

Frank J. Papa* and Feiming Li

Evidence of the preferential use of disease prototypes over case exemplars among early year one medical students prior to and following diagnostic training

DOI 10.1515/dx-2015-0024

Received July 30, 2015; accepted October 25, 2015

Abstract

Background: Two core dual processing theory (DPT) System I constructs (Exemplars and Prototypes) were used to: 1) formulate a training exercise designed to improve diagnostic performance in year one medical students, and 2) explore whether any observed performance improvements were associated with preferential use of exemplars or prototypes.

Methods: With IRB approval, 117 year one medical students participated in an acute chest pain diagnostic training exercise. A pre- and post-training test containing the same 27 case vignettes was used to determine if the subjects' diagnostic performance improved via training in both exemplars and prototypes. Exemplar and Prototype theory was also used to generate a unique typicality estimate for each case vignette. Because these estimates produce different performance predictions, differences in the subjects' observed performance would make it possible to infer whether subjects were preferentially using Exemplars or Prototypes.

Results: Pre- vs. post-training comparison revealed a significant performance improvement; $t=14.04$, $p<0.001$, Cohen's $d=1.32$. Pre-training, paired t-testing demonstrated that performance against the most typical vignettes>mid typical vignettes: $t=4.94$, $p<0.001$; and mid typical>least typical: $t=5.16$, $p<0.001$. Post-training, paired t-testing again demonstrated that performance against the most

typical vignettes>mid typical: $t=2.94$, $p<0.01$; and mid typical>least typical: $t=6.64$, $p<0.001$. These findings are more consistent with the performance predictions generated via Prototype theory than Exemplar theory.

Conclusions: DPT is useful in designing and evaluating the utility of new approaches to diagnostic training, and, investigating the cognitive factors driving diagnostic capabilities among early medical students.

Keywords: Dual Processing Theory; Exemplar Theory; Prototype Theory.

Introduction

One often overlooked impediment to diagnostic accuracy is the 'ill-defined' nature of most human diseases. More specifically, while all diseases are associated with a list of characteristic signs and symptoms (S/S), most lack a clearly defined, criteria meeting set of necessary and/or sufficient S/S with which to confidently confer a diagnosis at the bedside [1]. In the absence of well-defined bedside diagnostic criteria, an ill-defined disease can subsequently present with many different combinations of its characteristic S/S, with each unique combination representing an equally valid case presentation of that disease [1]. For example, while a patient with an ill-defined disease associated with 10 characteristic S/S can only produce one possible combination of those 10 S/S, a patient presenting with nine of those 10 characteristic S/S can present, via combinatorial mathematics, with 10 distinct combinations of S/S. A patient presenting with only eight of those 10 characteristic S/S yields 45 distinct combinations, while seven of 10 yields 120, six of 10 yields 210, and five of 10 yields 252 combinations. Theoretically, an ill-defined disease with 10 characteristic S/S could potentially manifest with well over 600 distinct yet equally valid patient presentations. Simply put, the ill-defined nature

*Corresponding author: Frank J. Papa, DO, PhD, University of North Texas Health Science Center – Medical Education, 3500 Camp Bowie, Fort Worth, 76107 TX, USA, Phone: +8177352988, Fax: +8177350417, E-mail: frank.papa@unthsc.edu

Feiming Li: University of North Texas Health Science Center, Center for Learning and Development, Fort Worth, TX, USA

of human diseases makes differential diagnosis a very difficult, inherently error prone task.

What cognitive mechanisms enable physicians to optimize the likelihood of a correct diagnosis when faced with an ill-defined disease that can present with numerous distinctly different, yet equally valid case portrayals? Medical educators are increasingly utilizing dual processing theory (DPT) as a research framework to explore the cognitive factors enabling the performance of differential diagnosis (DDX) in the hopes of improving training, assessment, and thereby diagnostic performance against ill-defined diseases [2–6].

Background

Dual processing theory

DPT suggests that two distinct, yet interrelated cognitive systems play critical roles during ill-defined categorization or classification tasks such as DDX (i.e. system I and system II – also referred to as type I and type II) [7–9]. As applied to DDX, system I is theorized as a rapid, non-analytical, pattern recognition or ‘similarity-driven’ approach to diagnostic reasoning while system II enables a more conscious, analytical or ‘rule-based’ approach. Each system in turn, is comprised of at least two core sub-components: 1) system-specific information processing mechanisms, and 2) system-specific knowledge. Given the many considerations required to provide but a brief overview of system I, and how it served as a framework for this investigation, there will be no further description of system II.

System I, a similarity-driven approach to diagnostic reasoning, is theorized to utilize information processing mechanisms which compare and contrast its ‘knowledge of what a given disease looks like’ with the collective set of signs and symptoms in a new case presentation. System I’s knowledge of individual diseases is further theorized as represented in memory in two distinct forms or knowledge structures called ‘disease exemplars’ and ‘disease prototypes’. The following provides a more detailed, yet overly simplified description of how system I’s two core knowledge structures are used to perform DDX.

