory for Patient Information as a Function of Experience in Mental Health

JESSECAE K. MARSH<sup>1</sup>\* and WOO-KYOUNG AHN<sup>2</sup>

<sup>1</sup>Department of Psychology, Lehigh University, Bethlehem, USA

<sup>2</sup>Department of Psychology, Yale University, New Haven, USA

*Summary:* Mental health clinicians are tasked to diagnose and treat the millions of people worldwide seeking help for mental health issues. This paper investigates the memory clinicians have for patient information. We hypothesize that clinicians encapsulate mental health knowledge through experience into more abstract concepts, as in other domains changing what clinicians remember about patients compared with non-professionals. We tested memory for realistic patient–therapist interactions in experienced clinicians, intermediately trained graduate students, and laypeople. Clinicians recalled fewer facts than intermediate trainees and as many as laypeople. Furthermore, clinicians reported more abstracted information than all other participants, providing the first empirical demonstration of knowledge encapsulation in the memory of mental health clinicians. We discuss how our results fit into the existing literature on clinical expertise in mental health and the implications of our findings for future research relevant to mental health care. Copyright © 2012 John Wiley & Sons, Ltd.

Mental disorders affect a large portion of the world population with an estimated at least 450 million people worldwide currently experiencing a mental illness. These disorders lead to great burden in the lives of those affected, with mental disorders being four of the six leading reasons people live in disability (World Health Organization, 2001). Mental health clinicians play a significant role in addressing these problems. Although the specific tasks may vary by clinician and his or her theoretical orientation or training, the responsibilities of clinicians may include conducting assessments, diagnosing mental disorders, creating case conceptualizations, estimating prognosis, developing treatment goals, and providing overall care for their patients (Hunsley & Lee, 2010). Understanding the cognitive processes underlying these tasks may help improve the mental health care process (Garb, 2005). The current study explores one such cognitive aspect, namely memory for patient information, which plays an important role in creating case conceptualizations and therapy plans (Hill, 2005). In particular, we compared experienced professional mental health clinicians, intermediately trained clinical graduate students, and novice laypeople in terms of their memory for patient information.

How might mental health clinicians remember information about their patients differently than people with less experience in the mental health field? We propose three alternatives.

First, mental health clinicians may better remember information about patients compared with people with less experience. This outcome is likely because experts in general are found to have increased memory for coherent, domain-relevant information (Chase & Simon, 1973; de Groot, 1966; Ericsson & Charness, 1994; Glaser & Chi, 1988; Hatano & Osawa, 1983).

It is yet to be shown, however, whether mental health clinicians also have better memory for patient information. The closest demonstration is by Brailey, Vasterling, and Franks (2001), who found that clinicians recalled more information specifically related to Diagnostic and Statistical

Manual of Mental Disorders (DSM) diagnoses than graduate students. Their participants were explicitly cued to recall patient information relevant to DSM diagnostic criteria. Thus, their results could have been obtained because clinicians know the DSM diagnoses better than graduate students rather than because they have better memory for various aspects of patient information, including those that are not directly relevant to formal diagnostic criteria, but are still essential to diagnosis and prognosis (e.g. risk factors; Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001). Whether mental health clinicians spontaneously recall more patient information in general has not yet been demonstrated.

Second, the memory of mental health clinicians may not differ from the lesser-experienced, given previous studies showing that in a variety of clinical tasks, clinicians do not necessarily outperform lesser-experienced participants (for reviews, see Dawes, 1994, and Garb, 1998; see Garb, 2005, for a review of these issues as it relates to interrater reliability). For example, patients seen by either professional therapists or laypeople given basic instruction in therapy demonstrated equal post-therapy improvement (e.g. Beutler, Machado, & Neufeldt, 1994; Durlak, 1979; Strupp & Hadley, 1979). Accuracy of professionals in making different clinical judgments (e.g. diagnosis) did not appear consistently better than inexperienced participants (e.g. Ebling & Levenson, 2003; see Garb, 1989, for a review and for critiques Berman & Norton, 1985; Durlak, 1979; Faust & Zlotnick, 1995; Strasser & Gruber, 2004). Although some research suggests training improves diagnosis (e.g. Brammer, 2002; Lambert & Wertheimer, 1988; see Spengler, White, Ægisdóttir, & Maugherman, 2009, for a review), others found no such effect (e.g. Ganzach, 1997; Hillerbrand & Claiborn, 1990). In short, the lack of consistent and reliable expertise effects in mental health suggests that memory differences may not be found either.

We believe that there is a third alternative that is the most likely description for how experience in mental health influences the memory of professional clinicians. The general expertise literature shows that as experience is gained in a domain, relevant information is reorganized at a more abstract level and in larger, more information-rich units (see Feltoch, Prietula, & Ericsson, 2006, for a review). We propose that a

\*Correspondence to: Jesseciae K. Marsh, Department of Psychology, Lehigh University, 17 Memorial Drive East, Bethlehem, PA 18015, USA.  
E-mail: jessecae.marsh@lehigh.edu.

similar restructuring of domain-relevant knowledge into abstract, higher order concepts will occur in professional clinicians. As a result, when clinicians recall information about patients, they will differ in the type of information they recall. Specifically, they will recall more abstract information than people with less experience in mental health.

A similar process has been proposed in an analogous domain to mental health, namely medicine. Physicians are described as restructuring their knowledge into a series of higher-order concepts, termed encapsulated knowledge, as they gain experience within a field (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1992; van de Wiel, Boshuizen, Schmidt, & Schaper, 2000). The restructuring involves appropriate sets of symptoms or syndromes being clustered into a higher-order concept through the causal relationships that underlie and connect them (Schmidt & Rikers, 2007). For example, the individual symptoms of, 'high fever, shaking chill, sweating, feelings of prostration, shortness of breath, and a high pulse rate', reflective of infection could be encapsulated as the single medical condition of 'sepsis' that implies a bodywide response to infection (example from Rikers, Loyens, & Schmidt, 2004, p. 1038).

Within medicine, knowledge encapsulation has been credited with producing the apparently paradoxical memory finding of expert physicians recalling less information about patients than advanced medical students, and in fact recalling similar amounts to untrained novices (Schmidt & Boshuizen, 1993; see also, Claessen & Boshuizen, 1985; Eva, Norman, Neville, Wood, & Brooks, 2002; Patel, Groen, & Frederiksen, 1986; but see van de Wiel, Schmidt, & Boshuizen, 1998). This inverted U-shaped pattern in the amount of recalled information has been termed the 'intermediate effect' because it is people with intermediate experience that show the best memory. The intermediate effect is explained by knowledge encapsulation in that physicians recall a few higher-level facts about a patient (e.g. 'sepsis') as opposed to the individual facts that compose this concept. In contrast, people with intermediate amounts of medical training (e.g. medical students) have basic biomedical knowledge that aids memory but have yet to build the knowledge structures that would support this re-conceptualization of knowledge, and therefore recall many individual facts about a patient. People who are even more inexperienced in medicine recall few pieces of information because of their lack of knowledge in the domain (Patel & Groen, 1986; Schmidt & Boshuizen, 1993; van de Wiel, Boshuizen, & Schmidt, 2000). Physicians recall even less about patients in their own subspecialty area than outside their specialty area because of higher knowledge encapsulation in their own specialization (e.g. Rikers, Schmidt, & Boshuizen, 2002; Rikers, te Winkel, Loyens, & Schmidt, 2003). For example, a cardiologist would recall less about a patient with a heart defect than a neurologist would (example adapted from Rikers, Schmidt, & Boshuizen, 2002).

The knowledge encapsulation found in medical health professionals, however, may not generalize to mental health professionals. As explicitly noted in the contemporary versions of the DSM (American Psychiatric Association [DSM-IV], 1994; American Psychiatric Association [DSM-IV-TR], 2000), most recognized mental disorders do not

have well-understood and agreed-upon etiologies. Furthermore, there is a debate as to what counts as a separate mental disorder (e.g. whether generalized anxiety disorder exists separately from major depressive disorder). These ambiguities may dissuade the production of encapsulated units.

