FAMILIARITY OF CONTEXT AND FEATURES IN A MEDICAL DIAGNOSIS TASK

THE ROLE OF FAMILIARITY OF CONTEXT AND FEATURES IN A MEDICAL DIAGNOSIS TASK

By

MEREDITH E. YOUNG, B.Sc. Hons, B. A.

A thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Doctorate of Philosophy

McMaster University

DOCTORATE OF PHILOSOPHY (2009) McMaster University

(Psychology, Neuroscience & Behaviour) Hamilton, Ontario

TITLE: The Role Of Familiarity Of Context And Features In A Medical Diagnosis Task

AUTHOR: Meredith E. Young, B.Sc. Hons (McMaster University),

B. A. (McMaster University)

SUPERVISOR: Dr. Geoffrey R. Norman

NUMBER OF PAGES: ix, 156.

Abstract

Medical diagnosis is a complex task, that requires integrating several sources and types of information: a patient's description of their symptoms, lab results, and perhaps even 'gut feelings' regarding potential diagnoses. From a cognitive psychology perspective, diagnosis is a type of categorization and as such has been typically divided into processes that are deliberate, rule oriented, and available to conscious control (often called analytic processing) and processes that are rapid, outside of conscious awareness, and typically based on similarity (often referred to as non-analytic processing).

Traditionally, similarity has referred to whole-case similarity between a current and previously encountered case. However, this pattern matching to an entire previous case does not differentiate between diagnostic and non-diagnostic information, a distinction that is made clear in the rules taught to medical professionals. Since medicine does rely extensively on diagnostic rules, the research presented in this thesis will examine the effect of similarity of features relevant to the application of a diagnostic rule as well as the effect of similarity from patient identity, which is mnemonically salient but diagnostically irrelevant.

The work presented in this thesis specifically examines the role of similarity in the categorization, or diagnostic decisions of novices. Medical students start training with the best available rules, standard diagnostic rules likely to be used in future practice. In the experiments reported in this thesis, participants are trained to competence on diagnostic rules using prototypical written case vignettes of simplified psychiatric diagnoses. Participants then evaluate cases in which clinical information supports two

possible diagnoses, but in which either diagnostic features or diagnostically irrelevant identity information is similar to those seen in training. The results of these experiments indicate a strong reliance on familiar of rule-relevant symptom descriptions (i.e. similarity within the application of a diagnostic rule), supporting an adaptive role of similarity within the application of an analytical decision rule. Further, the influence of familiar diagnostically irrelevant information (i.e. similarity within the context of patient identity) demonstrates the maintenance of non-diagnostic information within memory, and the possibility of matching to a previous exemplar on rule-irrelevant features. Familiarity, whether diagnostically relevant or not, increases the probability that clinically relevant features are mentioned in support of a diagnosis, which may indicate the disambiguation of features following previous experience with that feature, and a strong influence of familiar but non-diagnostic information on the interpretation of features.

This thesis supports a model of medical decision making in which there is an effect of similarity to previous instantiations of clinically relevant features. That is, similarity is a basic component of decision making that is not limited to matching on entire previous instances. Previous research has suggested that analytic and non-analytic reasoning are competing or fundamentally separate processes, whereas the demonstration of similarity within the application of a diagnostic rule suggests that not only is similarity an adaptive strategy for learners, but the differentiation between similarity based and rule based processes may be less clear than previously suggested.

Preface

This dissertation follows a stapled thesis format. I, Meredith Young, am the first author for all of the journal papers presented within. The ideas and work, including the design, data collection, data analysis, and manuscript preparation, associated with each paper are primarily my own. Drs. Geoffrey R. Norman and Lee R. Brooks were co-authors. In their role as my supervisors, they provided guidance, critical feedback, and suggestions in regards to experimental design, data analysis, and write up of the papers.

Acknowledgements

How does one sum up the amount of help, inspiration, mentorship and friendship that has helped me get through the last few years? It is hard not to sound too sappy, but what follows are honest expressions of gratitude to those who have lead me down this path, and have helped make this 5-year process a success.

To Geoff, I owe my appreciation. You were willing to take a chance on someone who was excited about research questions but lacked direction, and allowed me freedom to explore my ideas. You are responsible for the existence of this document, and the soon-to-be-doctor that authored it. You demonstrate incredible passion for research, and are always willing to chat about ideas, random thoughts, ungrounded speculation, and rough spots. Plus you normally pay for the beer.

To Lee, I owe my gratitude and appreciation for mind-blowing conversations. Although most under his supervision are occasionally challenged with the task of understanding what he is saying, I cherish the moments when it all 'clicks', even if it is occasionally weeks, months, or years later. Lee, you are a great mentor, and have taught me to slow down, think deeply, never take anything for granted, and chase great ideas. You have been a patient and dedicated member of my 'team', and I thank you.

To Kevin, who always pushes me to do better, know more, and stretch farther. I have great respect for you. You are a brilliant researcher, a compassionate listener, a fantastic editor, and gosh darn it, such a nice guy. I would be so lucky as to be considered roughly in the same league as someone like you. You have been a great support, a great

reality check, and someone who always pushes me for more. Best of luck, wherever life takes you.

To Karin, my formal supervisor and amazing mentor. You have been a great help, and have provided me with some truly amazing experiences. Your door is always open, your mind always critical, and you have really become a friend. You've helped me grow as a researcher, always pushed me to think big, and have mentored me through my thesis, radio shows, tv shows, and far too many embarrassingly bad blogs. You are a great researcher, and manage to inspire and excite everyone around you about science, and allow your students the freedom to grow on their own. I don't feel like 'thanks' is enough, but it will have to do for now.

To Scott, the best not-member-of-my-committee that I could ask for. You've taught me that gin and tonic tastes best in pint-size, that PRPs can be cool, and that open minds really do make science go 'round. You've been an amazing help, and a great friend. I will miss you, your lovely wife, your cute kid, and your open door terribly. But, I'll be back. We do after all have that project to finish.

Every single person who leaves the Department of Psychology, Neuroscience & Behaviour (before and after the name change), owes a substantial part of their degree to the most helpful folks on the planet: Nancy, Sally, Wendy, Ann and Gary. You are all amazing at what you do, and thanks for all of your help (and Nancy, thanks for the gossip!).

I would like to thank all of the graduate students who have helped me through my studies – you've kept me sane, grounded, and well tanned by keeping me company at the

Phoenix. Special thanks to those who have been with me since the start: Jim, Ayesha, Paul, and Ferrinne. Thanks especially to Ferrinne who shared my life in grad school, re-taught me the benefit of play, and taking the time to smell lilacs. Thanks also to all of those who know more about me than they probably care to share. Thanks to Paul Brunet, Danny Krupp, Heather Poole, Maria D'Angelo, Rob Berger (who taught me the secrets to truly enjoying a solid high-five), Phillow, Andrea Currie (who taught me to see another perspective), Melissa Deller (who taught me you really can do anything you set your mind to), Sophie Fanourgiakis, Nikki Woods, Ferrinne Spector, Jonathan Oler, and everyone else for your support, walks, hugs, and frequent laughter.

How do I go about thanking the person who has held me up, and helped me grow. Stefan, there was no way that this was possible without you. From your ability to make me smile no matter what, to your quiet and heartfelt abilities to make me feel strong from your ability to listen, to your ability to critique my experimental design. You have helped me become something that I never thought possible. I love you, and thank you.

My success can also be attributed to my family. My parents, Marie and Dennis, instilled a love of learning, and taught me the value of not only hard work, but of loving what you do. They taught me high standards, strong morals, and a drive to keep learning. You have supported me in so many ways, and I have no capacity to thank you properly. Mom, you were more insistent that my committee regarding the progress of my thesis, and Dad, your check-ins on my sanity were always brilliantly timed. I love you both so much, and hope I've made you proud.

Nathan and Elaine, who knew we would take this path. It has been remarkable sharing it with you, and know that your brilliance and hard work will take you both to the tops of your fields, wherever you decide to land. I love the both of you, and wish the best.

I wish I had witty words of inspiration to close, but I don't. I simply want to say that the last few years have been brilliant, and I have truly enjoyed my time in the department. In closing, thanks to you all, and I'm off!

Table of Contents

List of Tables

List of Figures

Chapter 1

General Introduction

General Introduction

Medicine is a domain in which errors are high cost, influencing everything from quality of life of a patient to probability of survival. It is an area in which understanding how decisions are made is of particular value – knowing how the decision process proceeds allow for insight into how it might possibility be improved. Drawing on strong predictive models from cognitive psychology provides a systematic way to investigate how complex decisions in medicine are made, how formal knowledge (such as medical rules) and intuition can interplay, and how one learns to integrate individual experience with the use and application of diagnostic rules. Clinical reasoning and concept learning have had a long and interconnected past, with early researchers attempting to identify and catalogue general heuristics and strategies used in the process of clinical diagnosis (e.g. Elstein, Shulman & Sprafka, 1978). A clear benefit exists in combining the study of concept learning and categorization with medical decision making: medicine is a series of complex concepts and categories learned in adulthood, which are rich in variability (no two patients are the same, nor does their illness present in an identical manner), full of specialized vocabulary, and exists within complex causal and explanatory biological systems.

It is now generally accepted that clinicians use multiple strategies and types of knowledge to arrive at a diagnosis. Various authors have proposed that diagnostic reasoning is based on defining features, (reviewed in Medin, 1989; Regehr & Norman, 1996), regression approaches to weighting clinical features (Freidman, Elstein, Wolf, Murphy, Franz, Heckerling *et al*, 1999; Wigton, 1988), Bayesian models of probabilities

(Weinstein, Fineberg, Elstein, Frazier, Neuhauser, Neutra & McNeill, 1980), illness scripts (Feltovitch & Barrows, 1984) and encapsulated knowledge (Schmidt & Rikers, 2007 for a review). These models all include the assumption of consciously accessible knowledge structures accessed for and by analytical thinking processes. A major distinction is made between analytical strategies that apply various kinds of formal knowledge and formal judgment rules, and non-analytic strategies based on an unconscious match between the current problem and previous cases or "exemplars" retrieved from memory (proposed by Brooks, 1978; Brooks, 1990; reviewed in Norman, Young & Brooks, 2007). Research in both cognitive psychology (e.g. Homa, Sterling & Trepel, 1981; Brooks & Hannah, 2006; Hannah & Brooks, 2006) and clinical reasoning (Brooks, Norman & Allen, 1991; Hatala, Norman & Brooks, 1999) have demonstrated a role for similarity to exemplars derived from specific prior experiences.

An instance model proposes that there is a rapid similarity match to a previously encountered case, and this previously seen case aids in current decision making. This has typically been investigated using similarity to previously seen whole cases, and this matching to whole prior instances leaves little room for the coordination of rule-based knowledge and previous experience, an issue exemplified in medicine. As a student learns the diagnostic rules for a given diagnosis, they must learn to identify the variability with which a diagnostic feature can present. In the case of a novice, it would make intuitive sense to rely on specific manifestations of features that have been encountered previously – it would be an adaptive use of previous experience with particular features within the diagnostic rule. The research presented in this thesis will attempt to address the

role of similarity both within the context of the application of a diagnostic rule, and within the influence of non-diagnostic information.

Themes and goals of the present research:

The influence of specific previous experience on current decision making is an overarching theme of my graduate research program and this doctoral thesis. In a series of four studies in three empirical chapters, I investigated the role of specific previous experience within the application of an authoritative diagnostic rule and demonstrated the reliance on familiar non-diagnostic information. I used psychiatry as a domain of study in order to investigate the influence of familiarity along these disparate dimensions, as clear distinctions are made between clinically relevant (rule based) and clinically irrelevant (patient identity information) information. Based on the overarching theme of familiarity, I address three main goals: 1) to examine medicine as an example of complex decision making, and the benefits of theories of categorization, 2) to identify the difficulties and benefits of using psychiatry to develop multidimensional stimuli to study medical decision making, 3) the exploration of the differentiation between similarity to whole prior instances and similarity to individual manifestations of features.