Exemplars and prototypes

A disease exemplar is generally defined as a record of the signs and symptoms associated with each individual, previously experienced case [10, 11]. Exemplars are recorded

in Episodic memory; the cognitive space responsible for retaining the details of prior experiences (see Figure 1). Each and every stored disease exemplar adds to the physician’s knowledge of what that disease looks like. Diagnostic reasoning (categorization/classification) via exemplars is theorized to occur via System I information processing mechanisms comparing the similarity of the set of S/S in a new case presentation with the S/S of all stored, individual exemplars (prior case presentations) representative of all previously experienced diseases. These mechanisms determine which single stored exemplar (and its S/S) exactly matches, or is most similar to, the pattern of S/S in the new case. The mechanism(s) then ascribes the best matching exemplar’s diagnosis as the diagnosis for the new case.

A disease prototype is generally defined as a single, averaged portrayal of the S/S that characterize a given disease [12–14]. Prototypes are stored in Semantic memory; the cognitive space responsible for storing an abstracted representation of a given concept (e.g. “what disease ‘X’ looks like”). The formation of a disease’s prototype can begin via readings and lectures listing the disease’s characteristic S/S. However, as the number and variety of experiences with cases presenting as that particular disease increases, the individual’s prototype begins to incorporate estimates of the frequency with which each of its associated S/S occurred in a population of case portrayals (e.g. approximately 90% of patients with disease ‘X’ have S/S “a”, while 60% have S/S “d”). These frequency estimates lead to a more robust representation of disease ‘X’s’ prototype via the additional representation of those S/S that occur more or less frequently with the disease at hand (see Figure 2). Furthermore, these frequency estimates can enable the identification of those S/S useful for

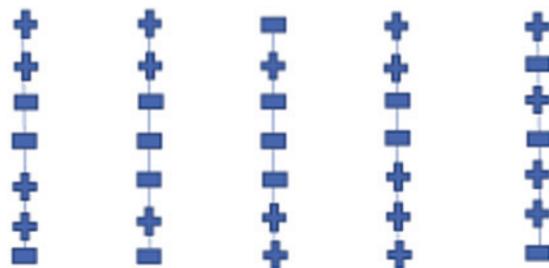


Figure 1: Storage of case exemplars of a given disease in Episodic memory – per Exemplar Theory.

The figure portrays how five distinct case exemplars, representative of a given disease, are stored in Episodic Memory. Seven findings (S/S) are associated with each case exemplar. For each of the seven S/S, a positive (+) or negative (-) symbol indicates whether a given S/S was present (+) or absent (-) during the patient encounter.

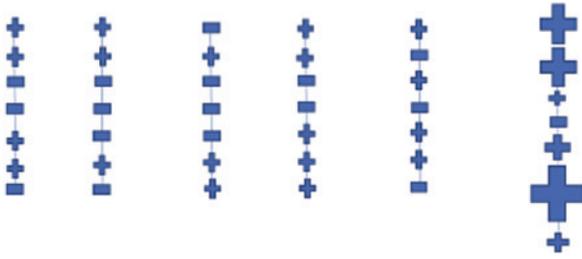


Figure 2: Storage and transformation of case exemplars into a Disease Prototype in Semantic memory – per Prototype theory. The Figure portrays how five distinct case exemplars (stored in Episodic memory and representative of a given disease), are used to create a Prototypical portrayal of that disease, and, how that disease’s ‘Prototype’ is represented in Semantic memory. The Prototype (on the far right) is created by averaging the number of times each of the disease’s seven characteristic S/S was present across all of the five case exemplars (represented on the left). A single, summarized disease ‘Prototype’ is subsequently stored in Semantic memory. On viewing the prototype, note that the bigger the Prototype’s positive (+) symbol, the more frequently that particular S/S was present across all representative case exemplars. Note that the Prototype’s largest positive sign (next to bottom S/S) reflects the fact that this S/S was present in all five exemplars. Further note that the Prototype’s next most frequently encountered S/S was both the first and second S/S (i.e. the top two S/S) with each present in four of the five exemplars. Finally, note that the Prototype lists a negative sign (middle of the seven S/S) indicating that none of the five exemplars was found to have that S/S in any of the patient encounters.

differentiating one disease from one or more competing diseases [15]. Diagnostic reasoning (categorization/classification) via prototypes occurs via System I information processing mechanisms comparing the similarity of the set of S/S in a new case with the prototypical portrayal of each disease stored in Semantic memory. The diagnostic category (disease name) associated with the best matching disease prototype serves as the diagnosis for the new case presentation.