Nonetheless, a few studies in mental health have shown intermediates performing differently from other groups, suggesting the plausibility of this account. Eells, Lombart, Kendjelic, Turner, and Lucas (2005), for instance, found that highly experienced clinicians constructed better case formulations than novices, but lesser experienced clinicians constructed the worst case formulations. The same trend was found for diagnosis (Witteman & van den Bercken, 2007). Unlike the aforementioned demonstrations of intermediate effects in memory in the medical domain, these studies found worse performance in intermediates. Yet, these results could be consistent with the medical domain because the better case formulations and diagnoses of professionals could also be resulting from their encapsulated knowledge. Thus, if their memory for patient information is measured, experienced mental health professionals may also show lower recall of patient information just as the medical experts. (See Hauser, Spada, & Rummel, 2007; Hauser, Spada, Rummel, & Meier, 2006 for other preliminary results.)

These varying possibilities about the memory of clinicians necessitate a systematic empirical test. We predict that experienced mental health professionals should encapsulate knowledge in ways similar to other domains, resulting in clinicians recalling a greater number of abstract, encapsulated units than other participants. If this encapsulation occurs, then an intermediate effect would obtain such that clinicians would recall fewer basic level facts than intermediates and may look similar in recall to laypeople.

## METHOD

### Participants

We recruited participants with three levels of experience with mental health disorders. Table 1 shows summary demographic information for each group.

#### *Professional experience group: Clinicians*

The professional experience group consisted of clinical social workers ( $n=8$ ), psychiatrists ( $n=2$ ), and psychologists ( $n=11$ ), who had been licensed for at least 10 years, following the commonly used proxy for expertise (e.g. Chi, Glaser, & Farr, 1988). In order to maximize familiarity with diagnoses used in our experimental materials, namely alcohol abuse and anorexia, we only recruited clinicians who specialized in treating either eating ( $n=11$ ) or substance use disorders ( $n=10$ ). This way, we could also compare the memory of clinicians for cases within their specialization area and outside of their specialization area. Clinicians indeed reported being more familiar and seeing more patients diagnosed with the disorders in their area of specialization than the other disorders (Table 2).

Participants in this group were recruited through advertising flyers sent to clinics specializing in either eating disorders or substance use, email advertisements sent to

Table 1. Demographic information for each participant group

	Clinicians	Trainees	Undergraduates	Age-matched laypeople
Demographics				
<i>N</i>	21	20	21	21
Sex (female/male)	14/7	14/6	10/11	14/7
Mean age in years ( <i>SD</i> )	51.7 (5.8)	26.7 (1.4)	18.9 (0.9)	51.7 (5.5)
Total age range	41–67	24–29	18–21	41–63
Highest degree obtained				
Bachelor's	0	5	–	0
Master's	7	15	–	14
PhD / PsyD	9 (PhD), 3 (PsyD)	0	–	3 (PhD)
MD	2	0	–	0
Other	0	0	–	2 (EdD), 1 (MBA), 1 (JD)
Experience in mental health issues				
Mean years seeing patients ( <i>SD</i> )	26.7 (5.7)	2.33 (0.9)	–	–
Range in years of seeing patients	14–38	1–4.5	–	–
Mean number of psychology courses taken ( <i>SD</i> )	–	–	1.43 (0.8)	1.67 (0.7)
Theoretical orientation				
Cognitive–behavioral	7	12	–	–
Eclectic	6	3	–	–
Psychoanalytic	6	3	–	–
Behavioral	3	3	–	–
Other	3	3	–	–

Note. – indicates items that were not asked on the given group.

listings of professionals who self-identify as specializing in the treatment of these disorders, and emails sent to mailing lists of national professional associations for clinicians who specialize in treating these disorders. Clinicians were paid at a rate of \$85 per hour for their participation.

#### Intermediate experience group: Trainees

This group consisted of clinical psychology graduate students.<sup>1</sup> Students were recruited who were at least in their third year of graduate training ( $M = 3.65$ ;  $SD = 0.75$ ; range = 3–5 years) and had been responsible for the individual treatment and diagnosis of patients in a clinical setting to ensure basic exposure to patient interactions. Trainees reported seeing patients but significantly fewer than clinicians (Table 2). Participants in this group were recruited through clinical psychology programs of six different universities in the USA. Trainees were compensated with a flat fee of \$30.

#### No-experience groups

The no-experience groups consisted of age-matched laypeople and undergraduates.

**Age-matched laypeople.** Because people tend to recall less information as they age (Salthouse, 2009) and what is recalled tends to be more gist-like in nature (e.g. Tun, Wingfield, Rosen, & Blanchard, 1998), we tested a group of laypeople matched in age with the clinicians group to ensure that gist-like memory, if found in the clinicians, would not be solely attributable to age differences. Laypeople were recruited through a campuswide advertisement

seeking participants in the age range of 41–67 years with at least a Master's level degree, the same age range and educational background as the clinician group. Those who had undergraduate majors in psychology or had worked in a mental health care setting were excluded. Participants were paid at the same rate as the clinicians (\$85 per hour).

**Undergraduates.** Because age-matched laypeople, although lacking professional experience, may have interacted with more people who fit diagnostic categories, which can in turn change their views of mental disorders (see Hooten, 2011), we additionally tested undergraduates as another no-experience group. Yale University undergraduates screened to have no previous experience working in mental health care participated as part of an introductory psychology course experiment requirement or for pay at a rate of \$10 per hour. They had completed little formal course work in Psychology, had not taken Abnormal Psychology, and most ( $n = 12$ ) had not taken a psychology course beyond Introduction to Psychology.

#### Materials

We developed two written descriptions of patients, termed the Coherent cases, to serve as the main experimental materials testing memory of participants for mental health information. These cases were abridged transcriptions from clinical training tapes and portrayed actual patients being interviewed by a therapist. One Coherent case described a patient diagnosed with the eating disorder anorexia nervosa, bingeing–purging subtype (referred to as the eating disorder case), and the other described a patient diagnosed with the substance abuse disorder of alcohol dependence (referred to as the substance abuse case). The narratives took the form of the therapist asking the patient about his or her major mental health concerns (e.g. 'I understand you have some

<sup>1</sup> An additional trainee participated in the experiment but was excluded from discussed analyses as an outlier. The recall performance of the participant on the Coherent stories fell more than three standard deviations outside the performance of the rest of the trainee group. However, including the data would not change the pattern of presented results.

Table 2. Mean number of patients treated and familiarity ratings for disorder categories across experience

	Clinicians in area of specialization	Clinicians out of area of specialization	Trainee
No. of patients seen			
Anorexia nervosa	238* (318)	19.8* (31)	0.50 (1.1)
Alcohol abuse	386* (433)	55.1* (63)	1.95 (3.0)
Eating disorders	477* (509)	35.2* (50)	1.90 (3.3)
Substance abuse disorders	511* (440)	72.5* (75)	5.25 (8.5)
Average specific	312.1** (378)	37.5 (51)	1.23 (3.6)
Average general	493.6** (463)	53.9 (65)	3.58 (4.6)
Familiarity: scale 0 (not familiar at all) to 10 (extremely familiar)			
Anorexia nervosa	9.55* (0.69)	3.4 (2.4)	5.28 (2.4)
Alcohol abuse	9.90* (0.32)	5.91 (1.2)	5.30 (2.4)
Eating disorders	9.73* (0.47)	4.10 (2.3)	5.35 (2.3)
Substance abuse disorders	9.60* (0.97)	5.82 (1.3)	5.10 (2.5)
Average specific	9.71** (0.56)	4.71 (2.2)	5.29 (1.8)
Average general	9.67** (0.73)	5.00 (2.0)	5.23 (1.7)

Note: Numbers in parentheses represent standard deviations. Patient load of clinicians may be underestimated. Many clinicians reported the number of patients in their area of expertise as 'hundreds' or 'thousands'. Conservative versions of these types of estimates were used such that an estimate of 'hundreds' was recorded as 100 and estimates of 'thousands' were recorded as 1000.

\*Significant difference between the clinician group and trainee group ( $ps < .01$ ).

\*\*Significant difference between in-area and out-of-area clinicians ( $ps < .005$ ).

difficulties eating and with your eating habits.'), and the patient describing the issues she or he was dealing with (e.g. 'I don't want to eat because it seems like as soon as I eat, I just gain weight.').