1. Medicine as categorization

Categorization is the grouping of perceptually different stimuli into a single category, or group. Objects can be categorized into naturalistic groups (e.g. individual examples and species all equally represent 'dog'), artificial groups (e.g. types of cars), goal-driven groups (e.g. things I will regret tomorrow). Some of the functions of categorization

include the ability to communicate using category labels (e.g. it is easier to refer to 'a dog' rather than the specific animal you have encountered), and allows us to treat perceptually different stimuli as identical (you can treat a new dog in a similar manner as dogs you have previously encountered). The categorization of objects or stimuli is a powerful analogy for medical diagnosis. Patients of varying appearance, with different manifestations of symptoms, are ultimately categorized with a particular diagnosis, and are typically prescribed treatments congruent with their diagnostic category.

Assuming that diagnosis is similar to a judgment of category membership, theories from cognitive psychology can be applied to understanding the complex processes that underlie diagnostic decision making. The research presented in this thesis uses medicine as a domain of study to investigate the influence of familiarity in decision making. Consequently the research presented here is equally informative to research into the mechanisms of categorization, and the learning and application of medical rules. With this intersecting research program in mind, the studies presented here will attempt to speak both to those interested in medical decision making specifically, and cognitive mechanisms of categorization more generally. Using theories proposed to explain categorization behaviour, I have enhanced our understanding of the role of familiarity in a variance-rich, and complicated decision task.

2. *Psychiatry as a domain of study*

The practice of medicine, and what skills are considered most valuable, varies greatly between specialty. Previous demonstrations of the role of similarity to a whole previous

case have used highly visual domains of medicine, such as dermatology (Brooks, Norman, & Allen, 1991; Kulatunga-Moruzi, Brooks, & Norman, 2001). Additionally, the recognition of visual features frequently proves difficult, even for experts (Brooks, LeBlanc, & Norman, 2000). The use of visually dominated domains of medicine provides perceptual variation in symptom presentation, and features clearly occur within a context (either the context of a patient (Brooks, LeBlanc, & Norman, 2000), or within the context of physical location of the feature (Kulatunga-Moruzi, Brooks, & Norman, 2001)). However, there are benefits to be gained from extending previous research on the influence of familiarity to written case vignettes.

Clinical features can be represented on multiple levels. Students are taught clinical signs and symptoms using specific medical language; for example students are taught that hallucinations are a characteristic feature of schizophrenia. This is considered the informational level of a clinical feature, or language of the medical rule, and having a category of behaviours that can be represented as 'hallucinations' facilitate communication and diagnosis. However, each clinical feature presents in an idiosyncratic way across patients – the specific hallucination experienced by one patient is unlikely to be identical in another. The individual manifestations of the features mentioned in the clinical rule are considered to be instantiations of the medical rule, or instantiated features. By representing both informational and instantiated features within written case vignettes, we eliminate the distinction between informational levels as verbal (features that can be identified on verbal report) and instantiated features as perceptual (the

physical appearance of a feature in a photo) – a distinction that might be assumed from research in highly visual domains such as dermatology.

The ability to manipulate informational and instantiated features within a written case vignette also allows for clear distinctions and delineations between features. As each clinically relevant feature is presented as a written description, features can be easily parsed and recombined – allowing for recombination without the cost of decreasing the validity of the patient case. Further, the ability to clearly delineate between features also allows a clear experimental advantage in the addition and manipulation of clinically relevant and irrelevant information, a distinction that is important to the research presented in this thesis.

The use of written case vignettes does have benefits in examining the role of specific previous experience in clinical decision making. For the purposes of this thesis, the role of specific previous experience, both clinically relevant and irrelevant, has been investigated in the domain of psychiatry. Psychiatry is an area of medicine in which the features are very ambiguous, highly varying, and often lack confirming medical tests. It is also not uncommon to have a patient with more than one diagnosis, which allows for the maintenance of validity of using mixed case designs, and which is a central experimental method in this thesis. Because of the variability and ambiguity, psychiatry is often reported to be quite difficult for learners. Finally, psychiatry is unique in the presence of additive rules, as dictated by the Diagnostic and Statistical Manual for Psychiatric Disorders IV (DSM-IV: American Psychiatric Association, 2000). The DSM-IV advocates that all clinical features identified should be weighed equally, and that a

minimum number of features must be present for a diagnosis to be made. This presence of rules that are authoritative, advocate equal weighting, and provide a minimum number of features to be identified provides a grounding for learners, but allows a unique opportunity to investigate the role of similarity within the context of strong diagnostic rules.

3. Similarity to whole instances and instantiated features

Previous research in cognitive psychology has provided strong evidence of an exemplar, or instance based model of categorization (as suggested by Brooks, 1978; Brooks, 1990, and supported by Estes, 1976; Hintzman, 1986; Medin, 1989; Medin & Schaffer, 1978; Nosofsky, 1989), and evidence has accrued for an instance based model of clinical reasoning (Brooks, Norman, & Allen, 1991; Hatala, Norman & Brooks, 1999; Medin, Altom, Edelson & Freko, 1982; Young, Brooks, & Norman, 2007). As previously mentioned, an instance model proposes that when faced with a novel object (or patient) a rapid similarity-based match occurs to a similar previously seen case. If the two exemplars are sufficiently similar, they are categorized (or diagnosed) as being members of the same category (or diagnostic category). While similarity to a whole previous case can be functionally adaptive, little distinction is made within this model between what is mentioned in the clinical rule, and is therefore rule-relevant, and that which is not mentioned in the clinical rule, and is therefore rule-irrelevant.

Differential reliance on familiar manifestations of clinically relevant features is a demonstration of the role of similarity within the application of a diagnostic rule. The role of familiar instantiated features has been demonstrated within artificial animals

(Brooks & Hannah, 2006), and synonymous medical terminology (Dore, Weaver, & Norman, 2005) but this program of research is the first to demonstrate the adaptive use of similarity in the application of a diagnostic rule in novices. Through research presented in this thesis, a distinction is made between the influence of familiar clinically relevant and irrelevant information, and the use of similarity in the diagnostic decisions of novices. Further, the research presented within this thesis proposes that a clear distinction between rules and similarity, as distinct and separable processing strategies, may not be complete.

Overview and research questions

Literature review

As models of categorization function as the theoretical basis for research presented in this thesis, I will first offer a brief review and discussion of major models of categorization, both within the cognitive psychology and clinical reasoning literatures. This literature review provides the foundation of this dissertation by discussing traditional models of categorization, and our current understanding of decision making and clinical reasoning.

Can familiarity influence reliance on features?

In Chapter 3, I present the first empirical demonstration from my program of research. This study demonstrates the reliance on familiar feature instantiations in the diagnostic decisions of novices. This study was published in *Medical Education* (Young, Brooks, & Norman, 2007), and demonstrates an increased weighting of familiar feature instantiations compared to equally valid, but novel feature descriptions. This heavy

weighting of familiar clinical features is maintained following a 24-hour delay. Further, the influence of familiar symptom descriptions remains consistent when considering a subpopulation of participants who could consistently and accurately identify the clinically relevant features present in the written case vignette. This suggests that the reliance on familiar instantiated features cannot be explained by an inability to recognize novel manifestations of features, rather that participants appear to be assigning more diagnostic weight to the features that are expressed in a familiar manner.

This initial demonstration of the reliance on familiar feature instantiations is not without limitations. Feature instantiations were not completely systematized, therefore it may be possible that aspects of the unique feature descriptions were actually aiding in their translation, and so all features were not necessarily undergoing a re-descriptive process into the medical language of the rule.

Can familiarity influence diagnostic decisions in the context of a rule?

In Chapter 4, I present two empirical studies investigating the influence of indirectly cuing participants to use authoritative diagnostic rules on the overweighting of familiar clinical features. The familiar feature instantiations used in Chapter 4 were constructed to reflect a patient voice, and remove any terms that might be considered synonymous with the diagnostic rules. The use of symptoms that reflect a patient voice (closely mimicking how a patient would describe their own symptoms) can not only increase the difficulty of translation from the instantiated to informational level, but also rule out the possibility that differential difficulty of translation may have influenced the results reported in Chapter 3. Study 1 in Chapter 4 clearly replicates the findings of

Chapter 3, demonstrating a reliance on familiar feature instantiations, even when feature instantiations were carefully controlled.

Further, in Experiment 1 participants reported clinically relevant features in one text field, with no minimum or maximum criteria for number of features being set. In Experiment 2, participants reported clinically relevant features in four individual text fields, with the requirement that at least three clinically relevant features, not necessarily supporting the same diagnosis, had to be identified in order to make a diagnosis. This indirect cue resulted in more participants reporting 'counting' the features present in a case as a diagnostic strategy, and participants still assigned more diagnostic weight to familiar symptom descriptions than to equally valid novel descriptions. Further, results presented in this chapter indicate that the role of similarity within the application of a diagnostic rule is not transient, and familiarity results in the increased probability of a feature being reported as clinically relevant.

Can familiar non-diagnostic information influence decision making?

Distinct from the studies presented thus far in the thesis, the experiment presented Chapter 5 suggests that non-diagnostic information can influence diagnostic decision making in novices. Further, the influence of familiar patient identity information is not transient, and may increase across increased delay periods (up to 1 week). Finally, the research presented in this chapter suggests that familiar non-diagnostic information can influence the probability that a clinically relevant feature is mentioned in support of the diagnostic decision made. Patient identity information was carefully constructed to be

clinically irrelevant, and so research presented in this chapter suggests a match between a current and previous case can be made based on irrelevant stimulus dimensions.

General Discussion

In this chapter (Chapter 6) I present an integrative synthesis of the research presented in these three empirical chapters. Here, I also present a discussion of implications, limitations, and future directions of this work.

Chapter 2

Review of the literature

Review of the literature

This thesis used theories and models drawn from cognitive psychology to understand decision making in a medical diagnostic task. Theoretical grounding in categorization is a foundation of research presented in this dissertation, as such, this literature review focuses on differential theories of both categorization and clinical reasoning.

Imagine yourself as a physician in a family medicine clinic. You have a very diverse practice including members of different cultural groups, differing education levels and ranging from newborns to seniors. During each patient encounter, your patient tries to explain their reason for the visit using their own vocabulary ('I'm feeling a little off', 'I've had a headache I just can't shake', 'The baby just hasn't been herself recently, she's been fussy'), you re-describe their statements into signs and symptoms found in medical vocabulary, seek more information to clarify and narrow a potential diagnosis, decide what clinical tests need to be run (if any), and make recommendations for treatment. Once complete, you repeat this range of tasks six to eight times an hour, for a total of 30-45 patient cases a day, five days a week. Medicine is difficult - it requires large amounts of specialized training, including formal knowledge and practice experience, the ability to rapidly narrow down possible diagnoses, and the ability to recognize clinically relevant features when patients present them in their own idiosyncratic ways of communicating. However, all of these components of medical decision making are done using the same memory, cognitive, and decision strategies and

structures used in everyday decision tasks, and so parallels should be found when comparing medical decision making to everyday decisions studied using a cognitive psychology approach.

A learner in medicine is faced with particular challenges: learning a new lexicon filled with complex medical language, a new category structure including causes and biomedical relationships, and new associations between items they already know – learning just how many diseases can present with a headache, a sore stomach, or shortness of breath. This learning of language, rules and relationships takes years, and functions to focus the attention of the learner on aspects of a case that may be important. However, the best available rule does not always help to recognize the variability in symptom presentation; a skin lesion can look different on patients of different ages, races, and across stages of progression. It is possible then, that formal medical knowledge, as learned in medical school, performs best when combined with ample experience. A good rule supplemented with many examples in memory may lead to the best possible performance, and may be what aids learning of diagnostic categories.