The importance of knowledge in differential diagnosis

Over the course of several decades, medical education researchers have generally agreed that it is the knowledge stored in Episodic or Semantic memory (i.e. knowledge of what diseases look like, as exemplified by the two distinct constructs referred to as disease exemplars and prototypes) which is primarily responsible for diagnostic accuracy or error, rather than information processing mechanisms [16]. However, there has long been intense disagreement as to whether exemplars or prototypes serves as the knowledge structure preferentially called upon during categorization tasks such as DDX [14]. Evidence that one knowledge structure is more closely associated with diagnostic accuracy or

error would be very useful for: 1) constructing more efficient and effective instructional and assessment approaches for students, residents and practitioners, and 2) identifying, supplementing and/or correcting the malformed or biased knowledge structures leading to diagnostic errors. In recent years, functional magnetic resonance imaging (fMRI) has been used to directly identify which regions of the brain are activated during categorization tasks in the hopes of linking activated brain regions (believed to store semantic or episodic memories) to the performance of a given task [17]. Unfortunately, fMRI is not an easily deployable research methodology. Subsequently, determinations as to whether exemplars or prototypes are preferentially utilized by clinicians are made indirectly, and largely predicated upon the different ‘performance predications’ offered by Exemplar and Prototype theories.

Case typicality as a predictor of diagnostic performance

In addition to evidence demonstrating that knowledge (rather than information processing mechanisms) is the primary driver of diagnostic performance, there is also evidence demonstrating that diagnostic accuracy is a function of a case’s typicality (a phenomena referred to as the typicality/performance gradient) [18, 19]. That is, the more typical the case, the more likely it will be correctly diagnosed; the less typical, the less likely. In formulating an understanding of how a case’s typicality assignment is used to predict the probability that it will be correctly or incorrectly diagnosed, it is critical to understand that Exemplar and Prototype theories define and ascribe a measure of a case’s typicality very differently [14, 20, 21]. Thus, subject performance more in accord with the predictions of Exemplar vs. Prototype theories makes it possible for researchers to draw indirect inferences as to whether exemplars or prototypes are preferentially utilized in a given population of subjects (e.g. students in their pre-clinical training, clinical training, residency or practitioners). The following broadly describes the method by which Exemplar and Abstraction Theorists define typicality within the context of differential diagnosis. Differences in these typicality assignments lead to difference in how each theory predicts subject performance against a panel of test cases.

Exemplar theories tend to define typicality in terms of the ‘frequency’ with which, or number of times, a case presenting with a specific set or combination of S/S has been previously experienced, and subsequently, the number of times that particular set of S/S has been stored in Episodic memory (see Figure 3). Given ill-defined disease ‘X’ with 10 characteristic S/S and tens to hundreds of distinct

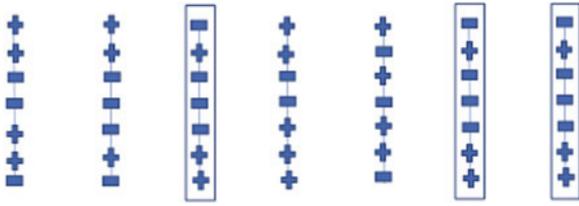


Figure 3: The generation of an Exemplar theory based assignment of a case's typicality.

The figure represents how Exemplar theory posits that the typicality of a given case is a function of the number of times that a case consisting of a specific set of S/S has been previously encountered, and subsequently, the number of times it is stored or instantiated in Episodic memory. Of the seven case exemplars portrayed above, note that the third, sixth and seventh case (each enclosed in a rectangle) have exactly the same set of S/S. Thus, this particular set of S/S would be said to represent the most typical case portrayal of that disease because it is instantiated in memory on three occasions. Note that each of the remaining four cases are both distinct from each other, and, different from the case instantiated on three prior occasions. Each of these four distinct cases would be considered a much less typical portrayal of the respective disease because each is only instantiated on one occasion. A test case exactly like one of the four distinct cases would much less likely be correctly diagnosed than a test case exactly like the one instantiated in Episodic memory on three previous occasions.

combinations of S/S by which a case representative of disease 'X' might present for diagnosis, a more typical case exemplar would be one whose specific combination of S/S (a, c, d, f, g) was experienced in more patient case presentations than an exemplar consisting of a different and yet less frequently experienced combination of S/S (b, c, e, f, j). Exemplar theories would predict that a new test case containing S/S exactly like, or very similar to a 'very typical' (more frequently instantiated) exemplar would more likely be correctly diagnosed than a new test case containing S/S exactly like, or very similar to a 'less typical' (less frequently instantiated) exemplar.

Considering that tens to hundreds of different exemplars are possible for ill-define disease 'X', Exemplar theories would predict that relatively inexperienced individuals with very few stored exemplars (such as the subjects in this investigation) would not have experience sufficient to determine which specific exemplars were more or less frequently encountered, and thereby, more or less typical portrayals of disease 'X'. That is, it would be unlikely that they experienced one specific case exemplar on more than one occasion considering the tens to hundreds of possible exemplars that could represent a case of disease 'X'. Therefore, Exemplar theories would predict that inexperienced individuals would be unlikely to correctly diagnose many new test cases of a given ill-defined

disease. Furthermore, inexperienced individuals would demonstrate little to no evidence of a typicality/performance gradient when challenged to diagnose a panel of new test cases representing any given combination of S/S with which disease 'X' might present.