We created an additional patient case, termed the Jumble case, to serve as an exploratory material testing whether experience influenced the memory of participants for non-coherent, mental health-related information. This fictional patient description followed the same interview format and structure and was the same length as the Coherent cases. Symptoms endorsed by the hypothetical patient were taken from several different DSM disorders (American Psychiatric Association [DSM-IV-TR], 2000), with not enough information being endorsed to meet diagnostic criteria for any single disorder. Because of the way the Jumble case was created, clinicians may lack a priori theories and encapsulated knowledge for this case. However, many patients with real-life mental disorders tend to meet criteria for multiple disorders simultaneously, meaning clinicians may rarely meet a textbook case of a single mental disorder, like the Coherent cases used in the current study. Thus, experienced clinicians may be adept at processing even incoherent mental health information into meaningful, encapsulated chunks. For that reason, the Jumble case is not an ideal control condition, like the 'random games' used to study chess experts (Chase & Simon, 1973); even the Jumble cases can be made meaningful by clinicians. Consequently, the effects of experience we have found may be attenuated, rather than missing, for this case.

As such, we created a fourth case depicting a job interview to serve as a measure of baseline recall for person information when no mental health component is involved. We chose the dialog between an individual interviewing for a job and her potential employer in order to keep the overall written presentation and format as similar as possible to the patient cases. This way, we can simultaneously partial out in our analyses any potential disadvantages (e.g. aging effect) as well as advantages (e.g. potential transfer of more experience in interviews) that clinicians may have over

others and measure only the effect of mental health materials. This Baseline case had the same overall length and structure as the three previously described cases. All stories are available from the authors (see Marsh, 2008, for the Jumble and Baseline cases).

## Procedure

All participants first read patient cases and performed some tasks in relation to the patient cases. After a filler task, they were asked to free recall the patient cases and performed cued recognition for the patient cases. These main memory tasks will be described first, followed by a description of the filler and control tasks.<sup>2</sup>

### Reading and processing the patient cases

Participants began the experiment by reading the three patient cases (e.g. eating disorder, substance abuse, and Jumble cases) in a counterbalanced order. To encourage thorough reading of the cases, participants were told to provide a 'judgment or diagnosis of what may be happening with this person' and a treatment plan for each case or 'what steps you believe should be taken for this person to better his/her current situation'.<sup>3</sup> There was no effect of story order for any of the memory tasks,  $ps > .12$ , so all results were collapsed across counterbalancing order.

<sup>2</sup> It should be noted that although some of the predictions discussed here are inspired by research on the development of memory over experience in medical domains, the primary aim of this study is not a comparison of mental health to said domains. Thus, the current methods were not directly taken from existing studies but rather developed to be suitable for cases involving mental health disorders.

<sup>3</sup> These instructions could focus participants on reading the cases for this information alone, which could influence the type of information recalled in our tasks. Despite this possibility, we think it is important to give some direction so that all participants read the cases with the same goal in mind. Without this, clinicians may read the cases naturally with this goal, whereas novices may read the cases with a different goal. Also, because all participants received this instruction, any differences between groups should not be reflective of this manipulation.

Then, participants were asked to list features of each case that 'were not explicitly mentioned in the narratives, but that you believe would be true of the described people'. This feature generation task was included to develop lures for a later recognition task. In addition, this task would have encouraged participants in all groups, not just experts where it may come naturally, to think beyond the facts of the case, providing a stronger test of relative differences between trainees and experts.

#### *Surprise free recall*

To measure the quantity and type of information people would spontaneously recall about mental health patients, participants received a surprise free recall task. They were instructed to list from memory as many features as they could about each patient case. For this and all other tasks, participants were cued to respond for a specific case by the name of each patient along with the order in which the case was read [e.g. 'Story 1 (the story of Chris)'] in order to minimize the possibility of confusing responses across cases.

#### *Recognition task*

Participants could differ in free recall because of differences in responding (e.g. feeling the need to only list important information for diagnosis) rather than differences in underlying memory. To address this problem, participants completed separate cued recognition memory tests for each of the three patient cases. Participants read a list of 25 statements for each patient and were asked to identify whether the item was explicitly stated in the original case (i.e. an 'old' item) or if the item had not been explicitly stated (i.e. a 'new' item). For each case, we included 10 old and up to 15 new items. To increase the difficulty and variety of the new items, statements were created that were related to a case fact but were incorrect on a detail of the fact ('avoids weighing herself' when the patient stated she frequently weighs herself;  $n=5$ ), as well as statements that were associated with the correct diagnosis (Coherent cases) or with disorders more generally (Jumble case) but not included in the cases ('hides drinks around his house' for the substance abuse case;  $n=5$ ). In addition, to increase the difficulty of the task, we used five custom items that were generated by each participant in a previous inference task. The inferences used for a given participant were randomly selected from the list of responses from that participant and reformatted to match the other recognition items (e.g. removing capitalization, correcting spelling errors, etc.). If a participant had originally generated five or fewer inferences, then all of his or her inferences were used. We did not find differences across experience groups for the different types of new items so we have collapsed across the new items and do not present separate analyses by item type.

To analyze the recognition data, for each participant, we computed a corrected recognition (CR) score, the 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 correct hits to false alarms). Perfect performance is indicated by a score of 1, with the lowest possible performance score equal to 0. Scores for the

recognition task were calculated out of the total number of items each individual participant received.

#### *Filler and control tasks*

Participants read the Baseline memory case after reading the three patient cases and before the free recall task. After reading the Baseline case, participants immediately completed a free recall task for information provided in the story. This task served to measure the baseline memory of participants.

Finally, participants answered a series of post-test questions. To see if any of the experience groups had felt under-informed, participants were asked to generate any questions they would have liked to ask the patient that 'would have helped you understand what was going on with each person'. A one-way analysis of variance (ANOVA) across the experience groups found no difference on the number of questions generated for the Coherent cases ( $p=.18$ ,  $\eta_p^2=0.058$ ). Then, participants answered questions about their age, gender, and education, as well as their familiarity for the four disorder categories on a scale of 0 (not familiar at all) to 10 (extremely familiar). Undergraduates and age-matched laypeople reported how many psychology courses they had taken and if they had experience working or volunteering in any type of mental health setting (e.g. volunteering in a camp for children with emotional problems) to double-check our screening criteria. Trainees reported in what capacities they had seen patients (e.g. individual sessions with in-session supervision) and if they had diagnosed patients in a clinical setting. Both clinicians and trainees reported the extent to which they had treated patients generally and specifically how many patients they had seen for the four disorder categories of interest, as well as indicated their primary theoretical orientation.<sup>4</sup>

#### *Other procedural details*

The experiment was conducted in two different formats. For participants local to the Connecticut area, an in-person format using the RSVP computer software to present all instructions and cases on one computer screen and an Excel spreadsheet to record all responses on an adjacent computer was used ( $n=26$ ). In order to recruit additional participants from a larger geographic region, online formats were created using SurveyMonkey and Qualtrics ( $n=57$ ). The online version looked as identical as possible to the in-person version. The only notable difference was that the online survey software allowed the participants to respond in text boxes in the survey software instead of on a separate computer. The survey programs did not allow participants to go back to previous screens of the experiment or open another window containing the experiment. The experimenter scheduled appointment times with the online subjects so she could be available to tailor the recognition task responses. There were no differences in participant responses using the Web or in-person formats for any of the memory tasks, all  $ps > .14$ , so all data are collapsed across formats.

<sup>4</sup> Although we collected theoretical orientation information and the current results could depend on different theoretical orientations, there were not enough clinicians sampled to do analyses across orientation in our data sample.

All participants completed the experiment at their own pace (mean completion time = 86 minutes;  $SD = 27$ ). There was no significant difference among the four experience groups in the time taken to complete the entire experiment,  $p = .15$ ,  $\eta_p^2 = 0.07$ , or in the elapsed time from reading the patient cases to the free recall task (e.g. encoding time),  $p = .69$ ,  $\eta_p^2 = 0.019$ . Encoding time was not a significant covariate of any memory measures.

## CODING SCHEME CREATION AND USE

A goal of this study was to measure the amount and type of information remembered by different experience groups. As such, two coders (one blind to the purposes of this study) developed a master coding guide that divided each story into the series of facts that were presented (similar to the analyses used in memory for texts, Bransford & Johnson, 1972). Figure 1 shows an example of how a section of the Jumble case was divided into fact units according to the master coding guide. Free recall responses were compared with these coding guides to determine what piece of information they represented from the cases. Descriptions of the creation of the coding scheme and the coding process can be found in the Appendix.