For the remainder of this thesis, medical diagnosis will be discussed as a categorization task – the grouping of different patients, with slightly differing disease presentation, into a diagnostic group or category. This parallel between diagnosis and categorization will allow for a more explicit comparison of clinical reasoning and diagnosis to a theoretical background in categorization and cognition, which will be reviewed in this chapter. Within such a comparative framework, clinical reasoning and categorization will be explored from the perspective of several categorization theories:

the use of rules, the use of prototypes, the use of examples, the combined use of rules and examples, the role of theory, and dual process models of reasoning. While the theories of categorization are discussed, they will be discussed in light of properties and characteristics of medicine and clinical reasoning.

Knowledge and reasoning with rules

Early psychological accounts of categorization took a definitional approach – that we are able to identify objects in the world based on defining features. For example, one can define a bachelor as an unmarried male, and a triangle as a closed geometric object made up of three straight sides. The definitional approach posits that we categorize objects in the world by using features that are individually necessary (an object must have this feature to be considered a member of the category) and collectively sufficient (if an object has the defining features, it is therefore a member of the category) (Barton & Komatsu, 1989; Bruner, Goodnow, & Austin, 1956; Katz, 1972; Katz & Fodor, 1963; Komatsu, 1992; Medin, 1989; Murphy, 2002; Pothos & Hahn, 2000). This idea has been historically associated with Aristotle's view of an item's 'essence' or core nature (Lakoff, 1987; for a more recent perspective see Ahn, Kalish, Gelman, Medin, Luhmann, Atran, Coley, & Shafto, 2001). Early experimental evidence demonstrated when a definitional feature is present in a complex set of stimuli, individuals do appear to learn the necessary feature for categorization (Hull, 1920).

The proposition of categorization relying on rules that specify necessary and sufficient features, makes three important predictions. First, it implies that there is a

difference in the value of properties of an object that are considered to be definitional, and features that are not (Komatsu, 1992). The rule used to categorize identifies the characteristic features of an object that are valuable for categorization (three sides for a triangle, being male for a bachelor) and does not have any room for 'irrelevant' components (the colour of the triangle, or where you meet the bachelor). In this sense, the definitional features are the loci of attention, and so the definitional properties of an object are either identified as present or not. This leads to the second implication of a definitional approach, that category membership is discrete. An object is either a member of a category or not – four sides on a geometric form make it a square or a rectangle, not a triangle. There is no room within a definitional perspective for partial membership within a category. The third implication of a definitional approach is that all members of a category are equally valid – all triangles are created equal and there is no 'better' bachelor so long as they possess the necessary and sufficient features for membership.

The account of categorization as derived from rules that specify necessary and sufficient features, is one that is quite congruent with our sense of the world, fitting our intuition of order, structure, and definitional properties as characteristic of our decision processes: when undergraduate students are asked if a 'rule' exists that would allow you to determine whether something is a member of a complex category, response rates are very high (88% endorsing yes for table: 64% for dog, 56% for furniture, etc), and students protest the statement that a rule may not exist in naturalistic categories (Brooks, Squire-Graydon, & Wood, 2007). Rules clearly do exist for geometric objects, but even our previous definition of bachelor has exceptions – a young baby boy and the Pope are

all examples that fit our rule for 'bachelor' but few would actually consider an infant or the pope a good example. Additionally, definitional features become even more complex when considering concepts such as 'justice' or 'game' (Wittgenstein, 1957), and experts in their own fields rarely agree on the definitional features of a category (e.g. Mayr, 1982).

The presence of rules is a place where medicine and typical categorization differ: naturalistic categories may have few examples where necessary and sufficient features exist, whereas medicine is constructed to have categories (diseases) that are characterized by very specific features (signs and symptoms). However, there are atypical presentations of a disease – something that does not perfectly reflect the list of features or symptoms characteristic of disease. The existence of such atypical or uncharacteristic examples violates one of the assumptions of the definitional approach – there are some members of a category that are better examples than others, therefore violating the statement that if necessary features are present, all examples are equally valid. This possibility of graded membership (some examples of a category are better than others) led to the development of a prototype hypothesis of categorization (e.g. Rosch & Mervis, 1975), suggesting a stratification of members of a category based on their representativeness (Tversky & Kahneman, 1974).

Knowledge and reasoning with prototypes

Beginning in the 1970's, several lines of research converged to challenge the definitional approach to categorization. When asked to provide the defining features of a

category, participants noted features that were shared by some members of the category, but not all (Rosch & Mervis, 1975), calling into question whether natural categories could be explained through necessary and sufficient features. Additionally, if participants were asked whether particular items or images of examples were members of a category (Are robins birds? Are ostriches birds?), participants were slower at agreeing to the statement, or classifying the item, if asked about less 'typical' items (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1978; Rosch, Simpson & Miller, 1976). Further, the more typical members were produced earliest when individuals were asked to name members of the category (Rosch, Simpson & Miller, 1976; Smith, Balzano & Walker, 1978), and those mentioned first tend to show the highest proportion of feature overlap (Rips, 1975). If all members of a category are equally valid, and all contain necessary and sufficient features, this differential treatment between category items would not be seen.

The presence of better or worse, typical or atypical, members of a category resulted in a shift from a definitional approach to one that acknowledged graded category membership. Members of a category no longer had to share necessary features, and instead could be conceptualized as sharing a family of characteristics – where all members shared some features, but the notions of necessity and sufficiency were abolished (Rosch & Mervis, 1975; for a review see Murphy, 2002; Minda & Smith, 2001). This view, named the family resemblance view, proposed that categories were mentally represented as the average, or central tendency of the family resemblance group – meaning that the category was represented with some form of the 'best' category member. The idea of a 'best' member could represent the single best member of a

category (e.g. robin for the category bird), a mental 'average' of all members (e.g. the average of all encountered birds), or the central tendency of a category (e.g. perhaps the most frequently encountered bird) (Homa, Sterling, & Trepel, 1981; Homa & Vosburgh, 1976; Posner & Keele, 1968; Reed, 1972). A new item that needed to be categorized was compared to this mental representation of central tendency, and if the new item is similar enough it would be considered a member of the category. The introduction of the prototype view suggested that categorization functioned not only on more relaxed category rules, but also on the similarity of an object to a mental representation.

The prototype view has also been investigated in clinical reasoning. If you consider individual diseases as examples of illnesses within organ systems (e.g. gastric ulcers as an example of gastrointestinal disorders) prototypicality effects emerge when students are explicitly asked to rate different diseases that fall in the same organ system (Bordage & Zacks, 1984). Moderate levels of agreement are found between the typicality ratings of individuals, and on recall, participants report the more typical examples sooner than the ones rated as less typical (Bordage & Zacks, 1984; a parallel to Rosch, Simpson, & Miller, 1976; Smith, Balzano, & Walker, 1978). The idea of prototypical members, or prototypical diseases, can also be considered within the framework of medical education. However, it is important to note that prototypicality does not have to be limited to the disease level – although some diseases are perceived to be more typical than others (Bordage & Zacks, 1984), prototypicality effects may manifest both at the disease level and at the level of disease presentation. Medical rules exist because there is some systematicity in how disorders present themselves – perhaps analogous to a family

resemblance structure. Additionally, most medical education materials present typical, or even prototypical, examples of individual diseases to learners. Medical textbooks and encyclopedias show the 'typical' case of mumps (an individual disease) or the 'prototypical' example of a moon-shaped face (an individual feature or symptom), presumably as a way to provide distinct examples of differing categories to aid learning (Avrahami, Kareev, Bogot, Caspi, Dunaevsky, & Lerner, 1997). The prototype view offers clear explanatory power for the differential treatment of members of a given category, however it may not be the only perspective that can explain graded category membership, and some gaps still remain in a full explanation of categorization and medical decision making.

Gaps left by both rules and prototypes

In using the perspective of everyday categorization as a parallel to medicine, it is sometimes easy to oversimplify the task at hand. It is easy to identify a cat, to know which item would be good to pour hot tea into, and to rapidly identify the letters on a keyboard. Because such tasks are easy for us, we assume they are simple. As previously discussed, category membership is not easily captured by rules, and is more complex than it first appears. One explanation for our assumption that categorization is simple is the illusion of the expert (e.g. Carbon, 2008) – once we become fluent at a task, we assume it to be simple, and it is difficult for us to take the perspective of a learner or novice. We do not find it hard to identify a cat, so it must be a simple task. Physicians can read an ECG printout, so it must be straightforward and unambiguous.

In the discussion of categorization so far, it has been implied that features can be easily identified, and are in an object ready to be 'found'. In the definitional approach, features are present or not, and the presence of necessary and sufficient features defines category membership. In the prototype view, an item is compared to a prototype, or a group of family resemblance features, if it is sufficiently similar it is considered to be in the same category. Both the definitional and prototypical approach have so far failed to acknowledge the complexity and variability with which features can present themselves, and does not recognize that the variations in the manifestations of the features can have cue value in themselves (Brooks & Hannah, 2006) and require reorganization, translation, interpretation, and often clarification. This perspective has been supported by research within categorization, which demonstrates that the features identified in an object can be learned flexibly, and even as a consequence of categorization (i.e. features are learned following categorization, rather than features driving categorization: Schyns & Rodet, 1997). The idea that features can be flexibly interpreted, and the identification of features may be a large part of learning seems reasonable – we rarely know what we are looking for when first learning (novices with chest x-rays: Christensen, Murray, Holland, Reynolds, Landay, & Moore, 1981; novices sexing chickens: Biederman & Sciffrar, 1987; novices with dermatological lesions: Norman, Brooks, Coblentz, & Babcock, 1992). Novices cannot always recognize the features that experts are able to point out, and the presence of a single feature indicating another category can greatly influence categorical decisions (Brooks & Hannah, 2006; Hampton, 1995) and such weighting of a single feature is unlikely to be seen in either a definitional or prototype approach.

Research in clinical reasoning has identified flexibility and ambiguity in the identification and interpretation of features (Brooks, LeBlanc, & Norman, 2000; Brooks, Norman, & Allen, 1991; Groves, O'Rourke, & Alexander, 2003; Hatala, Norman, & Brooks, 1999; LeBlanc, Brooks, & Norman, 2002; LeBlanc, Norman, & Brooks, 2001 Zarin & Earls, 1993). Difficulty with detection and identification of a feature has been demonstrated to be mediated by the diagnostic hypothesis currently being considered (LeBlanc, Brooks, & Norman, 2002; LeBlanc, Norman, & Brooks, 2001), and can even vary depending on the presence of non-diagnostic information such as lesion location (Brooks, Norman, & Allen, 1991), or patient identity (Hatala, Norman, & Brooks, 1999). We have yet to discuss a model of categorization that can explain the influence of rule irrelevant information. Additionally, features can vary – lesions can look very different on patients of different ages, ethic groups, and health states, and the models of categorization discussed so far cannot appropriately explain a physician's ability to identify features that vary in their appearance, or how a student learns to identify a feature in its varied presentations. Prototype and definitional models assume that features are easily identified, and present in an unambiguous way. Further, traditional prototype models allow little maintenance of the variability in the individual members of the category to be represented in memory. New items are compared to the category average, and if sufficiently similar, the new item is identified as a member, and assimilated into that average. The same way that a mathematical average does little to conserve the original values in the calculation, a prototype may do very little to preserve the variability present in a naturalistic or diagnostic category.