Prototype theorists define typicality in terms of the 'degree to which' the S/S in a case approximates a given disease's prototype (see Figure 4). A disease's prototype is generally defined as representing: 1) a list of the disease's characteristic S/S, 2) an estimate of the frequency with which its characteristic S/S occur with the disease at hand (i.e. ranging from more to less frequently), and 3) containing information sufficient to enable an individual to discern which of these more or less frequently encountered S/S could be used to help differentiate the prototype at hand from competing prototypes.

In general, Prototype theories would posit that a more typical case portrayal contains more of that disease prototype's characteristic S/S, more of that disease's higher frequency S/S, and more of its distinguishing S/S. A less typical case would contain fewer of that prototype's characteristic S/S, fewer higher frequency S/S, and fewer distinguishing S/S. Subsequently, Prototype Theorists would therefore predict that the more closely a new test case approximated the characteristic, frequent

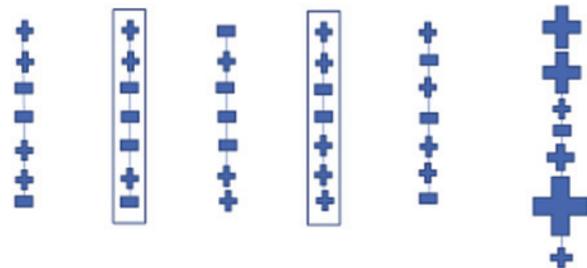


Figure 4: The generation of a Prototype theory based assignment of a case's typicality.

The figure represents how Prototype theory posits that the typicality of a given case is a function of the degree to which it 'approximates' the signs and symptoms associated with that disease's 'prototypical portrayal'. Thus, a test case that more closely approximates its respective disease's prototypical portrayal would much more likely be correctly diagnosed than one that less closely approximated its respective disease's prototypical portrayal. Note that of the five cases portrayed above, the fourth from the left is a closer approximation of the disease's prototype (far right) than the second of the five cases. Prototype theory would predict that a test case with the same set of S/S as the fourth case would much more likely be correctly diagnosed than a test case with the same set of S/S as the second case. It is worth noting that according to Prototype theory, a case's typicality is not necessarily a function of the number of times that a specific case with a specific set of S/S has been instantiated in Episodic memory.

and discriminating S/S of a given disease's prototype, the more likely it would be correctly diagnosed, with a less typical test case less likely to be correctly diagnosed [15]. Prototype theories would subsequently predict that relatively inexperienced individuals (the subjects in this study) would still demonstrate evidence of a typicality/performance gradient when challenged to diagnose a panel of new test cases representing any given combination of S/S with which disease 'X' might present. Thus, Prototype theories predict performance against a new test case in a manner very different than Exemplar theories.

Study preview

This study utilized a DPT/System I framework to achieve two objectives. First, to design an approach to diagnostic training utilizing both exemplars and prototypes in the hopes of enabling students to achieve significant improvements in diagnostic performance against nine common and important ill-defined disease differentials for the problem of acute chest pain. Use of the same set of test cases in both pre- and post-training would enable the direct determination of any performance improvements. Second, to determine whether exemplars or prototypes were preferentially utilized to diagnose new test cases. Pursuit of the second objective would involve inferences based upon: 1) differences in each test case's typicality measure as assigned by Exemplar and Prototype theories, 2) the use of these assignments to predict the subjects' performance, and 3) the determination of which theory best predicted the observed performance.

The subjects in this study were year one, third month medical students. Given that each of the nine ill-defined diseases differentials serving as the study's training objective could theoretically be represented via tens to hundreds of distinctly different case portrayals, it was assumed that these early medical students likely had little if any case-based experiences stored in Episodic memory to serve as exemplars for each of these differentials. Given the study's training exercise was designed to provide the subjects only four distinctly different training cases for each differential, and, the assumption that any previously stored case exemplars (and each one's particular set of S/S) would be represented in Episodic memory on only one occasion, Exemplar theories would posit that no exemplar would come to represent a more or less typical portrayal for each of the nine ill-defined differentials. Exemplar theories would therefore predict little if any evidence of a typicality/performance gradient against a panel of test cases (i.e. no test case would more or less likely be correctly diagnosed than any other test case).

However, in this investigation a previously validated, prototype theory-oriented, mathematically-based method (involving the use of S/S frequency estimates) was used to assign a typicality measure to each of the study's training and test cases. Prototype Theorists would predict that test cases with a higher typicality estimate were more likely to be correctly diagnosed than those with a lower typicality estimate even among students with relatively little experience with the diseases at hand. Evidence that the subjects' performance increased as the typicality of test cases increased (i.e. demonstration of a typicality/performance gradient) would make it possible to draw the inference that the subjects' diagnostic performance was driven by their preferential use of disease prototypes.