Each recall item was coded as to whether it represented a 'fact' or a 'conjectured feature'. Items coded as facts represented information presented in the original case. Facts were further classified as either specific (i.e. correct recall of specific details of the fact) or non-specific recalls (i.e. missing specifics) of the original fact (see Table 3 for examples). Recall items that were not explicitly stated in the narratives and represented abstracted information about the patient were coded as 'conjectured features'. For example, 'sees beer as different from liquor—not as damaging [part of denial]' was coded as a conjectured feature because the patient did not state any of this information about himself. As can be seen in Table 3, the conjectured features are

INTERVIEWER: You've said that you have been having some sleep problems. Tell me a little about that. CHRIS: Yeah, it's just that I don't really sleep at night anymore. It seems that when I try to go to bed, you know, at night, I can't.	1
INTERVIEWER: Do you ever get a chance to sleep? CHRIS: Mmm, yeah. INTERVIEWER: When would that be? CHRIS: It seems I sleep a lot during the day.	2
INTERVIEWER: Is every day like this for you?	3
CHRIS: No, not every day. Sometimes I seem to have just tons of energy. I can work all day on things and not get tired.	4
I end up only sleeping for a couple of hours because I have so much energy.	5

Figure 1. Example coding parsing. Numbered boxes indicate text that was combined as a single text unit. The text units labeled 1 and 2 show examples of several sentences being combined into one story fact. Text units 3, 4, and 5 are examples of one passage of text that was separated into multiple parts

not just instances of recalling bits of a fact without recalling the details (such items were coded as non-specific facts) but rather represented information that went beyond the given text. Most of the conjectured features are stating underlying causal mechanism (e.g. 'using compensatory behaviors to control her weight'), providing diagnostic interpretations of surface behaviors (e.g. 'alcoholic'), or providing a higher-order feature that subsumes individually stated features (e.g. 'socially isolated'). Thus, they fit the descriptions of Schmidt and Rikers (2007) of encapsulated knowledge seen in physicians (i.e. higher-order concepts composed of symptoms and causal mechanisms interrelating the symptoms).

Across the three patient cases, most responses were coded as facts or conjectures (81.5% and 17.5%, respectively). The rest of the responses (1.1%) were either intrusions (e.g. recalling a fact from the alcohol abuse case as in the eating disorder case) or comments about the style of the interviewer (e.g. 'interviewer prompted too often'). These responses were not included in the following analyses because of their low frequency.

## RESULTS

We first present performance on Coherent cases for facts and conjectured features. Then we present performance on the Jumble case, followed by performance in and out of specialty area. To account for any differences in performance by age, we used Baseline free recall memory performance as a covariate in all analyses comparing free recall across experience groups.<sup>5</sup>

### Recall of factual information on coherent cases

We hypothesized that if clinicians restructure information about patients via knowledge encapsulation, they should recall less information about patients than the intermediate experience group. We obtained this basic finding, as shown in Figure 2. We averaged across the number of facts recalled in the two Coherent cases to create an average Coherent case fact recall score. Experienced clinicians recalled fewer facts than intermediate trainees and actually showed recall performance that was similar to inexperienced undergraduates and age-matched laypeople. A one-way analysis of covariance (ANCOVA) on the average Coherent fact scores with Experience (undergraduate, age-matched layperson, trainee, clinician) as a factor and Baseline memory as a covariate found a significant main effect of Experience,  $F(3,78) = 4.93$ ,  $p = .039$ ,  $\eta_p^2 = 0.10$ . The covariate of Baseline memory was also significant,  $F(1,78) = 33.1$ ,  $p < .001$ ,  $\eta_p^2 = 0.29$ . Specific planned independent  $t$ -tests found clinicians recalled significantly fewer facts than trainees [ $t(39) = 4.36$ ,  $p < .001$ ,  $d = 1.36$ ], whereas undergraduates and clinicians did not differ,  $p = .35$ ,  $d = 0.29$ . Age-matched

<sup>5</sup> A one-way ANOVA with Experience (undergraduate, age-matched layperson, trainee, and clinician) as a between-subject variable did not find a significant main effect of Experience for baseline memory,  $p = .16$ ,  $\eta_p^2 = 0.062$ . Further specific  $t$ -tests found no significant differences between clinicians and undergraduates ( $p = .90$ ,  $d = 0.04$ ), trainees ( $p = .076$ ,  $d = 0.57$ ), or age-matched laypeople ( $p = .49$ ,  $d = 0.21$ ).

Table 3. Examples of coded recall items

Patient case	Fact from master coding guide	Participant responses		
		Fact-specific recall	Fact-non-specific recall	Conjectured feature
Substance abuse case	Drinks two and a half cases of beer a week	'Drinks 2.5 cases of beer per week'	'Drinks several cases of beer'	'Compulsive drinker/alcoholic'
Eating disorder case	Laxatives to purge when at work	'Uses laxatives to purge if she's at work'	'She reported use of laxatives'	'Using compensatory behaviors to control her weight'
Jumble	Lost friends because of irritability	'Has lost friends due to his irritability'	'Has lost some friends because of the way he acts'	'Socially isolated'

Note: Items in quotes are actual participant responses.

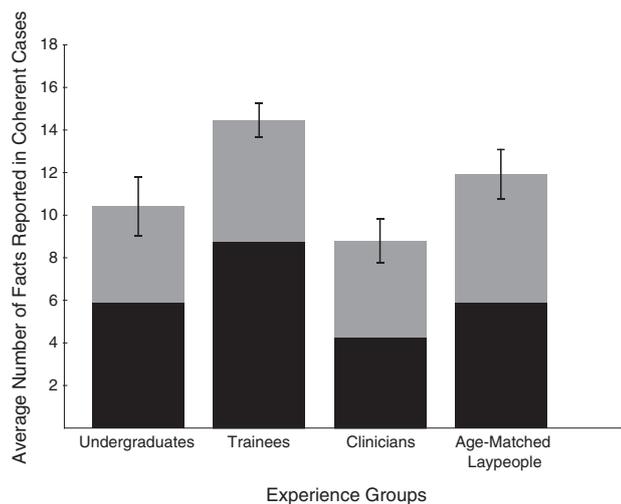


Figure 2. Recall performance for facts in the Coherent cases. Note: The black section of each bar represents the number of specific facts recalled, and the gray section represents the number of non-specific facts recalled. Error bars are standard errors of the total number of recalled facts (specific and non-specific fact recall combined)

laypeople did not differ from novices in the number of recalled facts ( $p = .41$ ,  $d = 0.26$ ) but reported almost significantly more facts than clinicians,  $t(40) = 2.01$ ,  $p = .051$ ,  $d = 0.62$  (see Table 4).

We further analyzed specificity of the facts that participants recalled by looking at the total number of facts that were recalled with the original details intact (i.e. specific

facts, represented as the black sections of bars in Figure 2). The same basic intermediate effect was found for specific fact recall. Trainees recalled the greatest number of specific facts, followed by undergraduates, and then age-matched laypeople. Clinicians recalled the fewest specific facts. A one-way ANCOVA over the mean number of Coherent case specific fact recalls with Experience as a between-subjects factor and Baseline memory as a covariate found a significant main effect of Experience,  $F(3,78) = 4.09$ ,  $p = .009$ ,  $\eta_p^2 = 0.13$ . The covariate of Baseline memory was also significant,  $F(1,78) = 12.7$ ,  $p = .001$ ,  $\eta_p^2 = 0.14$ . Planned independent  $t$ -tests found that trainees recalled significantly more specific facts than clinicians [ $t(39) = 4.23$ ,  $p = .001$ ,  $d = 1.32$ ], as well as significantly more than age-matched laypeople [ $t(39) = 2.48$ ,  $p = .017$ ,  $d = 0.78$ ] and undergraduates [ $t(39) = 2.12$ ,  $p = .040$ ,  $d = 0.66$ ], replicating the intermediate memory effect in the medical domain. Clinicians recalled marginally fewer specific facts than undergraduates [ $t(40) = 1.77$ ,  $p = .084$ ,  $d = 0.55$ ] and did not differ from age-matched laypeople,  $t(40) = 1.60$ ,  $p = .12$ ,  $d = 0.50$ .