Knowledge and reasoning with examples

An exemplar theory of categorization proposes that categories are represented as distinct instances, or examples, in memory, in contrast to an abstracted central tendency proposed by the prototype model of categorization (Brooks, 1978; Brooks, 1990; Estes, 1976; Hintzman, 1986; Medin, 1989; Medin & Schaffer, 1978; Nosofsky, 1989, 1988a, 1988b, 1985). When we encounter a new object, a decision about category membership is made not through a comparison to a prototype, bur rather by forming an implicit analogy between the current case and a specific previous case. If the case is sufficiently similar to one previously encountered, the new object is considered a member of the category. It is important to note here that this comparison of previous to current cases is assumed to be an unconscious, non-deliberate process, and therefore is not available for conscious scrutiny (e.g. Medin & Schaffer, 1987; Croskerry, 2000; Elstein & Schwarz, 2002). One clear advantage of an exemplar model is that it allows for the maintenance of variability of category members through the addition of new instances in memory. The maintenance of variability may allow more accurate categorization of atypical category members – they are maintained as whole instances rather than being assimilated into a category average (Brooks, 1990). The exemplar model can equally explain the typicality effects used to support a prototype approach – items that are closer to the prototype are observed more frequently, and so are represented more frequently in memory and are more easily accessible, therefore would be mentioned first in a recall task.

In addition to broadly explaining the possible influence of variability across items, an exemplar approach suggests the encoding of whole instances – irrelevant and relevant information, target features and accompanying context. This encoding of contextual cues can influence categorical decisions, and interact with judgments of typicality. When participants are asked to judge the typicality of a robin and a peacock from an American perspective, a robin is seen as quite typical, however when asked to take a Chinese perspective, peacocks are rated most typical (Barsalou, 1988; Baraslou & Swewell, 1985). This flexibility of typicality cannot be well explained using a prototype perspective, nor can the strong role of contextual information (role of contextual information in language see: Labov, 1973; in data interpretation: Hutchinson & Alba, 1997; in memory: Hockley, 2008; in spatial memory: Chun & Jiang, 2003), the generation of categories on the fly (such as 'things to take out of the house in a fire' or 'birthday presents' as suggested by Barsalou, 1985), or the role of correlated features (Medin, Alton, Edelson, & Frecko, 1982). Additionally, similarity to prior exemplars seems to be a good account of the development of categories in children (e.g. Sloutsky, 2003).

Many demonstrations in support of an exemplar theory of categorization have been conducted using medical materials (Brooks, Norman, & Allen, 1991; Hatala, Norman, & Brooks, 1999; Medin, Altom, Edelson, & Freko, 1982; Young, Brooks, & Norman, 2007). An instance model has traditionally been considered to function on the whole-case level, with global similarity functioning as the metric for comparison to previous instances (Brooks, Norman, & Allen, 1991; Kempainen, Migeon, & Wolf, 2003;

Norman, 2005; Norman, Young, & Brooks, 2007). Research in dermatology has demonstrated a strong role for specific previous experience in the diagnostic decisions of both novices and expert physicians (Brooks, Norman, & Allen, 1991; Kulatunga-Moruzi, Brooks, & Norman, 2001), and similar patient descriptions in ECG interpretation biased the medical decisions of residents (Hatala, Norman, & Brooks, 1999). Further, policy makers across fields frequently call for students to follow the 'rule of thumb' that similar cases should be treated in a similar manner (Lovegrove, 1989 for arguments in support of precedence in law; Norman, 1988 for similar arguments in product design).

There is evidence for the reliance on exemplars in medical decision making (Brooks, Norman, & Allen, 1991; Dore, Weaver, & Norman, 2005; Hatala, Norman, & Brooks, 1999; Kulatunga-Moruzi, Brooks, & Norman, 2001; Young, Brooks, & Norman, 2007), and medical expertise may be the result of an accumulation of a large, and varied, memory store of medical examples. While there is strong support for the exemplar model in medicine, reliance to previously seen whole cases cannot be a complete explanation. If it were, we would be able to train physicians by exposing them to large banks of specific examples, and little else (for a discussion on the importance of basic science knowledge for creating coherence among symptoms see Woods, Brooks, & Norman, 2005). Additionally, exemplar based reasoning makes little distinction between relevant and irrelevant information (Brooks, 1978; 1990), there has been little definition of what counts as sufficiently similar, and what counts as a similar stimulus dimension may vary by context or task (Sloman & Rips, 1998). Although, it seems that do we rely on these rapid, unconscious matches to previously seen cases, particularly when resources are low

(Gailliot & Baumeister, 2007; Masicampo & Baumeister, 2008; Vohs, Baumeister, Schmeichel, Twenge, Nelson, & Tice, 2008), similarity matching to previously seen whole cases may be an incomplete explanation.

Knowledge as rules and examples

Medical students attend many years of medical school learning basic science, medical rules and diagnoses, treatment options, and clinical practice. Following instruction with the best available medical rules, students move onto years of supervised practice. If the rules are the best that they can be, what can be gained by practice? It is possible that the required learning amounts to the building up of many examples of the rules, in addition to learning how the language of diagnosis applies to individual patient cases – in essence, a combination of formal knowledge and experience. Perhaps learning to practice medicine requires exposure to the variability that features can present, and clinical education aids in bridging the gap between knowing the medical rules and being able to diagnose patients in clinic.

Brooks and Hannah (2006) propose that there are two levels with which we can discuss features. The *informational level* of a features is the level that is named in the medical rule – crushing chest pain, hallucinations, and a polygonal shaped papule are all features presented in the language of the diagnostic rule. But each of these features can vary across individual items or patients. The *instantiated* level is the unique perceptual manifestation of the feature presented in a given case – a patient saying that his chest feels like it is squished in a vise is an instantiation of the feature 'crushing chest pain', is

the unique appearance of a skin lesion on a particular patient is an instantiation of a polygonal shaped papule. It is possible that learning the instantiations, or the variability with which features can present, is one of the real difficulties of medical learning. We provide students with the best available medical rules, but perhaps examples are necessary to make up for the insufficiency of the language in the rule. Examples help students learn the variability in how symptoms can present, and examples may aid in the efficient use of the diagnostic category.

The distinction between information and instantiated levels of features is a theme within this research program, and a distinction that is clear in medicine. The presence of medical rules (informational level), and of the idiosyncratic ways that patients can describe their symptoms (instantiated level) makes medicine an ideal domain of study to examine the difficulty of complex decision making and the potential interaction between similarity and formal medical rules.

Knowledge and theories

In concept learning, it has been argued that theoretical knowledge is the 'glue' that holds a series of exemplars together in memory (Murphy & Medin, 1985; see also Gentner & Brem, 1999; Lakoff, 1987; Medin & Wattenmaker, 1997; Woods, Brooks, & Norman, 2005), and so the combination of examples and experience (or naïve theories) underlies much of everyday learning. It appears as though our knowledge of an item includes aspects that cannot simply be captured by perceptual similarity – we can infer that whales and sheep have spleens, but this knowledge has little to do with surface

similarity (Gelman & Wellman, 1991 for a demonstration of generalization across perceptually dissimilar items in children).

Murphy and Medin (1985) and Medin and Wattenmaker (1997) argue that concepts and categories cannot be maintained by similarity alone, that personal and causal theories are necessary (Keil, 1991). However, general knowledge effects have been hard to formalize (Dreyfus & Dreyfus, 1986; Heit, 1997; Heit & Bott, 1999; McDermott, 1987; Oaksford & Chater, 1991, 1998; Pickering & Chater, 1995). This may be where medicine provides a good venue to examine the combined role of experience and formal knowledge – the presence of strict diagnostic rules may create the structure that can focus attention to the relevant dimensions of a particular exemplar. Formal knowledge, in the form of good medical rules, allows for the diagnostic categories to be maintained by both a theory (or medical rule) and a bank of similar cases, patients, or exemplars – thus providing a framework and a set of examples to hold together a more complete mental representation of a diagnostic category.

Dual Process Models:

It is now generally accepted that clinicians use multiple strategies and types of knowledge to arrive at a diagnosis. Various authors have proposed that diagnostic reasoning is based on defining features, (reviewed in Medin, 1989; Regehr & Norman, 1996), regression approaches to weighting clinical features (Wigton, 1988; Freidman, Elstein, Wolf, Murphy, Franz, Heckerling *et al*, 1999), Bayesian models of probabilities (Weinstein, Fineberg, Elstein, Frazier, Neuhauser, Neutra & McNeill, 1980), illness

scripts (Feltovitch & Barrows, 1984), and encapsulated knowledge (Schmidt & Rikers, 2007 for a review). These models all include the assumption of consciously accessible knowledge structures accessed for and by analytical thinking processes. A major distinction is made between analytical strategies that apply various kinds of formal knowledge and formal judgment rules, and non-analytic strategies based on an unconscious match between the current problem and previous cases or exemplars retrieved from memory (proposed by Brooks, 1978; Brooks, 1990; reviewed in Norman, Young, & Brooks, 2007). Research in both cognitive psychology (e.g. Brooks & Hannah, 2006; Homa, Sterling, & Trepel, 1981), and clinical reasoning has demonstrated a role for similarity to exemplars derived from specific prior experiences (Brooks, Norman, & Allen, 1991; Hatala, Norman, & Brooks, 1999; Young, Brooks, & Norman, 2007).

The categorization of processing strategies as either analytic or non-analytic (or rules versus similarity) could be considered a direct application of dual-processing models of reasoning in cognitive psychology (as reviewed in Evans, 2008). Psychologists have uncovered ample evidence for distinct reasoning processes with very different characteristics. Analytic reasoning has been characterized as slow, energy-intensive, conceptually logical, conscious, and deliberate. In contrast, non-analytic processes are largely unconscious, rapid, highly contextualized, efficient and automatic. There are physiological correlates which seem to support these different reasoning strategies, and functional MRI has shown that the two processes implicate different parts of the brain (Goel, Buchel, Grith, & Dolan, 2000; Geol & Dolan, 2003). In particular, analytic processes tend to be associated with areas of the brain related to short term or

working memory; non-analytic processes do not (Houde, Zago, Crivello, Moutier, Pineau, Mazoyer, & Tzourio-Mazoyer, 2001). Recent evidence has demonstrated an energy dependence for reasoning strategies – that non-analytic reasoning strategies are more likely used when resources are at a low (Masicampo & Baumeister, 2008; Gailliot & Baumeister, 2007; Vohs, Baumeister, Schmeichel, Twenge, Nelson, & Tice, 2008).

The use of rules and similarity are often argued as dichotomous processes (see Pathos, 2005 for a discussion of the rules versus similarity distinction, see Sloman & Rips, 1998 for a discussion of similarity, see Croskerry, 2002; Norman, Young, & Brooks, 2007 for a review of analytic and non-analytic processing, see Hammond, Hamm, Grassia, & Pearson, 1997 for a review of intuition versus rationality, see Evans, 2008 for a review of dual processing, see Dijksterhuis & Nordgren, 2006 for a review of the benefits of thinking 'unconsciously'). Here we propose that accounting for the influence of individual manifestations of features mentioned in the rule necessitates a more sophisticated approach. Rather than suggesting the similarity based, or non-analytic, reasoning models are flawed and error prone (Bornstein & Emler, 2001; Elstein & Schwarz, 2002; Kempainen *et al*, 2003), we propose that characteristic features and specific instances can play a complimentary role (Ark, Brooks, & Eva, 2006; Brooks & Hannah, 2006; Gentner & Medina, 1998; Sloman, 1996; Whittlsea, Brooks, & Westcott, 1994; for a combined approach see Gentner & Medina, 1998; Hahn & Chater, 1998), and perhaps instantiated features are what allows the best analytical model (as provided by medical rules) to work.

Conclusions:

While the comparison of generating a diagnosis to everyday categorization is a helpful analogy, research in clinical reasoning typically demonstrates that multiple strategies can be used depending on the context (e.g. Regehr, Norman, 1996). There is evidence that clinicians rely on base rates (and students are explicitly taught using rules of thumb such as 'when you hear hooves, think of horses, not zebras': Weinstein, Fineberg, Elstein, Frazier, Neuhauser, Neutra, & McNeill, 1980), differentially weight clinical features (Freidman, Elstein, Wolf, Murphy, Franz, Heckerling, *et al*, 1999; Wigton, 1988), rely on basic science knowledge (Woods, Brooks, & Norman, 2005) and personal causal theories (Kim & Ahn 2002).