Materials and methods

Subjects

Following an IRB approved protocol and informed student consent, 117 medical students very early into their medical training (first year; third month of semester one) volunteered to serve as study subjects. These students were recruited because their curriculum had not yet formally exposed them to: 1) a prototypical portrayal of these nine diseases, 2) any case examples of the diseases used in the study, nor 3) training in the process of DDX. Further, recruitment of these very early medical students also made it possible to assume that their pre-matriculation experiences provided them little opportunity to: 4) accumulate and store many, if any case exemplars for the diseases studied (in terms of complete, bedside based historical and physical S/S), 5) instantiate more than one copy of any specific exemplar, and 6) instantiate exemplars exactly matching the practice or test cases employed in this study.

Context

The problem of acute chest pain and nine if its more common and important differentials (Myocardial Infarction, Coronary Ischemia, Dissecting Thoracic Aneurysm, Pericarditis, Upper Gastrointestinal etiology, Pneumonia, Pneumothorax, Musculoskeletal etiology and Pulmonary Embolus) served as this study's instructional context. Despite their commonality and importance, all nine differentials selected have ill-defined bedside diagnostic criteria, and yet as portrayed in this study, had from eight to 12 characteristic S/S attributed to each disease. For example, the characteristic S/S associated with Myocardial Infarction included: pain described as dull, sub-sternal location of pain, pain described as acute in onset, radiation of pain to neck or arm, shortness of breath, nausea/vomiting, diaphoresis, weakness, tachycardia, tachypnea, crackles in lungs and an S4 gallop. Given no defining or criteria meeting set of clinical S/S by which the diagnosis of Myocardial Infarction could be confidently determined at the bedside, tens to hundreds of distinct yet equally valid case portrayals could be used as training or testing cases.

Case typicality assignment

Every case selected for use as a practice or testing case was assigned two typicality measures with one derived from the application of Exemplar theory and the other via Prototype theory. A previously validated research tool was used to assign a Prototype theory derived typicality estimate to each of the 36 practice cases and 27 test cases [15, 18, 19, 22]. In brief this tool, an Expert System Shell, utilizes the knowledge of board certified practitioners with expertise in the problem at hand (in the form of estimates of the frequency with which a given disease is associated with each of its characteristic S/S) to create a frequency-based portrayal of each disease's prototype for the problem at hand. Based upon the tool's internal representation of each disease's prototype, the tool uses Monte Carlo procedures to generate tens to hundreds of case vignettes for each disease. Additional algorithms are used to screen the generated cases for evidence of internal validity, with cases passing this screening process subsequently assigned an estimate of the degree to which each case approximates the prototypical portrayal of its respective disease.

The resultant typicality measure is then used to arrange all the validated cases for each differential 'along a case typicality gradient' (i.e. arranged from most thru least typical). For each of the nine differentials, four distinctly different cases at different points along the typicality gradient were selected to serve as practice cases, thus resulting in a total of 36 practice cases. Similarly, for each of the nine differentials, three distinctly different cases at different points along the typicality gradient were selected to serve as test cases, thus resulting in a total of 27 test cases. None of the resulting seven cases selected to represent each differential were duplicates.

Given the study subjects had neither formal curricular training nor little pre-matriculation exposure to the diseases addressed in this study, it was assumed that cases 'exactly like' the 36 practice and 27 test cases selected via the previously described procedure were unlikely to have been previously instantiated in the subjects' Episodic memory. The subjects' exposure to these 36 practice and 27 test cases during the study would cause these specific cases to be instantiated in Episodic memory on only one occasion (i.e. either during training or testing). From the perspective of Exemplar theory, all practice and testing cases should therefore be assigned the same typicality measure (i.e. a frequency of instantiation in Episodic memory equal to one). Based upon this rationale, all practice and test cases would fall on the same point along an Exemplar theory based case typicality gradient.

Implications of an Exemplar and Prototype theory based approach to case typicality assignments

As previously mentioned, while both Exemplar and Prototype theories produce an estimate of a case's 'typicality', the typicality measure assigned to a given case via either theory will be very different. From an Exemplar theory perspective, given all the cases used in this study would fall on the same point along a typicality gradient (i.e. an approach based upon the frequency with which a given specific case had been previously instantiated in Episodic memory), no test case would more likely be correctly diagnosed than any other test case. From a Prototype theory perspective, given all the cases used in this study fell on a different point along a typicality gradient (i.e. an approach based upon the degree to which a given specific case

approximated its disease's prototype), a test case that more closely approximated its disease's prototype would more likely be correctly diagnosed than a test case that less closely approximated its disease's prototype. The creation of two distinct typicality gradients would produce two distinct performance predictions. An analysis of the subjects' observed performance would likely be more congruent with one prediction or the other. The theory that best accounts for the subjects' observed performance would therefore be better positioned to argue that the form (structure) of its theoretical construct was preferentially being used to drive the observed performance.