Could clinicians still remember specific details concerning their patients despite not spontaneously reporting them? To examine this possibility, we analyzed the recognition task. A one-way ANOVA on the average Coherent CR scores found a significant main effect of Experience,  $F(3,79) = 3.82$ ,  $p = .013$ ,  $\eta_p^2 = 0.12$ . Planned independent  $t$ -tests found that CR scores for clinicians were significantly lower than scores for trainees [ $t(39) = 2.03$ ,  $p = .049$ ,

Table 4. Mean memory performance for free recall and cued recognition performance across experience groups

	Clinicians	Trainees	Undergraduates	Age-matched laypeople
Baseline memory case				
Overall recall	15.6 (6.9)	19.1 (6.6)	15.1 (5.4)	16.8 (6.7)
Average across coherent cases				
Facts	8.81 (4.7)	14.5** (3.6)	10.4 (6.4)	11.9 (5.4)
Specific facts	4.24 (3.0)	8.73** (3.7)	6.17 (4.0)	5.88 (3.6)
Conjectured features	3.67 (3.0)	2.08* (1.2)	2.12* (1.4)	1.69** (1.3)
Corrected recognition scores	0.65 (0.18)	0.74* (0.11)	0.62 (0.13)	0.73 (0.14)
Jumble case				
Facts	9.43 (5.6)	13.20 (4.4)	9.52 (5.0)	9.86 (4.8)
Specific facts	4.14 (4.0)	7.40 (3.7)	5.24 (3.4)	4.71 (3.1)
Conjectured features	3.62 (3.3)	2.45 (1.9)	1.57* (1.6)	1.71* (1.7)
Corrected recognition scores	0.72 (0.18)	0.84 (0.14)	0.77 (0.18)	0.75 (0.18)

Note: Numbers in parentheses represent standard deviations. Notation denotes a significant difference when compared with clinician data for that type of item. \* $p < .05$ , \*\* $p < .01$ ; italicized =  $p < .1$  and  $p > .05$ .

$d=0.63$ ], and they did not differ from the scores of undergraduates ( $p=.54$ ,  $d=0.19$ ; Figure 3). Furthermore, the CR scores of age-matched laypeople were marginally higher than the clinician CR scores [ $t(40)=1.71$ ,  $p=.095$ ,  $d=0.53$ ] and did not differ from the trainees,  $p=.84$ ,  $d=0.06$ .

### Conjectured feature recall

We predicted that experienced clinicians would be recalling fewer facts about patients in part because they would be recalling a few abstracted pieces of encapsulated information. To examine our prediction, we analyzed the number of conjectured features reported across groups. As Figure 4 shows, clinicians did report more conjectured features than all other participant groups. A one-way ANCOVA with Experience as a between-subjects factor and Baseline memory as a covariate completed over the average Coherent conjectured found a main effect of Experience,  $F(3,78)=4.52$ ,  $p=.006$ ,  $\eta_p^2=0.14$ . The covariate of Baseline memory was not significant,  $p=.62$ ,  $\eta_p^2=0.003$ .

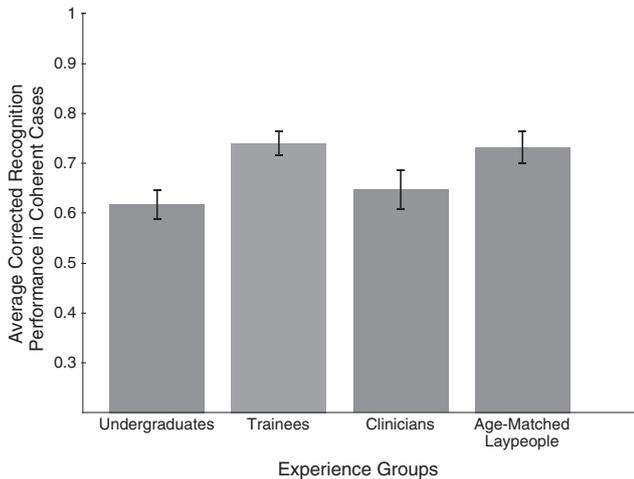


Figure 3. Corrected recognition performance for the Coherent cases

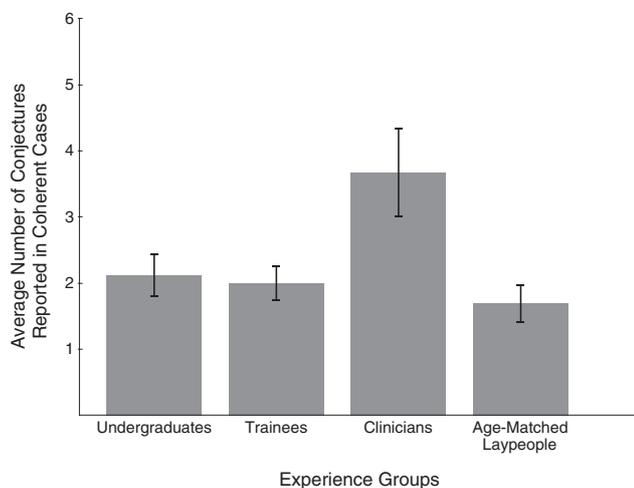


Figure 4. Recall performance for conjectured features in the Coherent cases

Planned independent  $t$ -tests found that clinicians produced significantly more conjectures than trainees [ $t(39)=2.20$ ,  $p=.034$ ,  $d=0.69$ ], undergraduates [ $t(40)=2.11$ ,  $p=.04$ ,  $d=0.65$ ], and age-matched laypeople,  $t(40)=2.76$ ,  $p=.009$ ,  $d=0.85$ .

### Recall in the jumble case

We found fewer significant differences between groups on the Jumble case, suggesting that any advantages or strategies provided by experience are somewhat lost on an incoherent case. Separate one-way ANCOVAs with Experience as the between-subjects variable and Baseline memory as a covariate did not find a significant main effect of Experience for the number of recalled facts ( $p=.23$ ,  $\eta_p^2=0.053$ ) or for the number of recalled specific facts ( $p=.093$ ,  $\eta_p^2=0.078$ ), and a one-way ANOVA with Experience as a between-subjects variable did not find a significant main effect for cued recognition performance ( $p=.11$ ,  $\eta_p^2=0.071$ ). A one-way ANCOVA over the conjectured feature responses with Baseline memory as a covariate did find a significant main effect of Experience [ $F(3,78)=3.86$ ,  $p=.012$ ,  $\eta_p^2=0.12$ ], but independent  $t$ -tests showed that this main effect was not because clinicians reported more conjectures than trainees ( $p=.17$ ,  $d=0.44$ ) but because they reported more than undergraduates [ $t(40)=2.59$ ,  $p=.013$ ,  $d=0.80$ ] and age-matched laypeople,  $t(40)=2.39$ ,  $p=.021$ ,  $d=.74$ .

### Subspecialty effects

Finally, we compared recall performance for each clinician on the case in their area of specialization with the case outside their area of specialization (e.g. eating disorders, substance use). From the medical literature (Rikers, Schmidt, & Boshuizen, 2002; Rikers et al., 2003), we would expect the eating disorder specialists to recall fewer facts and more conjectures for the eating disorder case (their in-area case) than the substance abuse case (their out-of-area case) and vice versa for the substance abuse specialists. Our results somewhat departed from this pattern. We combined the fact and conjectured feature data together to create a single total recall measure (e.g. Brailey et al., 2001; Eva et al., 2002). A mixed ANOVA with Specialization Area (eating disorders versus substance use) between-subjects and Coherent case (anorexia case versus alcoholic case) within-subjects found a significant interaction [ $F(1,19)=6.71$ ,  $p=.018$ ,  $\eta_p^2=0.26$ ], indicating that clinicians were responding differently depending on whether the case was inside or outside their area of expertise. There were no significant main effects, all  $ps > .4$ ,  $\eta_p^2s < 0.04$ . To directly compare the memory of clinicians in and out of their specialization area, we compared recall across specializations for the case in the specialty area of clinicians with the case outside their specialty area through a paired  $t$ -test. Clinicians reported a greater number of total recall items for their in-area case than their out-of-area case,  $t(20)=2.65$ ,  $p=.015$ ,  $d=0.58$ , as can be seen in Table 5.