It is simplistic to think that one model of reasoning will capture all situations, however, research in this thesis will present data on two interrelated issues: first, the role of familiarity in the use of diagnostic rules. Here, I will argue that instantiated features may be what permits the use of an analytic model (in this case, authoritative diagnostic rules), when faced with the variability presented in the real world. Second, I will present data suggesting that familiar non-diagnostic information can influence current diagnostic decisions, as evidence that previous experience can influence current decisions – the calling to mind of previous experience (or instance, as encoded in memory) does not necessarily rely differentially on information that falls on a critical dimension. Therefore we can demonstrate a reliance on previous information based on non-diagnostic information, thus implying the role of previous experience in current decision making, and so demonstrate support for an instance model of reasoning. While the role of specific

previous cases has been demonstrated in visual materials (dermatology: Brooks, Norman, & Allen, 1991, EEG: Hatala, Norman, & Brooks, 1995, and static patient images: Brooks, LeBlanc, & Norman, 2000), the role of instantiated features and unique instances has yet to be well demonstrated in verbal materials. If reliance on previous experience can be demonstrated using spoken (or written) descriptions of symptoms, then this would provide further evidence of an instance model of decision making, and provide a normative and functional use of similarity in novices.

We do not claim that the phenomena discussed in this thesis are limited to medicine – there are indications of what is frequently labeled as non-analytic reasoning, or reasoning based on previous experience, in other professions such as firefighting (Klein, 1998), law (Lovegrove, 1989), business (Schwenk, 1984), human environment interaction (Kirlik, Rothrock, Walker, & Fisk, 1996), problem solving (Grant & Spivey, 2003), and everyday categorization (Brooks & Hannah, 2006) and decision making (Jacoby, Lindsay, & Toth, 1992). With indications of reliance on previous experience in other domains, the results discussed in this thesis may generalize to other areas of complex decision making. The difficulty of variability is an issue faced by learners across multiple domains, and the adaptive use of previous instances or instantiations may be a common tactic for learners in complex environments.

References

Ahn, W. – K., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., Coley, J. D., & Shafto, P. (2001). Why essesences are essential in the psychology of concepts. *Cognition, 82*, 59-69.

Ark, T. K., Brooks, L. R., & Eva, K. W. (2006). Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Academic Medicine, 81,* 405-409.

Avrahami, J., Kareev, Y., Bogot, Y., Caspi, R., Dunaevsky, S., & Lerner, S. (1997). Teaching by examples: Implications for the process of category acquisition. *The Quarterly Journal of Experimental Psychology, Section A, 50*, 586-606.

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629-654.

Barsalou, L.W. (1988). The content and organization of autobiographical memories. In U. Neisser & E. Winograd (Eds.), Remembering reconsidered: Ecological and traditional approaches to the study of memory (pp. 193-243). Cambridge: Cambridge University Press.

Barsalou, L.W., & Sewell, D.R. (1985). Contrasting the representation of scripts and categories. *Journal of Memory and Language, 24,* 646-665.

Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. *Journal of Psycholinguistic Research, 18,* 433 – 447.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual learning task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 640-645.

Bordage, G., & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Medical Education, 18,* 406-416.

Bornstrein, B. H., & Emler, A. C. (2001). Rationality in medical decision making: a review of the literature on doctors' decision-making biases. *Journal of Evaluation in Clinical Practice, 7,* 97-107.

Brooks L. R. (1978). Non-analytic concept formation and memory for instances. In: Rosch & Lloyd, eds. Cognition and Categorization. Hillsdale, NJ: Lawrence Erlbaum.

Brooks L. R. (1990). Concept formation and particularizing learning. In: Hanson, ed. Information, Language and Cognition. UBC Press, Vancouver.

Brooks, L. R., & Hannah, S.D. (2006) Instantiated features and the use of "Rules". *Journal of Experimental Psychology: General, 132,* 133-151.

Brooks, L. R., LeBlanc, V. R., & Norman, G. R. (2000). On the difficulty in noticing obvious features in patient appearance. *Psychological Science, 11,* 112-117.

Brooks, L.R., Norman, G.R., & Allen. S.W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278-87.

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory & Cognition, 35*, 1-14.

Bruner, J. S., Goodnow, J., & Austin, G. A. (1956). A study of thinking. New York: Wiley.

Carbon, C. – C. (2008). Famous faces as icons. The illusion of being an expert in the recognition of famous faces. *Perception, 37*, 801-806.

Christensen, E. E., Murry, R. C., Holland, K., Reynolds, J., Landay, M. J., & Moore, J. G. (1981). The effect of search time on perception. *Radiology, 138,* 361-365.

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 29,* 224-234.

Croskerry, P. (2000). The cognitive imperative: Thinking about how we think. *Academic Emergency Medicine, 7,* 1223-1231.

Croskerry, P. (2002) Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Academic Emergency Medicine, 9,* 1184-1204.

Dijksterhuis, P., & Nodgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1*, 95-109.

Dore, K., Weaver, B., & Norman, G. (2005). Does a rose by any other name smell as sweet? The importance of variation in case presentation to diagnostic reasoning. Talk presented at the meeting for Research in Medical Education, Washington DC.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human*

intuition and expertise in the era of the computer. New York: The Free Press.

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *British Medical Journal, 324,* 729-732.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83,* 37-64.

Evans, J. S. B. T. (2008). Dual-Processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255-78.

Freidman, C. P., Elstein, A. S., Wolf, F. M., Murphy, G. C., Franz, T. M., Heckerling, P. S., *et al.* (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation: A multi-site study of 2 systems. *Journal of the American Medical Association, 282,* 1851-1856.

Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: linking blood glucose to self control. *Personality and Social Psychology Review, 11,* 303-327.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition, 38,* 213–244.

Gentner, D., & Brem, S. K. (1999). Is snow really like a shovel? Distinguishing similarity from thematic relatedness. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty-first annual conference of the cognitive science society* (pp. 179–184). Mahwah, NJ: Erlbaum.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65,* 263-297.

Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage, 12*, 504-514.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*, B11-22.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: guiding attention guides thought. *Psychological Science, 14,* 462-466.

Groves, M., O'Rourke, P., & Alexander, H. (2003). Clinical Reasoning: the relative contribution of identification, interpretation, and hypothesis errors to misdiagnosis. *Medical Teacher, 25,* 621-625.

Hahn, U., & Chater, N. (1998). Similarity and rules: distinct? Exhaustive? Empirically distinguishable? *Cognition, 65*, 197-230.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1997). Direct comparison of the efficiency of intuitive and analytical judgments in expert judgment.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of memory and language, 34,* 686-708. *IEEE Transactions on Systems, Man and Cybernetics, SMC-17,* 753-770.

Hatala R., Norman G. R., & Brooks, L. R. (1999). Research basic to medical education: influence of a single example on subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine, 11,* 110-117.

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 7–41). London: Psychology Press.

Heit, E., & Bott, L. (1999). Knowledge selection in category learning. In D. L. Medin

(Ed.), *Psychology of learning and motivation*. San Diego, CA: Academic Press.

Hintzman, D. L. (1986). Schema-abstraction in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Hockley, W. E. (2008). The effects of environmental context on recognition and claims of remembering. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34,* 1412-1429.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 322–330.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 418–439.

Houde, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B., & Tzourio-Mazoyer, N. (2001). Access to deductive logic depends on right ventromedial prefrontal area devoted to emotion and feeling: Evidence from a training paradigm. *Neuroimage, 14*, 1486-1492.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs, 28,* 1 – 85.

Hutchinson, J. W., & Alba, J. W. (1997). Heuristics and biases in the "eyeballing" of data: the effect of context on intuitive correlational assessment. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23,* 591-621.

Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: attention, awareness, and control. *American Psychologist, 47*, 462-6.

Katz, J. (1972). Semantic Theory. New York: Harper & Row.

Katz, J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language, 39*, 170-210.

Keil F. C. (1991). The emergence of theoretical beliefs as constraints on concepts. In S. Carey & R. Gelman (Eds.) *The epigenesis of mind: Essays on biology and cognition* (237-256). Lawrence Erlbaum Associates; Philadelphia, PA.

Kempainen, R. R., Migeon, M. B., & Wolf, F. M. (2003). Understanding our mistakes: a primer on errors in clinical reasoning. *Medical Teacher, 25,* 177-181.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory based representations of mental disorders predicts their diagnostic reasoning and memory. *Journal of Experimental Psychology: General 131,* 451-476.

Kirlik, A., Rothrock, L., Walker, N., & Fisk, A. D. (1996). Simple strategies or simple tasks? Dynamic decision making in 'complex' worlds. *Human Factors and Ergonomic Society, 40th Annual Meeting,* Philadelphia, 2-6 September, 184-88.

Klein, G. (1998). Sources of power. Cambridge, MA: MIT Press.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112,* 500-526.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine, 13*, 110-6.

Labov, W. (1973). Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

LeBlanc, V. R., Brooks, L. R., & Norman, G. R. (2002). Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Academic Medicine, 77,* s67-s69.

LeBlanc, V. R., Norman, G. R., & Brooks, L. R. (2001). Effect of a diagnostic suggestion on diagnostic accuracy and clinical feature identification. *Academic Medicine, 76,* s18-s20.

Lovegrove, A. (1989). Judicial decision making, sentencing policy, and numerical guidance. New York: Springer-Verlag.

Masicampo, E. J., & Baumeister, R. F. (2008). Towards a physiology of dual-process reasoning and judgment: Lemonade, willpower and expensive rule-based analysis. *Psychological Science, 19*, 255-260.

Mayr, E. (1982). The growth of biological thought. Cambridge, MA: Harvard University Press.

McDermott, D. (1987). A critique of pure reason. *Computational Intelligence*, *3*, 151–160.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8,* 37-50.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Wattenmaker, W. D. (1997). Category cohesiveness, theories, and cognitive archeology. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization.* Cambridge, UK: Cambridge University Press.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27,* 775-99.

Murphy, G. L. (2002). The big book of concepts. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Norman, D. A. (1988). The design of everyday things. Cambridge, MA: MIT Press.

Norman, G. R. (2005). Research in clinical reasoning: past history and current trends. *Medical Education, 39,* 418-427.

Norman, G. R., Brooks, L. R., Coblentz, C. L., & Babcock, C. J. (1992). The correlation of feature identification and category judgments in diagnostic radiology. *Memory & Cognition, 20,* 344-355.

Norman, G. R., Young, M. E., & Brooks, L. R. (2007). Non-analytic models of clinical reasoning: the role of experience. *Medical Education, 41,* 1140-1145.

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Journal of Experimental Psychology: Perception and Psychophysics*, *38*, 415–432.

Nosofsky, R. M. (1988a). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 700–708.

Nosofsky, R. M. (1988). Similarity, frequency, and category representation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 54–65.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Journal of Experimental Psychology: Perception and Psychophysics*, *45*, 279–290.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language*, *6*, 1–38.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world.* Hove, UK: Psychology Press.

Pathos, E. M. (2005). The rules versus similarity distinction. *Behavioural and Brain Sciences, 28,* 1-49.

Pickering, M., & Chater, N. (1995). Why cognitive science is not formalized folk psychology. *Minds and Machines*, *5*, 309–337.

Posner, M. I., & Keele, S. W. (1968). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304–308.

Pothos, E. M., & Hahn, U. (2000). So concepts aren't definitions, but do they have necessary or sufficient features? *British Journal of Psychology, 91,* 439-450.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382–407.