Training protocol

All training and testing activities were delivered over the WEB via each subjects' own PC. During their 3 h session, all students were given 1.5 h to study the characteristic (prototypical) S/S associated with each of the nine differentials along with an opportunity to use their knowledge of these rudimentary prototypes as the basis for developing and applying their evolving diagnostic capabilities against 36 distinct and randomly ordered practice cases. Each of these 36 practice cases were presented in the form of two brief paragraphs with one describing the case's historical findings and the other describing the case vignette's pertinent physical findings (see Table 1; an example case vignette). Again, four different case portrayals were used as practice cases for each of the nine differentials. Note that from a Prototype theory's perspective, each of the four practice cases used to represent a given disease had a different typicality measure (ranging from more thru less typical), while from an Exemplar theory's perspective, all four practice cases had the same typicality measure.

All 36 practice cases were presented as multiple choice questions using a 'please select your best diagnosis from among the nine possible disease etiologies' format. Immediately following their diagnosis for a given case, subjects received feedback that differed based upon whether they selected a correct or incorrect diagnosis. If correct, they were immediately given the opportunity to review both the case and those S/S in the case that were also listed as characteristic of that disease's prototype. If incorrect, they were immediately given

Table 1: Example case vignette.

HX: A 46-year-old female presents with a chief complaint of chest pain described as gradual in onset. The patient relates that it occurred at the supermarket and has been ongoing for the past 3 h. The pain is described as stabbing. The patient identifies the pain as right sided and laterally located. There was no radiation. Associated findings: cough and flu symptoms for the past 3 days, gradual difficulty breathing, and a temperature of 100.3. Neither exertion nor food appeared to alter the chest discomfort. The patient has a history of prolonged bed rest.

PE: HEENT findings are unremarkable. RESPIRATORY findings: tachypnea at 27 breaths per minute, coarse wheezing, fine basilar crackles, coarse ronchi, a productive cough with yellow sputum, and unilateral diminished breath sounds. CARDIOVASCULAR findings: tachycardia at 128. ABDOMINAL findings are unremarkable. EXTREMITY findings are unremarkable. MUSCULOSKELETAL findings: unremarkable. DERMAL findings are unremarkable.

the opportunity to review both the case and those S/S in the case that were also listed as characteristic of the prototype representing the disease that should have been selected as their diagnosis. These practice cases and feedback could serve as learning opportunities in three ways. First, instantiation of each practice case as an exemplar of its disease. Second, prototype enhancement via the opportunity for students to note which S/S occurred more or less frequently across the four practice cases used to represent a given disease. Third, further prototype enhancement via an opportunity to note whether certain S/S enabled one disease prototype to be more easily distinguished from a competing disease prototype.

Testing protocol

Immediately prior to, and following their 1.5 h training exercise, all students were given a 45 min pre-training test and a 45 min post-training test containing the same 27, randomly ordered test case vignettes (using the same format as the previously described training cases). These 27 test cases consisted of three different case portrayals of each disease. Again, while Exemplar theory assigned the same typicality measure to each, Prototype theory assigned a different typicality measure to each case (least, mid and most typical) for each of the nine acute chest pain differentials). All 27 test cases were presented as multiple choice questions using a 'please select your best diagnosis from among the nine possible disease etiologies' format. Student performance was determined by assigning a 1 for a correct diagnosis and 0 for an incorrect diagnosis. Evidence of a pre- vs. post-training improvement in diagnostic performance was determined by calculating the percentage of correct diagnoses out of 27 test items for each student. Evidence of a typicality/performance gradient (as predicted by Prototype theory) was determined by assigning the nine least typical case portrayals for each disease to a set termed 'least typical', the nine mid typical cases to a set termed 'mid typical' and the nine most typical cases to a set termed 'most typical'. Subject performance against each of the three sets of cases (least, mid and most typical) was calculated in terms of the percent of cases in each set correctly diagnosed.

Hypotheses

#1: Comparison of pre- vs. post-training diagnostic performance metrics will demonstrate a significant improvement in the subjects' diagnostic capabilities. #2: In accordance with Prototype theory, the subjects' diagnostic performance will increase as the typicality of the three case sets (least, mid and most typical) increases. Evidence of little or no performance improvement as the typicality of the three test case increased would support the inference that the subjects' diagnostic performance was primarily driven by their use of disease exemplars.

Results

Hypothesis #1: Overall, 117 students achieved a significant improvement in diagnostic performance following a 1.5 h

diagnostic training exercise consisting of both disease prototypes and exemplars (41.70% of cases correctly diagnosed at pre-training to 60.26% post-training); $t=14.04$, $p<0.001$. In terms of effect size, this improvement was highly significant; Cohen's $d=1.32$.

Hypothesis #2: During the pre-training test phase, the percent of cases correctly diagnosed for least typical, mid, and most typical cases was 31.7%, 41.9%, 51.5%, respectively. Paired t-testing demonstrated performance against most typical>mid: $t=4.94$, $p<0.001$; and mid typical>least: $t=5.16$, $p<0.001$). During the post-training test phase, the percent correctly diagnosed for least typical, mid and most typical cases was 52.3%, 57.7% and 69.8%, respectively. Paired t-test performance again demonstrated significant differences with most typical>mid: $t=2.94$, $p<0.01$; and mid typical>least: $t=6.64$, $p<0.001$. Thus, evidence of a prototype-derived, typicality/performance gradient was seen in both the pre- and post-training phase of the study.