We can separate the total recall data into facts and conjectured features to see if the overall pattern holds in both sub-types of recall items. The same general pattern shown in the overall data was obtained within each subtype of recall item. Descriptively, clinicians reported a somewhat greater

Table 5. Free recall performance averaged across coherent cases and specializations

	Clinicians in area of specialization	Clinicians out of area of specialization
Overall recall	13.4 (6.2)	11.6 (5.7)
Facts	9.19 (5.2)	8.43 (4.6)
Conjectured features	4.19 (3.4)	3.14 (3.2)

Note: Numbers in parentheses represent standard deviations.

number of conjectured features for their in-area case than for their out-of-area case ( $p = .11$ ,  $\eta_p^2 = 0.12$ ), as would be expected from knowledge encapsulation. Yet, clinicians also produced somewhat more facts for the in-area case than the out-of-area case ( $p = .28$ ,  $\eta_p^2 = 0.05$ ), unlike the pattern expected if knowledge encapsulation was greater in subspecialty areas. To summarize, greater experience within a specialization area led to greater amounts of total information recalled about a patient diagnosed within that specialty area rather than less as found in the medical domain. Conjectured features were recalled more for the in-area case, consistent with knowledge encapsulation. The apparently inconsistent results were because of the trend that even facts were recalled more for the in-area case. We discuss possible reasons for these results in the General Discussion.

## GENERAL DISCUSSION

Our findings support the hypothesis that because of a process of knowledge encapsulation, experience in mental health does change how people remember patient information. Experienced clinicians reported a greater number of conjectures about written patient descriptions than trainee graduate students, inexperienced undergraduate students, or age-matched laypeople. Also, clinicians recalled fewer facts than trainees, and they were less accurate than trainees in recognizing what facts were true of a patient. These findings are similar to those in the medical expertise domain (e.g. Schmidt & Boshuizen, 1993; Schmidt & Rikers, 2007; van de Wiel, Boshuizen, & Schmidt, 2000). The finding that experienced clinicians showed lower performance than age-matched laypeople on recall and recognition of facts indicates that this effect was not a byproduct of age alone. Finally, the effects of experience are seen within specialization areas in complex ways that need further exploration.

Current etiology and nosology of mental disorders are ambiguous, so what might clinicians be using to restructure and encapsulate their knowledge? Clinicians may use their own personal causal theories of the interrelation of disorder symptoms to structure and abstract mental health information (Ahn, Levin, & Marsh, 2005; de Kwaadsteniet, Hagmayer, Krol, & Witteman, 2010; De Los Reyes & Marsh, 2011; Kim & Ahn, 2002). Furthermore, clinicians seem to have developed their own classification systems independent of the DSM that they use to group disorders (e.g. Flanagan & Blashfield, 2007). These idiosyncratic

systems may allow clinicians to encapsulate knowledge and cluster symptoms into higher-order concepts.

## Challenges to existing literatures

Our results provide challenges for several extant findings in the expertise literature. First, we demonstrated an influence of experience in the mental health domain, a domain where it has been notoriously difficult to consistently find such effects. As mentioned in the Introduction, studies using measures such as therapy outcome have found no influences of experience (Garb, 1998) and studies assessing accuracy (e.g. diagnostic judgments) have found weak to no advantage for experience in mental health (Garb, 1989; Spengler et al., 2009). On the contrary, our study focused on the effects of experience on memory and found differences over experience level. Future research is needed to systematically examine the discrepancy and discover what our results mean to clinical practice.

Second, our results run contrary to the findings of Brailey et al. (2001) that clinicians will report more information than intermediate students. A key difference in our studies is that Brailey et al. cued participants to recall DSM-related diagnostic information. DSM diagnostic criteria do not necessarily represent the only relevant information a clinician may need for assessing or treating a patient (see De Los Reyes & Marsh, 2011). Therefore, investigating memory for diagnostic symptoms separately from other life information may not be reflective of how clinicians process information about patients during the therapeutic process. Because we did not cue participants as to what type of information to specifically recall in our free recall task and instead allowed participants to respond with any possible information, it is possible that participants could generate conjectures that spanned across diagnostic and non-diagnostic information, in turn allowing for greater restructuring of information and even fewer facts being recalled.

Third, our comparison of clinicians inside and outside their specialization area does not replicate similar investigations in physicians. Physicians reported the least amount of information for a patient in their area of specialization compared with any other patient (Rikers, Schmidt, & Boshuizen, 2002; Rikers et al., 2003; but see the following for cases where no effects for subspecialty were found, Patel, Groen, & Arocha, 1990; Rikers, Schmidt, Boshuizen, Linssen, Wesseling, & Paas, 2002). Our clinicians reported more information overall in their area of specialization than outside their specialization area, with trends of recalling more conjectures and more facts in their specialty area. One possible reason for the discrepancy is that encapsulated knowledge in the medical domain entirely subsumes facts (e.g. sepsis means having a given set of symptoms), whereas it may not in the mental health domain. For instance, in our anorexia case, the individual features 'weighs self often', 'has trouble stopping eating when she starts', and 'restricts eating' could be encapsulated into the more abstract feature 'fear of loss of control'. However, depending on the patient or theories of a clinician (e.g. Mumma, 1993), there may be multiple ways in which a higher-order concept can be instantiated in mental health such that a conjecture (e.g. 'fear

of loss of control') is not redundant to the symptoms that comprise the conjecture. Mental health experts may be particularly attuned to this lack of redundancy within their own specialization area and therefore report more conjectures and more individual facts in their area of specialized expertise, as we found. Needless to say, further exploration of these specialization comparisons are needed across a wider variety of specialization areas (e.g. childhood disorders, anxiety disorders) before firm claims can be made about differences in medical and mental expertise.

### Implications for the practice of mental health care

Considering real clinical situations, there are additional factors that could potentially increase or decrease knowledge encapsulation. In the following, we discuss some of these factors and the implications of our results for clinical practice and research on clinical decision making.

Our study utilized written patient descriptions to test the memory of participants for patient information; however, most information gained about patients comes through live person-to-person exchanges. What implications do our results have for the memory of clinicians in these real-life sessions? Our transcripts did not convey any of the facial expressions or postural cues that the patients may have displayed in talking about their mental health problems. Experienced clinicians may be able to gain additional cues to the problems of patients by this live experience that would not be immediately available to less experienced practitioners. If so, it is possible that the knowledge encapsulation we captured using written materials is an underestimation of what could happen in real-life cases when experienced clinicians utilize non-verbal cues. Alternatively, it is possible that laypeople may be more likely to make generalizations and conjectures based on minimal visual information (e.g. seeing a patient with possible anorexia that is below ideal weight may lead laypeople to jump to many stereotyping conjectures about the patient; see Corneille, Leyens, Yzerbyt, & Walther (1999) and Yzerbyt, Leyens, & Schadronek (1997) for a similar idea in the application of biases).

In addition, unlike in the current study, clinicians in real-life practice can utilize memory aides by taking notes about patients. Whether such memory aides would improve memory depends on what is recorded. For example, clinicians may note specifics about patients as a way of preventing loss of information in the therapy session. Alternatively, clinicians may note more conjectures that point toward specific diagnostic categories to evaluate (e.g. recording 'anorexic eating patterns'). If this is true, then clinicians may not benefit much in remembering specifics about patients (e.g. the age she previously stated she began purging) from looking at their case notes. Interestingly, whether clinicians note facts may be contingent on whether they realize that their memory tends to encapsulate information in favor of storing veridical, detailed information.