Regehr, G., & Norman, G. R. (1996). Issues in cognitive psychology: implications for professional education. *Academic Medicine, 71*, 988-1001.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behaviour, 14,* 665-681.

Rosch, E., & Mervis, B. C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyles-Brian, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of experimental Psychology: Human Perception and Performance, 2,* 491-502.

Schwenk, C. R. (1984). Cognitive simplification processes in strategic decision making. *Strategic Management Journal, 5*, 111-128.

Schyns, PG, & Rodet, L. Categorization creates functional features. 1997 *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 681-696.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological*

Bulletin, 119, 3-22.

Sloman, S. A., & Rips, L. J. (1998). Similarity as an explanatory construct. *Cognition, 65*, 87-101.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*, 246-251.

Smith, E. E., Balzano, G. J., & Walker, J. (1978). Semantic factors in cognition. Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124-1131.

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: a limited-resource account of decision making, self-regulation, and active initiation. *Journal of Personality and Social Psychology, 5,* 883-898.

Weinstein, M. C., Fineberg, H. V., Elstein, A. S., Frazier, H. S., Neuhauser, D., Neutra, R. R., & McNeil, B.J. (1980). Clinical Decision Analysis. Phillidelphia: WB Saunders.

Whittlesea, B. W. A., Brooks, L. R., & Wescott, C. (1994). After the learning is over: factors controlling the selective application of general and particular knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 259-274.

Wigton, R. S. (1988). Use of linear models to analyze physicians' decisions. *Medical Decision Making, 8*, 241-252.

Wittgenstein, L. (1957). *Philosophical investigations* (3rd ed.). Oxford, UK: Blackwell.

Woods, N. N., Brooks, L. R., & Norman, G. R. (2005). The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Medical Education, 39,* 107-112.

Young, M. E., Brooks, L. R., & Norman, G. R. (2007). Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Medical Education, 41,* 1146-1151.

Zarin, D.A., & Earls, F. (1993). Diagnostic decision making in Psychiatry. *American Journal of Psychiatry 150*, 197-206.

Chapter 3:

Found in translation: the impact of familiar symptom descriptions on diagnosis in novices.

Young, M. E., Brooks, L. R., & Norman, G. R. (2007). Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Medical Education, 41*, 1146-1151.

Found in translation: the impact of familiar symptom descriptions on diagnosis in novices.

Meredith E. Young[1], Lee R. Brooks[1], Geoffrey R. Norman[2]

[1]Department of Psychology, Neuroscience & Behaviour

[2]Department of Clinical Epidemiology and Biostatistics

McMaster University

Abstract

Introduction: The language that patients use to communicate with doctors is quite different from the language of diagnosis. Patients may describe tiredness and swelling; doctors, fatigue and oedema. This paper addresses the process by which novices, who have learned standard medical terms of symptoms, use lay descriptions of symptoms to reach a diagnosis. Data in this paper indicate that the familiarity of the language used to describe symptoms influences diagnosis in novices and diagnosis is, therefore, not a simple translation into standard terms that are the basis of diagnostic decision.

Methodology: Twenty-four undergraduate students were trained to diagnose four pseudo-psychiatric disorders presented in written vignettes. Participants were tested on cases that contained two equiprobable diagnoses, one of which had the symptoms expressed using previously seen descriptions. A deviation from 50/50 in reported diagnostic probabilities was expected if the familiar symptom descriptions biased diagnostic decisions. Twelve participants were tested immediately after training, and twelve following a 24-hour delay.

Results: Participants assigned greater diagnostic probability to the diagnosis supported by the familiar feature descriptions (F (1,242) = 19.35, $p < 0.001$, effect size = .40) on both immediate (52% vs. 41%) and delayed testing (51% vs. 38%).

Discussion: The findings indicate that diagnosis is not simply based on a process of translating from patient symptom descriptions to standard medical labels for those symptoms, which are then used to make a diagnosis. Familiarity of symptom description has an effect on diagnosis and therefore has implications for medical education, and for electronic decision support systems.

Keywords: Diagnosis, decision making, education, medical, undergraduate, methods, psychology, teaching.

Found in translation: the impact of familiar symptom descriptions on diagnosis in novices.

When medical students begin to study medicine, they are taught diagnostic rules, lists of cardinal signs, and lists of textbook features for a variety of illnesses. These learned lists are usually expressed in terms of the standard medical labels for features. However, patients rarely walk into the office using medical language to describe what ails them. Patients do not use terms like retrosternal chest pain, oedema, or bilateral weakness, unless they are also health professionals. Instead, they talk about stabbing chest pain, swollen ankles, or a loss of strength. Further, the actual presentation of features in a particular patient are rarely exactly what is shown in a textbook photograph: physical features may look a bit different on patients of different age, gender, race or physical condition. Medical students have to learn to recognize these unique descriptions as *instantiations* (Brooks & Hannah, 2006) of the features that are named by the diagnostic rule. Without such a 'translational' ability, students could not communicate with colleagues or relate the instantiations to the rules and disease processes discussed in their formal learning. However, the question raised in this paper is whether diagnosis is strictly dependent on such a translation process. That is, do novices translate the feature instantiations present in the case into standard medical language, and then make their decision by matching to the terms given in diagnostic rules that they have learned?

Certainly, there are alternatives to sole dependence on such a translation process. A companion paper by Norman, Young and Brooks (2007) describes a spectrum of

diagnostic strategies, whose endpoints they characterize as non-analytic and analytic processing. Analytic processing refers to systematic, deliberate, seeking of medically relevant features and using these features to make decisions on the basis of diagnostic rules. When strictly applied, this process is restricted to medically relevant features and is generally regarded as free of the cognitive biases and diagnostic errors (Croskerry, 2003) that have recently received much attention in the popular press (Groopman, 2007; Groopman & Croskerry, 2007), and has encouraged further development of computer assisted diagnostic support systems (Freidman, Elstein, Wolf, Murphy, Franz, Heckerling, *et al.*, 1999; Graber & VanScroy, 2003).

Non-analytic processing is often placed in opposition to this rule-based approach, proposing that decisions are heavily influenced by similarity to a prior case. Non-analytic processing is often used to refer to holistic case similarity (Croskerry, 2003), or similarity in patient information that in itself is non-diagnostic (Hatala, Norman, & Brooks, 1999). Similarity of whole cases or patient identifiers is not conclusive evidence, but it could serve an important mnemonic function when a diagnostician is struggling with a large number of possibilities or with time pressures. Under some circumstances, the use of overall case similarity has been shown to lead to more accurate diagnosis (Brooks, Norman, & Allen, 1991).

However, it is possible that non-analytic or similarity based processes have an influence even when dealing with individual, medically relevant features. We could easily imagine that if a patient describes her symptoms using familiar words, the translation to diagnostic language would be facilitated. It is also possible that this familiar

language would suggest a diagnosis, or be taken as more convincing evidence than would the same feature expressed in novel form. Such an effect of familiar feature instantiations has been demonstrated in cognitive psychology (Brooks & Hannah, 2006), and the impact of familiar terms has been demonstrated in medical education when the alternative novel terms are simple synonyms. For example, Dore *et al* (2005) have demonstrated that novice diagnosticians rely more heavily on familiar synonymous features (i.e. sleeplessness and insomnia) than on equally valid, novel synonyms (i.e. inability to sleep, wakefulness) (Dore, Weaver, & Norman, 2005).

There are several reasons why familiar feature instantiations may impact diagnostic decisions. First, the difficulties faced by a novice diagnostician could be strictly a matter of recognizing the synonymity of different instantiations, or low confidence in recognition. A novice diagnostician could have difficulty recognizing a novel presentation as an instantiation of a term in a decision-making rule. For example, students may spend a lot of time learning that Cushing's syndrome is associated with a "moon-shaped face" and that acute myocardial infarction often presents with retrosternal chest pain, but may be unable to recognize a moon-shaped face (LeBlanc, Norman, & Brooks, 2001) or determine that a chest pain "that feels like my chest is put in a vice" is the same as the textbook description. Second, the variety between symptom presentations could trigger more direct effects of familiarity. A symptom instantiation that the student has heard before may be closely associated with the diagnostic context in which it was previously seen. For example, a previous depressed patient complaining of 'tossing and turning all night' may sway the clinician to the diagnosis of depression when a

subsequent patient expresses herself in the same way. In any of these cases, the process is not strictly rule based, instead having some of the characteristics of non-analytic reasoning. The latter kind of familiarity effects could be an intermediate phenomenon between the traditional dichotomous poles of analytic and non-analytic processing. In either case, the novelty or familiarity of a unique presentation may affect the diagnosis eventually made.

In this paper, we will first investigate the impact of familiar symptom descriptions on novice diagnosticians. The specific area of study for this paper is psychiatric diagnosis, a field in which verbal descriptions from patients is a key source of information, and in which there is a large variability in the description of any one symptom. We will then determine if our evidence allows us to select between two possible sources of the impact on diagnosis of familiar feature instantiations – to determine whether familiarity of the instantiation of a feature directly impacts the diagnostic decision of novices, or whether it is only having an indirect effect by allowing novices to better translate familiar instantiations into rule language. Any familiarity effects suggest that the novice has more than the rules to learn, thus knowing more about the source of these effects might guide the development of educational strategies to help with this initial and complex learning.

Methods

Participants

A group of twenty-four year 1 psychology undergraduate students participated in this study in return for course credit. Undergraduate students were chosen above medical students in order to ensure participants would have limited knowledge regarding psychiatric diagnoses. This study was approved by the McMaster University Research Ethics board.

Stimuli

Four pseudo-psychiatric disorders were created for the purposes of this experiment. The psychiatric disorders with modified diagnostic rules included Schizophrenia, Mania, Obsessive Compulsive Disorder and Paranoid Personality Disorder. Each disorder was characterized by four unique symptoms, limiting the possibility of confusion between the diagnostic categories. Subjects were told that the disorders did not reflect the real psychiatric disorders of the same name, so to focus on the diagnostic rules presented in the experiment. Psychiatry was chosen for this study because it involves high levels of variability in verbal symptom presentation (i.e. the many different ways delusions or obsessive thoughts can be described) whereas other areas of medicine (e.g. cardiology, nephrology) are less dependent on patient's verbal descriptions of their symptoms.

Each diagnostic feature (e.g. hallucinations) was presented in a variety of instantiations within the case vignettes (for example: hearing voices when no one else is in the room, seeing visions before going to bed, etc.). The various descriptions were

constructed to differ in their 'ease of translation' into their feature label, from near synonymous terms to full behavioural descriptions. A range of difficulty of translation was included in order to expand the work of Dore *et al* (2005) that demonstrates the impact of familiar synonymous terms (Dore, Weaver, & Norman, 2005). By including near-synonymous terms (e.g. needs only two to four hours of sleep and decreased need for sleep), and complex behavioural descriptions (e.g. due to her husband's job loss, she has started working both day and night shifts, and does not feel tired), we were able to evaluate the impact of familiarity using more complex and more realistic instantiations.

Procedure

All materials were presented on a computer, programmed with RunTime Revolution version 2.5 (RunTime Revolution Ltd. Edinburgh, Scotland).

Learning Phase: Participants were shown the four features that were diagnostic of a pseudo-psychiatric disorder (e.g. hallucinations, delusions, disorganized speech and disorganized behaviour for schizophrenia). They were asked to study the list and then to identify four features that were characteristic of the disorder from a list of 16 features. Upon completing this task, the next disorder was presented in the same way and this sequence repeated until all four diagnostic categories had been presented. To complete the learning phase, participants had to complete a quiz, with which they were required to identify all features for each disorder. The pass score was set at 15/16. A pass resulted in the participant moving into the practice phase. If participants did not pass the quiz, they re-started the learning phase.