Discussion

This investigation's first objective was to assess the utility of a DPT/System I theoretical framework as a means of improving the diagnostic capabilities of early year one medical students confronted with ill-defined diseases. The findings demonstrated that a 1.5 h training activity involving both disease prototypes and multiple exemplars led to a highly significant effect size performance improvement in their DDX capabilities (Cohen's $d=1.32$). Contemporary medical education has yet to create a codified, evidence-based approach to training to or assessing the diagnostic capabilities of medical students, residents or practitioners. These findings, in conjunction with those emerging from other DPT-oriented research [4, 5, 19, 23] suggest that a DPT framework could contribute to the development of a codified approach to diagnostic training and assessment.

It is extremely difficult for researchers to draw direct inferences regarding the cognitive factors enabling the development and improvement of diagnostic capabilities. These findings suggest that a DPT-based research framework may provide an opportunity to gain meaningful, albeit indirect insights into the roles its theorized constructs and processes play in the development of diagnostic competence. Specifically, Exemplar theories would predict little to no evidence of a typicality/performance gradient among students with very little prior experience, while Prototype theories would predict evidence of such a gradient. Our findings demonstrated a typicality/

performance gradient in the pre-training phase of the study and thereby provided indirect evidence that the construct driving their performance was likely in the form of a rudimentary prototype for each of the diseases studied. Our findings also demonstrated a typicality/performance gradient in the post-training phase of the study thereby providing additional, although once again, indirect evidence that the construct driving their later performance was also in the form of disease prototypes.

Prototype theories postulate that the formation of increasingly robust prototypes begins with the instantiation that a given ill-defined concept has an associated set of characteristic findings. From the perspective of the task of differential diagnosis, the initial formation of a medical student's disease prototype begins with the belief, understanding and instantiation that the disease at hand is associated with a set of characteristic S/S. Prototype theories further postulate that exemplars serve as the basis for estimating the frequency with which the disease's characteristic S/S occur in a larger population of cases representing the disease at hand. These accumulating case exemplars subsequently enable the transformation of a rudimentary prototype consisting of a list of non-differentiated, albeit characteristic S/S into a frequency-based listing of the disease's characteristic S/S. This frequency based listing of characteristic S/S would enable learners to discern which of these more and less frequently observed S/S might be used to further differentiate the prototype at hand from competing prototypes.

Given our account of how a DPT/System I framework might be applied in the context of DDX, the findings of this study, and the notion that the construction of robust, frequency-based disease prototypes requires exposure to multiple and varied case exemplars, we urge the reader to neither perceive this study as arguing that prototype training is singularly important, nor discrediting the importance of exemplars. Rather, we suggest that the initial instantiation of a rudimentary prototype and its subsequent enrichment via multiple and varied case exemplars are both required to construct increasingly robust, frequency-based disease prototypes capable of expediting the development of diagnostic capabilities in early medical students.

Further, existence of a typicality/performance gradient during the pre-training phase of the study causes the authors to suggest that prior to matriculation to medical school, students may already have a tendency to utilize constructs in the form of prototypes as the basis for performing categorization tasks. Evidence of a post-training typicality/performance gradient causes the authors to more firmly suggest that the observed performance improvements were likely to also be prototype driven. If

correct, research into how the System I knowledge base structuring and information processing mechanisms of pre-clinical students might be trained to optimize the transformation of case exemplars into a more robust, frequency-based prototype for each disease could expedite the development of diagnostic competencies.

Finally we suggest that a DPT-based DDX research framework could also make it possible to provide evidence as to whether such an approach can also enhance the diagnostic capabilities of students in clinical training, residents and clinicians. We therefore recommend that contemporary medical educators take the time to familiarize themselves with DPT and in doing so, consider utilizing elements of this framework to explore new approaches to training and testing students, residents, fellows and practitioners.

While these findings and those of other DPT oriented investigations suggest that this theoretical framework may be capable of making meaningful and enduring contributions to medical education, DPT has significant limitations. Perhaps principal among them is the question of how to best deal with and represent constructs in the form of Exemplars and Prototypes. For example, we could only assume that the subjects in this study had few if any exemplars stored in Episodic memory prior to their pre-test, and furthermore, that any previously instantiated exemplars were unlikely to be 'exactly like' those that would be utilized as practice or test cases. These assumptions cannot be objectively quantified. Similarly, we could only assume that the subjects were using their encounters with multiple training exemplars as the basis for transforming their disease prototypes (initially instantiated only in terms of their characteristic S/S), into frequency-based prototypes. Once again, while this is the theoretical position of Prototype theorists, we cannot yet objectively validate the existence of these alleged knowledge structures. Nonetheless, there is little doubt that in the years and decades to follow, a new generation of neuro-cognitive researchers will create and use advanced, neural network based models of cognitive structures and processes, and unimaginably advanced fMRIs, to ground future 'models of mind and education' upon more objectively quantifiable models of neural structures and processes.