Our results suggest that it is important to study what exactly clinicians are losing information wise about a patient when they encapsulate ideas. Because of the primarily sequential manner of learning information about a patient, a clinician could begin to encapsulate information before

learning all possible information about a patient (for a similar process in causal induction, see Marsh & Ahn, 2006). Would the encapsulation process make case reformulation, or the process of creating a new diagnostic hypothesis, more difficult? For example, if a patient presented with the symptoms 'trouble concentrating at work', 'difficulty sleeping at night', and 'loss of interest in previous hobbies', a clinician could group these symptoms into a plausible preliminary encapsulation of 'depression issues'. Although pursuing whether a diagnosis of depression was warranted, imagine the patient admits to being preoccupied with worry about minor details of her life. If a clinician only remembers the abstract concept of 'depression' about this patient, then the new symptom may be interpreted as stemming from depression. However, if the clinician remembered the specific details that made up the depression encapsulation, then the symptoms of concentration problems and insomnia could be reinterpreted as fitting an encapsulation of 'anxiety issues', with the loss of interest in hobbies stemming from the preoccupying worry of the patient. If the details presented earlier had been lost, then encapsulation could provide the mechanism by which such reformulation may not be feasible. [See Medin & Ross (1989) for a similar discussion of the rationale for preserving specific exemplar information for the sake of future re-classifying of exemplars. Also, see Reiss, Levitan, & Szyszko (1982) and the accompanying diagnostic overshadowing literature for the influence of preliminary diagnosis formation.]

Future research should also focus on the benefits that can be derived from encapsulation in mental health. In our study, we cannot determine if the information conjectured by our clinician participants is actually valid because of the controversial etiologies as well as classifications of mental disorders. Would the same knowledge encapsulation benefits discussed in medicine for the creation of treatment plans that target root causes rather than merely surface symptoms (Schmidt & Rikers, 2007) translate to the ill-defined field of mental health? Future studies focusing on encapsulation, regardless of validity *per se*, could help inform the practices by which we train future mental health clinicians.

### CONCLUSION

The great expense to society involved in the health care of people with mental illnesses demands study of mental health practitioners. Our study has focused on the abstraction of information and its effects on the memory of mental health professionals for patients. We have demonstrated parallels in how memory functions in the mental health and medical health fields. Understanding in more depth how these experience differences develop in mental health and how they differ from other domains can provide a better understanding of how experience changes professionals within the mental health domain.

### ACKNOWLEDGEMENTS

The authors would like to thank Batel Isenstein and Haley York for their help in administering this project, and the first

author's dissertation committee for feedback on this project that was done in partial fulfillment of her dissertation. Preliminary versions of some findings were presented at the 30th Annual Conference of the Cognitive Science Society. Support for this research was provided in part by Yale University Psychology Department graduate research funds awarded to the first author, the NIH grant MH 57737 awarded to the second author, and Texas Tech University research funds provided to the first author.

## REFERENCES

- Ahn, W., Levin, S., & Marsh, J. K. (2005). Determinants of feature centrality in clinicians' concepts of mental disorders. In B. Bara, L. Barsalou, & M. Bucciarrelli (Eds), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 67–72). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn, text revision). Washington, DC: Author.
- Berman, J. S., & Norton, N. C. (1985). Does professional training make a therapist more effective? *Psychological Bulletin*, *98*, 401–407. doi:10.1037//0033-2909.98.2.401
- Beutler, L. E., Machado, P. P. P., & Neufeldt, S. A. (1994). Therapist variables. In A. E. Bergin & S. L. Garfield (Eds), *Handbook of psychotherapy and behavior change* (4th edn, pp. 229–269). Oxford: John Wiley & Sons.
- Boshuizen, H. P., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, *16*, 153–184. doi:10.1207/s15516709cog1602\_1
- Brailey, K., Vasterling, J. J., & Franks, J. J. (2001). Memory of psychodiagnostic information: Biases and effects of expertise. *The American Journal of Psychology*, *114*, 55–92. doi:10.2307/1423381
- Brammer, R. (2002). Effects of experience and training on diagnostic accuracy. *Psychological Assessment*, *14*, 110–113. doi:10.1037//1040-3590.14.1.110
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726. doi:10.1016/S0022-5371(72)80006-9
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81. doi:10.1016/0010-0285(73)90004-2
- Chi, M. T. H., Glaser, R., & Farr, M. (Eds). (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.
- Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education*, *19*, 61–67. doi:10.1111/j.1365-2923.1985.tb01140.x
- Corneille, O., Leyens, J., Yzerbyt, V., & Walther, E. (1999). Judgeability concerns: The interplay of information, applicability, and accountability in the overattribution bias. *Journal of Personality and Social Psychology*, *76*, 377–387.
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York, NY: The Free Press.
- de Kwaadsteniet, L., Hagmayer, Y., Krol, N. P. C. M., & Witteman, C. L. M. (2010). Causal client models in selecting effective interventions: A cognitive mapping study. *Psychological Assessment*, *22*, 581–592. doi:10.1037/a0019696
- De Los Reyes, A., & Marsh, J. K. (2011). Patients' contexts and their effects on clinicians' impressions of conduct disorder symptoms. *Journal of Clinical Child and Adolescent Psychology*, *40*, 479–485. doi:10.1080/15374416.2011.563471
- de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving* (pp. 19–50). New York, NY: Wiley.
- Durlak, J. A. (1979). Comparative effectiveness of paraprofessional and professional helpers. *Psychological Bulletin*, *86*, 80–92.
- Ebling, R., & Levenson, R. W. (2003). Who are the marital experts? *Journal of Marriage and the Family*, *65*, 130–142. doi:10.1111/j.1741-3737.2003.00130.x
- Eells, T. D., Lombart, K. G., Kendjelic, E. M., Turner, L. C., & Lucas, C. P. (2005). The quality of psychotherapy case formulations: A comparison of expert, experienced, and novice cognitive-behavioral and psychodynamic therapists. *Journal of Consulting and Clinical Psychology*, *73*, 579–589.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *The American Psychologist*, *49*, 725–747. doi:10.1037/0003-066X.49.8.725
- Eva, K. W., Norman, G. R., Neville, A. J., Wood, T. J., & Brooks, L. R. (2002). Expert-novice differences in memory: A reformulation. *Teaching and Learning in Medicine*, *14*, 257–263. doi:10.1207/S15328015TLM1404\_10
- Faust, D., & Zlotnick, C. (1995). Another dodo bird verdict? Revisiting the comparative effectiveness of professional and paraprofessional therapists. *Clinical Psychology & Psychotherapy*, *2*, 157–167. doi:10.1002/cpp.5640020303
- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds), *The Cambridge handbook of expertise and expert performance* (pp. 41–67). New York, NY: Cambridge University Press.
- Flanagan, E., & Blashfield, R. (2007). Clinicians' folk taxonomies of mental disorders. *Philosophy, Psychiatry, & Psychology*, *14*, 249–269. doi:10.1353/ppp.0.0123
- Ganzach, Y. (1997). Theory and configularity in clinical judgments of expert and novice psychologists. *Journal of Applied Psychology*, *82*, 954–960.
- Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin*, *105*, 387–396. doi:10.1037/0033-2909.105.3.387
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, *1*, 67–89.
- Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds), *The nature of expertise* (pp. xv–xxviii). Hillsdale, NJ: Lawrence Erlbaum.
- Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition*, *15*, 95–110. doi:10.1016/0010-0277(83)90035-5
- Hauser, S., Spada, H., & Rummel, N. (2007). The effects of practical experience on expertise in clinical psychology and collaboration. In D. S. McNamara & J. G. Trafton (Eds), *Proceedings of the 29th annual cognitive science society* (pp. 1061–1066). Austin, TX: Cognitive Science Society.
- Hauser, S., Spada, H., Rummel, N., & Meier, A. (2006). Expertise development in clinical psychology. In R. Sun, N. Miyake & C. D. Schunn (Eds), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1458–1463). Mahwah: Erlbaum.
- Hill, C. E. (2005). Therapist techniques, client involvement, and the therapeutic relationship: Inextricably intertwined in the therapy process. *Psychotherapy: Theory, Research, Practice, Training*, *42*, 431–442. doi:10.1037/0033-3204.42.4.431
- Hillerbrand, E., & Claiborn, C. D. (1990). Examining reasoning skill differences between expert and novice counselors. *Journal of Counseling and Development*, *68*, 684–691.
- Hooten, D. H. (2011). Essentialism beliefs about mental health categories: Examining the impact of experience and training. (Unpublished doctoral dissertation). Texas Tech University, Lubbock, TX.
- Hunsley, J., & Lee, C. M. (2010). *Introduction to clinical psychology*. Hoboken: Wiley & Sons.
- Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology. General*, *131*, 451–476. doi:10.1037//0096-3445.131.4.451
- Kraemer, H. C., Stice, E., Kazdin, A. E., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *The American Journal of Psychiatry*, *158*, 848–856. doi:10.1176/appi.ajp.158.987.848