Practice Phase. Participants were shown a series of 12 different cases, three for each of the different pseudo-psychiatric disorders, presented in random order. Each case included a patient description (including name, age, occupation and familial situation), and the four features characteristic of one of the four disorders. In the practice phase, features presented were instantiated in a manner consistent with the way patients might describe their symptoms (e.g. hearing voices). The practice phase served two purposes: it allowed participants to gain some expertise with these disorders by providing diagnostic feedback, and it exposed participants to a set of feature instantiations that will function as familiar instantiations in the test phase.

Test Phase. Participants in the 'immediate test' condition moved immediately to the Test phase, whereas participants in the 'delay' condition returned to the lab to complete the test phase after a 24-hour delay period. Participants diagnosed a total of 12 test cases.

All test cases included a total of 4 features presented in unique instantiations. Each case included 2 familiar feature instantiations (drawn from the practice cases) indicative of one disorder, and 2 novel familiar feature instantiations indicative of another disorder. Participants were asked to assign a diagnostic probability rating to each of the four disorders, and to report the diagnostically relevant features. Due to the presence of two supporting features for two different diagnoses, the 'expected' unbiased response would be a 50:50 split in diagnostic probabilities assigned. If familiarity of feature descriptions impacts upon the assignment of diagnostic probabilities, we would expect to see a deviation from 50:50 in favor of the diagnosis supported by the familiar feature descriptions.

Results

The diagnostic probability assigned to each diagnosis was recorded for each participant. A repeated measures analysis of variance (ANOVA) was conducted using the 2 plausible diagnoses (from the 50:50 design), as the within-subjects comparison, individual cases as the repeated measure, and the testing time (either immediate or delay) as a between-subjects comparison. Across both testing conditions, participants assigned significantly more diagnostic probability to the diagnosis supported by the familiar feature descriptions (F (1,242) = 19.352, $p < 0.001$, effect size (δ) = .40). There was no significant interaction between delay and the impact of familiarity (F (1,242) = .193, $p > 0.05$), indicating that the impact of familiar symptom descriptions did not change with delayed testing. Mean scores are shown in Figure 3.1.

Insert Figure 3.1 about here

In analyzing the features reported by participants, it was found that ten of twenty-four participants reported the features entirely in the instructed 'rule' language (i.e. translated from the specific instantiation presented). These ten participants reported nearly all features present in the test cases (99.8% of all features were reported), and reported them all in their translated form. This provides us with an indication that novices can – even when not instructed – translate correctly and effectively into diagnostic rule language. These participants did show an effect of familiarity comparable to that of the

whole group (51% vs. 42%), indicating that even 'perfect translators' are being influenced by familiar feature descriptions.

Discussion

This study demonstrates that a single, unique, prior instantiation of a medically relevant feature can affect a subsequent diagnosis. The impact of familiarity is thus not only at the level of holistic, non-analytic processing (Brooks, Norman, & Allen, 1991), but also at the time of feature interpretation. This study also indicates that this effect is not transient, given that a familiar feature instantiation will continue to bias diagnostic decisions after a 24-hour delay. If a single, unique experience impacts the diagnostic decisions of novices following a 24-hour delay, the we assume this experience is stored in long term memory (Kellogg, Newcombe, Kammer, & Schmitt, 1996; Squire, Knowlton, & Musen, 1993), which leads us to believe that familiarity could continue to impact diagnosis in longer delay periods, making it an influence to be taken seriously, at least for novices.

A sub-sample of subjects chose to report the diagnostically relevant features in the language of the diagnostic rule. These participants showed high accuracy in their 'translations', but still showed an influence of familiarity. These results indicate that diagnosis among these novices is not just a matter of recognizing instantiations of features and then matching their labels to the terms in the rules that had been learned. Future research should also examine of the impact of familiar feature instantiations in expert physicians. Past research has indicated that expert doctors are influenced by

familiarity of whole cases (Brooks, Norman, & Allen, 1991), so it is possible that there would also be an effect of familiarity at the level of feature instantiation.

A rule cues us to search for particular features, or to confirm the features already found. It is important to remember, no matter how good a rule is, it does not imply the extent to which that feature might vary in appearance or expression. This particular domain of study, psychiatry, is one that is rich in feature variability. If one considers just the symptom of hallucinations, a patient could describe an auditory, olfactory, tactile or visual hallucination – all of which fall into the same symptom category, but a category for which there is incredible variability in the individual presentation.

Diagnostic error has recently received much attention in the public domain. The call has been for more reliance on analytic, rule based approaches to medicine, and a call for instructional programs to teach physicians to 'watch out' for potential cognitive biases that stem from pattern-recognition processes (non-analytic). However, a growing literature challenges the idea that the move from a description of a symptom to a symptom label is an unambiguous one, and this observation has direct implications in medical education, and the development of diagnostic support systems. The current results suggest that caution is needed before pursuing decision support systems, as the system is dependent on the physician's ability to input the correct 'translation' into the decision support system (Friedman *et al.*, 1999; Graber & VanScory, 2003). The present study shows that, at least in novices, this process is far from unbiased. Without this skill, the diagnostic systems become ineffectual, which can be seen in the decreased accuracy of diagnoses generated by the system when medical students or physicians input the

symptoms (Friedman *et al.*, 1999). Indeed, since interpretation of features can be strongly influenced by a suggested diagnosis (LeBlanc, Norman, & Brooks, 2001), there is every reason to presume that the decision support system will be as prone to confirmation bias as the clinician who is supposed to be de-biased by using the system.

In conclusion, the process of medical diagnosis is a complex area, which we are slowly beginning to understand. The purpose of this paper was to investigate the impact of familiar instantiations on the diagnostic process of novices. We were successful in demonstrating such an effect for novices using materials that are plausible for psychiatric diagnosis and that lasted for at least 24 hours. This supports an interest in the much more concrete information that is provided by specific instantiations of features. From the moment a patient walks through a clinic door, she is communicating with the physician – her gait, colouration, mannerisms, and appearance are providing information that may influence a diagnosis in ways that are not captured by a strictly analytic understanding of the process of diagnosis.

References

Brooks, L. R. & Hannah, S.D. (2006) Instantiated features and the use of "Rules". *Journal of Experimental Psychology: General, 132,* 133-151.

Brooks, L.R., Norman, G.R., & Allen. S.W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278-87.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to prevent them. *Academic Medicine, 87,* 1-6.

Dore, K., Weaver, B., & Norman, G. (2005). Does a rose by any other name smell as sweet? The importance of variation in case presentation to diagnostic reasoning. Talk presented at the meeting for Research in Medical Education, Washington DC.

Freidman, C.P., Elstein, A.S., Wolf, F.M., Murphy, G.C., Franz, T.M., Heckerling, P.S., Fine, P.L., Miller, T.M., & Abraham, V. (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation. *Journal of the Americal Medical Association, 282,* 1851-1856.

Graber, M.A., & VanScroy, D. (2003). How well does decision support software perform in the emergency department? *Emergency Medicine, 20,* 426-428.

Groopman, J. (2007). What's the trouble? How doctors think. *The New Yorker*, January 29.

Groopman, J., & Croskerry, P. (2007). Is there a House in the Doctor? *The Globe and Mail* March 24: F8.

Hatala R., Norman G. R., & Brooks, L. R. (1999). Research basic to medical education: influence of a single example on subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine, 11,* 110-117.

Kellogg, R.T., Newcombe, C., Kammer, D., & Schmitt, K. (1996). Attention in direct and indirect memory tasks with short- and long-term probes. *American Journal of Psychology, 109,* 205-217.

LeBlanc, V.R., Norman, G.R., & Brooks, L.R. (2001). Effect of diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Academic Medicine, 76,* S18-20.

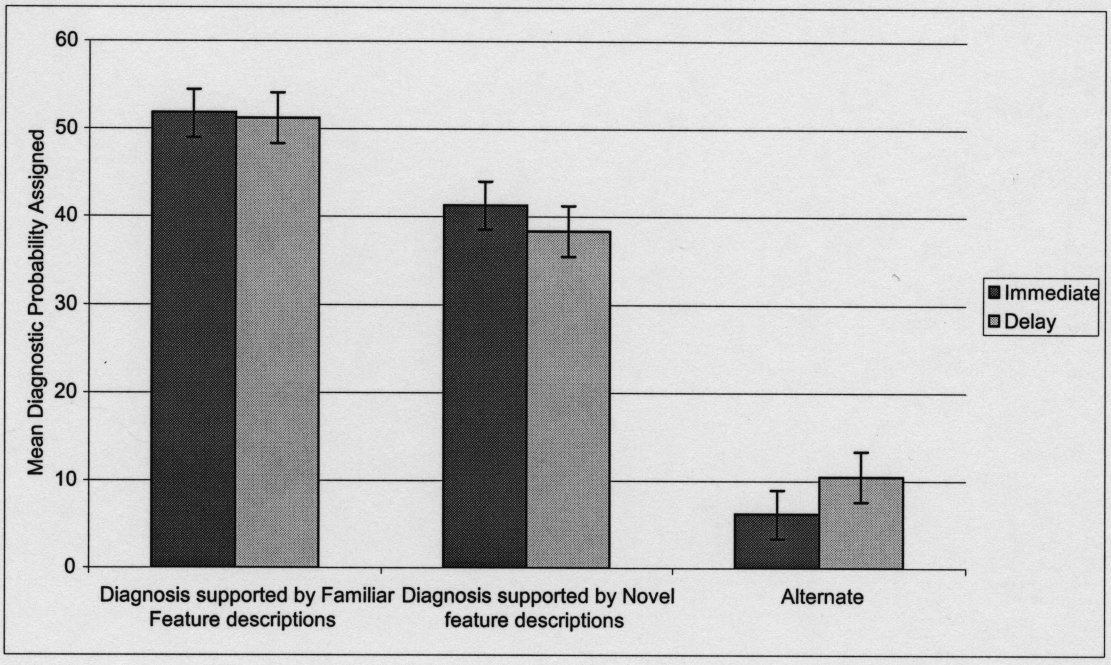
Norman, G. R., Young, M. E., & Brooks, L. R. (2007). Non-analytic models of clinical reasoning: the role of experience. *Medical Education, 41,* 1140-1145.

Squire, L.R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review in Psychology, 44,* 453-95.

Figure Captions

Figure 3.1: Mean diagnostic probability assigned to the diagnosis supported by familiar feature instantiations and the diagnosis supported by the novel feature instantiations. The two other diagnostic options are collapsed and shown in the 'alternate'. Data for the immediate test group (n=12) and the delay test group (n=12) are shown. Error bars indicate the standard error of the mean.

Figure 3.1:


60
50
40
30
20
10
0
Mean Diagnostic Probability Assigned
Immediate
Delay
Diagnosis supported by Familiar Feature descriptions
Diagnosis supported by Novel feature descriptions
Alternate

Chapter 4

Preference for familiar feature instantiations: Using rules in medicine

Experiments 1 and 2

This paper is in preparation for submission to: Journal of Experimental Psychology: Learning, Memory and Cognition.

Preference for familiar feature instantiations: Using rules in medicine

Meredith E. Young[1], Geoffrey R. Norman[2], Lee R. Brooks[1]

[1]Department of Psychology, Neuroscience and Behaviour,

[2]Department of Clinical Epidemiology and Biostatistics,

McMaster University

Abstract

When learning medicine, medical students are taught lists of symptoms and accompanying diagnoses. However, as they acquire expertise, they must also learn to recognize signs and symptoms and deal with substantial variability in presentation. This study investigates the impact of this variation in symptom presentations on diagnosis, and whether reliance on similarity could be reduced by the application of counting rules (by instituting a minimum number of features to be named in support of a diagnosis). Participants practiced diagnosis on written prototypical case vignettes, and were tested using cases in which two diagnoses were equally probable in terms of the learned rules (i.e. had two features supporting one diagnosis, two features supporting another). Participants assigned significantly more diagnostic probability to the disorder supported by familiar feature instantiations, even when participants were cued to count individual features while assigning probabilities. We discuss the normative value of this bias for familiar instantiations in natural materials, even when having an authoritative identification rule.