Conclusions

Given that most human illnesses have ill-defined bedside diagnostic criteria, and therefore can present with tens to hundreds of distinct and yet equally valid case presentations, training to and assessing the diagnostic capabilities

of medical students would appear to be a daunting task. Further, mounting evidence demonstrating that diagnostic error results in a significant number of unnecessary deaths annually confirms the urgency with which all involved in medical education and licensure pursue the development of a codified, evidence based approach to diagnostic training and assessment [24–26].

This study assessed the utility of DPT as framework for both training early medical students in DDX, and, as a means of investigating the cognitive factors enabling improvements in DDX capabilities. Our findings suggest that a DPT instructional framework can serve as an efficient and effective means of improving the DDX capabilities of very early medical students. The findings also suggest that constructs in the form of prototypes were likely preferentially utilized to diagnose test cases during pre- and post-training. While these later findings represent potentially important insights, more conclusive evidence that the formation of robust, frequency-based disease prototypes played a significant role in improving the diagnostic capabilities of early medical students is needed. In the meantime, we suggest that instructional designers and educators consider the use of both disease prototypes and multiple case exemplars as a basis for supporting students in the development of diagnostic competencies.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

- Papa F. Learning sciences principles that can inform the construction of new approaches to diagnostic training. *Diagnosis* 2014;1:125–9.
- Ark TK, Brooks LR, Eva KW. Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Acad Med* 2006;81:405–9.
- Ark TK, Brooks LR, Eva KW. The benefits of flexibility: The pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Med Educ* 2007;41:281–7.
- Eva KW, Hatala RM, Leblanc VR, Brooks LR. Teaching from the clinical reasoning literature: Combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ* 2007;41:1152–8.
- Monteiro SM, Norman G. Diagnostic reasoning: where we've been, where we're going. *Teach Learn Med* 2013;25(Suppl 1): S26–32.
- Sladek RM, Phillips PA, Bond MJ. Implementation science: a role for parallel dual processing models of reasoning? *Implementation Science* 2006;1:12.
- Stanovich KE. *Who is rational? studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum, 1999.
- Evans JS. Dual processing accounts of reasoning, judgment and social cognition. *Annu Rev Psychol* 2008;59:255–78.
- Sloman SA. The empirical case for two systems of reasoning. *Psychological Bulletin* 1996;112:3–22.
- Nosofsky RM. Attention and learning processes in the identification and categorization of integral stimuli. *J Exp Psychol: Learn* 1987;13:87–108.
- Medin DL, Schaffer MM. Context theory of classification learning. *Psychol Rev* 1978;85:207–38.
- Homa D, Sterling S, Trepel L. Limitations of exemplar based generalization and the abstraction of categorical information. *J Exp Psychol: Hum L* 1981;7:418–39.
- Smith JD, Minda JP. Distinguishing prototype-based and exemplar-based processes in category learning. *J Exp Psychol: Learn* 2002;28:800–11.
- Voorspoels W, Vanpaemel W, Storms G. Modeling typicality: extending the prototype view. In: Love BC, McRae K, Sloutsky VM, editors. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 2008:757–62.
- Papa FJ, Shores JH, Meyer S. Effects of pattern matching, pattern discrimination, and experience in the development of diagnostic expertise. *Acad Med* 1990;65:S21–22.
- Norman G. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;39:418–27.
- Mayes AR, Montaldi D. Exploring the neural bases of episodic and semantic memory: the role of structural and functional neuroimaging. *Neurosci Biobehav Rev* 2001;25:555–73.
- Papa FJ, Elieson W. Diagnostic accuracy as a function of case prototypicality. *Acad Med* 1993;69:S58–60.
- Papa FJ, Stone RC, Aldrich DG. Further evidence of the relationship between case typicality and diagnostic performance: implications for medical education. *Acad Med* 1996;71:S10–12.
- Osherson D, Smith EE. On typicality and vagueness. *Cognition* 1997;64:189–206.
- Smith JD. Prototypes, exemplars and the natural history of categorization. *Psychon Bull Rev* 2014;21:312–31.
- Papa FJ, Young JI, Knezek G, Bourdage R. An expert system-based differential diagnostic skills assessment and tutorial tool. *Computer Assisted Learning 1991 (CAL91)*, Lancaster, UK, April, 1991.
- Papa FJ, Oglesby MW, Aldrich DG, Schaller F, Cipher DJ. Improving diagnostic capabilities of medical students via application of cognitive sciences-derived learning principles. *Med Educ* 2007;41:419–25.
- Graber ML, Berner ES. Diagnostic error: is overconfidence the problem? *Am J Med* 2008;121:S2–23.
- Berner ES. Diagnostic error in medicine: introduction. *Adv Health Sci Educ Theory Pract* 2009;14(Suppl 1):1–5.
- Balogh ER, Miller BT, Ball JR, editors. *Committee of diagnostic error in health care. Improving diagnosis in health care*. Washington, DC: The National Academies Press, 2015.