- Lambert, L. E., & Wertheimer, M. (1988). Is diagnostic ability related to relevant training and experience. *Professional Psychology-Research and Practice, 19*, 50–52.
- Marsh, J. K. B. (2008). The interaction of knowledge and categorization in reasoning: Evidence from causal reasoning and mental health clinicians' reasoning (Unpublished doctoral dissertation). Retrieved from ProQuest Digital Dissertations. (AAT 3317165).
- Marsh, J. K., & Ahn, W. (2006). Order effects in contingency learning: The role of task complexity. *Memory and Cognition, 34*, 568–576. doi:10.3758/BF03193580
- Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189–223). Hillsdale, NJ: Lawrence Erlbaum.
- Mumma, G. (1993). Categorization and rule induction in clinical diagnosis and assessment. In G. V. Nakamura, D. L. Medin, & R. Taraban (Eds), *Psychology of learning and motivation* (Vol. 29, pp. 283–326). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(08)60143-7
- Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*, 91–116. doi:10.1016/S0364-0213(86)80010-6
- Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory and Cognition, 18*, 394–406. doi:10.3758/BF03197128
- Patel, V. L., Groen, G. J., & Frederiksen, C. H. (1986). Differences between medical-students and doctors in memory for clinical cases. *Medical Education, 20*, 3–9. doi:10.1111/j.1365-2923.1986.tb01033.x
- Reiss, S., Levitan, G. W., & Szyszko, J. (1982). Emotional disturbance and mental retardation: Diagnostic overshadowing. *American Journal of Mental Deficiency, 86*, 567–574.
- Rikers, R. M. J. P., Loyens, S. M. M., & Schmidt, H. G. (2004). The role of encapsulated knowledge in clinical case representations of medical students and family doctors. *Medical Education, 38*, 1035–1043. doi:10.1111/j.1365-2929.2004.01955.x
- Rikers, R. M. J. P., Schmidt, H. G., & Boshuizen, H. P. A. (2002). On the constraints of encapsulated knowledge: Clinical case representations by medical experts and subexperts. *Cognition and Instruction, 20*, 27–46. doi:10.1207/S1532690XCI2001\_2
- Rikers, R. M. J. P., Schmidt, H. G., Boshuizen, H. P. A., Linssen, G. C. M., Wesseling, G., & Paas, F. G. W. C. (2002). The robustness of medical expertise: Clinical case processing by medical experts and subexperts. *The American Journal of Psychology, 115*, 609–629. doi:10.2307/1423529
- Rikers, R., te Winkel, W., Loyens, S., & Schmidt, H. (2003). Clinical case processing by medical experts and subexperts. *The Journal of Psychology: Interdisciplinary and Applied, 137*, 213–223. doi:10.1080/00223980309600609
- Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging, 30*, 507–514. doi:10.1016/j.neurobiolaging.2008.09.023
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In D. A. Evans, & V. L. Patel (Eds), *Advanced models of cognition for medical training and practice* (pp. 265–282). New York, NY: Springer Verlag.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. *Memory and Cognition, 21*, 338–351. doi:10.3758/BF03208266
- Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education, 41*, 1133–1139. doi:10.1111/j.1365-2923.2007.02915.x
- Spengler, P. M., White, M. J., Ægisdóttir, S., & Maugherman, A. S. (2009). Time keeps on ticking: The experience of clinical judgment. *The Counseling Psychologist, 37*, 416–423. doi: 10.1177/0011000009332008
- Strasser, J., & Gruber, H. (2004). The role of experience in professional training and development of psychological counselors. In H. P. A. Boshuizen, R. Bromme & H. Gruber (Eds), *Professional learning: Gaps and transitions on the way from novice to expert* (pp. 11–27). Dordrecht: Kluwer Academic Publishing.
- Strupp, H. H., & Hadley, S. W. (1979). Specific vs. nonspecific factors in psychotherapy: A controlled study of outcome. *Archives of General Psychiatry, 36*, 1125–1136.
- Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*, 230–241. doi:10.1037/0882-7974.13.2.230.
- van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (2000). Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians. *European Journal of Cognitive Psychology, 12*, 323–356. doi:10.1080/09541440050114543
- van de Wiel, M. W. J., Boshuizen, H. P. A., Schmidt, H. G., & Schaper, N. C. (2000). The explanation of clinical concepts by expert physicians, clerks, and advanced students. *Teaching and Learning in Medicine, 11*, 153–163. doi:10.1207/S15328015TL110306
- van de Wiel, M. W. J., Schmidt, H. G., & Boshuizen, H. P. (1998). A failure to reproduce the intermediate effect in clinical case recall. *Academic Medicine, 73*, 894–900. doi:10.1097/00001888-199808000-00015
- Witteaman, C. L. M., & van den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment, 23*, 56–61.
- World Health Organization. (2001). *The world health report 2001—Mental health: New understanding, new hope*. Geneva: Author.
- Yzerbyt, V., Leyens, J., & Schadronek, G. (1997). Social judgeability and the dilution of stereotypes: The impact of the nature and sequence of information. *Personality and Social Psychology Bulletin, 23*, 1312–1322.

## APPENDIX

**Coding Scheme Creation**

Two coders independently analyzed the four stories used in the experiment in order to divide the stories into individual fact units. A fact unit was defined as a group of text in the story that conveyed a single piece of information about the person described. Fact units could range in size from fragments of sentences to groups of sentences, depending on the information conveyed. After making their own fact unit divisions, the two coders compared their coding guides and finalized through discussion what constituted a fact unit in each story. These finalized fact units constituted the master coding guide to be used for coding all of the collected recall data. Because the cases had been matched for length and content, the Coherent, Jumble, and Baseline cases received an identical number of fact unit codes (i.e. 42 units).

**Coding of Recall Items**

The two independent coders separately coded the free recall responses using the coding guide. The coders received lists of the responses for each case that were randomized and stripped of individuating participant information to avoid knowing from which participant group each response came. The coders assigned the free recall responses to one of four categories: facts, conjectured features, intrusions, or evaluations. Responses coded as facts also received a number corresponding to which fact unit from the coding guide it represented (e.g. Fact 1, Fact 42, etc.). Responses were coded as facts even if recall for the fact unit did not contain every detail accurately. Coders split responses listing information that fit more than one response category or fact unit in the same response into separate recall units. After

each coder had completed the coding separately, the coded responses were compared for agreement. Coders showed high overall agreement in parsing the responses into the four response categories, agreeing 87% of the time. Within the fact category, the coders agreed which of the 42 fact units a fact represented 95% of the time.<sup>6</sup>

After agreement analyses were completed, free recall responses were further analyzed to determine if any participant received credit more than once for the same fact unit from a given story. For example, a participant may have split information about one fact unit into multiple responses (e.g. a participant provided the separate responses of 'he has gotten DUIs' and 'has two DWIs', which both refer to the same fact unit from the coding guide). This response pattern would have resulted in the participant receiving credit for the same fact unit twice because the coders were blind to who produced each response. The coders jointly analyzed the responses for this type of repetitive responding and eliminated or combined duplicate responses as appropriate.

After establishing the final response categories, the responses coded as facts were coded for their level of detail in recalling the original fact. The two coders received all of the facts randomized and stripped of participant information. If a fact was a veridical recalling of the coded fact unit, then the fact was coded as a specific fact recall item (e.g. a participant response of 'he drinks about two and a half cases of beer a week' for the fact unit 'Drinks 2.5 cases of beer a week'). All of the other facts were coded as a non-specific fact recall item (e.g. a response of 'he drinks beer very often' for the fact unit 'Drinks 2.5 cases of beer a week'). Across the stories, the coders agreed on the detail scoring 89% of the time. Disagreements in any step of the coding were settled through discussion between the two involved coders.

<sup>6</sup> The original blind coder was not available to code the last set of collected data (data from 25 participants), so a new blind coder was used. This new coder was given a subset of the recall items coded by the original coders (i.e. 15% of their coded data) and agreed with 91% of the basic four category parsing and with 94% of the specific fact codes. Therefore, the reported agreement averages across all three coders.