Preference for familiar feature instantiations: Using rules in medicine

In medical school, students are taught rules specifying the signs and symptoms of many disorders, but in order to practice medicine they must be able to recognize the variability with which symptoms can present – no easy task for a learner. The initial focus of medical learning is on overarching, abstract, and generalized symptom labels and cardinal features, essentially learning the rules of medical diagnosis. Medical students learn about aspects of disease like 'acute onset', 'insomnia', and 'dyspnea' (shortness of breath). They learn that a 'mica-like' scale is a cardinal sign of psoriasis, and that a 'moon-shaped face' is a classic symptom of Cushing's disease. However, the general language meaning of terms like 'mica-like scale' and 'moon shaped face' is insufficient for competent diagnosis. Before students can become competent, they must become confident in identifying a mica-like scale that is associated with psoriasis, and how these scales may look different on individuals of different ethnicities, ages, and health states. They must learn to distinguish mica-like scales on a variety of individuals from papules or scales indicative of another dermatological disease, and what an atypical presentation of psoriasis related mica-like scales may look like. There is evidence that the identification of even prototypical features can be difficult or ambiguous for both students and medical experts (Brooks, LeBlanc, & Norman, 2000; Groves, O'Rourke & Alexander, 2003; Hatala, Norman, & Brooks, 1999; Young, Norman, & Brooks 2007; Zarin & Earls, 1993), so feature detection and identification cannot typically be an easy task for learners. Here we propose that learners rely on familiar manifestations of

diagnostically relevant features, both typical and atypical, as a normative strategy when faced with the insufficiency of the language in the rule for feature identification in medicine.

In natural category learning, and in medicine, features can be considered on two levels. The *informational* level of a feature is the abstract description of features that facilitate communication and learning, often characterized by the words physicians use to describe the symptoms of their patients when speaking with one another. The informational level is also typically used during instruction or teaching – they provide the language of rules. "Acute onset", "dyspnea" and "hallucinations" are all considered to be informational descriptions, as they convey little about the particular manifestation of the feature, but are rich in specialized meaning. *Instantiations,* on the other hand, are the unique manifestations of the informational feature. For example, reporting "hearing voices when no one else is in the room" would be considered an instantiation of the informational feature "hallucinations". The distinction between informational and instantiated levels of features (Brooks & Hannah, 2006; Young, Brooks, & Norman, 2007), is central to the research in this paper[1].

There have been several demonstrations of the role of familiar feature instantiations, with participants demonstrating a bias towards categorizing ambiguous stimuli (Dore, Weaver, & Norman, 2005; Young, Brooks, & Norman, 2007) and transfer

[1] Because the particular instantiations of a features used in this study are often manifested in verbal behaviour and are commonly reported verbally, the distinction between these two levels of features cannot be captured by calling one of them 'perceptual' and the other 'verbal' as may be implied by previous research (Brooks, LeBlanc, & Norman, 2000; Hannah & Brooks, 2006; Kulatunga-Moruzi, Brooks, & Norman, 2001).

stimuli (Allen & Brooks, 1991; Hannah & Brooks, 2006; Regehr & Brooks, 1993) as members of the category supported by the familiar feature. These studies can be taken as indicating an increased weighting, or increased reliance, on familiar features compared to equally reliable, equally 'good', novel instantiations. Further investigations have demonstrated an influence of individual feature manifestations within the context of a categorization rule. Expansions on the work of Brooks and colleagues, originally designed to show the effect of overall similarity to prior instances (Allen & Brooks, 1991; Regehr & Brooks, 1993), have demonstrated the role of close perceptual matching of individuating features between training and test stimuli (Thibaut & Gelaes, 2006), and that the influence of unique feature instantiations can be observed within a strict categorization rule in the absence of non-rule features (Lacroix, Giguere, & Larochelle, 2005; Thibaut & Gelaes, 2006). With these demonstrations indicating the importance of familiarity and similarity, even within the context of a strict categorical rule, it is possible that a stock of familiar manifestations of a feature is what allows a learner to translate the abstractions mentioned in a learning rule into the practice of effective categorization in the world.

'Familiar instantiated features', as used in this work, is not the same as 'overall similarity to prior instances.' Familiar feature instantiations can correspond to terms (informational features) in a rule, and therefore do not necessarily represent a fundamentally different process of categorization than the application of a rule. However, the research just summarized uses simple visual stimuli, typically in the form of artificial animals. In order to consider the generalizability of these findings, and to specifically

tackle the role of instantiated features in combination with authoritative rules in complex decision making, we propose the use of medicine as a realistic domain of study, a perspective that is currently atypical in the literature.

The practice of medicine is a unique and rich domain in which to tackle the consolidation of formal knowledge with the variability of real world categorization. Medicine is a categorization task that is learned in adulthood, and learned within a structured learning environment (i.e. formal instruction). We know that during categorization tasks people rely on general knowledge (Keil, 1991) and basic science or causal knowledge (see Schmidt, Norman, & Bishuizen, 1990 for a review) and the impact of these kinds of knowledge can be seen in medicine (e.g. Woods, Brooks, & Norman, 2005 for basic science knowledge, and Kim & Ahn, 2002 for personal or causal theories). It is possible that the reliance on individual previous experience, particularly in the form of unique feature instantiations, is an important component in what connects personal theories and formal knowledge to their use in everyday reasoning. Thus a strong distinction between rules and similarity, and the assumption that they occur as essentially different processes, may be a perspective that is incomplete (see Pothos, 2005 for a discussion of the rules versus similarity distinction).

We have chosen to investigate the role of familiar instantiated features in a domain, psychiatry, in which there is high feature variability, clear diagnostic rules with little category overlap, and in which feature identification is notoriously difficult. The general purpose of the studies included in this paper is to investigate the impact of familiar feature instantiations on categorization in novices. The materials used in these

studies have been created in order to mimic the complexity of concept learning in a realistic domain and to reflect the early training and decision making skills involved in the acquisition of expertise.

Medical training often begins with learning rules containing diagnostic criteria, often presented in the form of informational level features. Following the mastery of a diagnostic rule students encounter patients, or case descriptions of patients, where instantiated features in the form of unique symptom descriptions or a 'patient voice', are introduced. This aspect of medical learning, and concept learning, is quite generalizable, and so for the studies presented here, care has been given to mimic the structure of this learning situation. To this end, the materials have been carefully constructed to contain both informational features and a variety of corresponding instantiated features, are taught using prototypes, and feedback is provided to the participants. The materials have a variety of instantiations of each of the abstract terms (informational features) to simulate the challenges faced by students as they learn to identify examples of the terms in the rules. The instantiations are intended to be relatively easy to identify (i.e. named using the abstract terms in the rules) and unambiguous. Ambiguity between diagnostic categories was kept to a minimum. Each disease was drawn from a different axis of psychiatry, had no overlap in either informational level or feature instantiations with other diagnostic categories, and each diagnosis contained the same number of diagnostic features. These constructed diseases are simplified versions of true psychiatric diagnoses, but retain the general features of actual psychiatric learning. In these materials, the informational features are clearly important in learning since they are the terms in the

diagnostic rules initially given to the learners. However, since there is no overlap of informational features between diagnostic categories, they do not contribute to difficulty in learning to categorize cases. Instead the difficulty, as is true in many cases of everyday learning, lies in learning to apply what initially are general language terms to manifestations of those features in instances of the category: learning to identify examples of 'shortness of breath' that indicate asthma from shortness of breath indicating other diseases such as pneumonia, or heart failure.

In summary, the following studies will address the role of familiar feature instantiations in the diagnostic processes in novices. These experiments address some fundamental topics in concept learning, using materials and procedures that are close to a real-world diagnostic concept learning task.

Experiment 1

This experiment was designed to investigate the effect of familiar feature instantiations on the categorical, or diagnostic decisions of novices. This experiment examines the role of a form of similarity in the context of a medical rule –investigating whether familiarity of the instantiations of one or two diagnostically relevant features can influence a complex categorization task. In the following studies, participants were trained to competence on simple diagnostic rules, and later tested on more complex, or ambiguous cases. Adapted medical materials were used in order to mimic realistic, complex decisions reliant on multiple cues. Diagnostic categories were designed to be unambiguous, with no between-category overlap. Participants were trained to a base level

of competence through formal instruction, and practiced on prototypical case vignettes, receiving feedback throughout training. The materials constructed for training the participants were strictly prototypes for their disease categories; that is, all features, both at the informational and instantiated level, indicate only one diagnostic category, and all relevant features were present in each vignette. Each symptom was presented in the form of a unique instantiation, and participants were required to learn the relationship between the instantiation and the informational level of the feature. Instantiated features were created to mimic a 'patient voice', and were designed to be quite complex, and idiosyncratic. This variability notwithstanding, participants received repeated feedback that each of these unique instantiated features does represent the same informational level feature, and thus was indicative of the overall disease category. By the inclusion of both informational features (in learning the medical rules) and instantiated features (contained within case vignettes in subsequent training), we construct a simple model of the process of learning to identify features of cases in descriptive terms, terms that are critical for communication and diagnosis in medicine. This supportive training, and the complexity of the materials to be learned, was designed to reflect a realistic training challenge faced by anyone trying to learn complex categories who must learn both the diagnostic rules and the varied, perceptual manifestations of those diagnostic features. Additionally, authoritative diagnostic rules were taught, and learned by participants, so any influence of similarity is being demonstrated not only in spite of the presence of authoritative rules, but within their use.

Methods

Participants

Thirty-six first-year undergraduate students participated in this experiment in exchange for course credit. Eighteen students were tested immediately following training, and the remaining eighteen were tested following a 24-hour delay. This study was approved by the McMaster University Research Ethics Board, and written consent was obtained from all participants.

Stimuli

A set of four modified psychiatric disorders were created for research purposes. These disorders included mania, schizophrenia, obsessive-compulsive disorder, and paranoid personality disorder. The rules for diagnosis were simplified from the traditional definition of the psychiatric disorders of the same name, and participants were told that these disorders were fictional, and were not representative of real world illnesses.

Each disorder was characterized by four unique informational level features, and each disorder was drawn from a different area of psychiatry, see Table 4.1 for an example. These restrictions were imposed in order to ensure little between-disease category uncertainty or ambiguity. Each informational level feature was presented as an instantiated feature during the practice and test phases of the experiment in the context of a written case vignette. The feature instantiations were behavioral descriptions (e.g., her husband comments that he has found her alone in a room seemingly laughing at a joke,

even though no one is around) and phrases approximating how a patient might describe their symptoms (e.g., he reports that he has been working double shifts six days a week, but does not feel the need to sleep). These instantiated features were combined to create written case vignettes that were either prototypical cases (containing all features indicating one disorder) or mixed cases (containing features supporting two different diagnostic categories). The prototypical cases were seen only in training, and the mixed cases were seen only in the test phase.

Insert Table 4.1 about here

Procedure

All materials were presented using RunTime Revolution 2.1 (RunTime Revolution Ltd. Edinburgh, Scotland), and presented on a computer screen. The experimental protocol included a rule learning phase, a case practice phase and a test phase.

Rule learning Phase. For each disorder, participants were asked to study a list of the four features characteristic of that disorder. They were then required to correctly identify the four features characteristic of the studied disorder from a list of all 16 symptoms. If participants were unable to accurately identify the four informational features, they returned to the original study list. If they correctly identified the four informational features, they then saw the next diagnostic rule, and this sequence was repeated until all

four disorders were learned. In order to continue to the practice phase of the experiment, participants were required to complete a quiz in which they were required to correctly identify all features for each disorder and achieve a score of at least 15/16. If participants were unable to reach this criterion score, they restarted the rule learning phase. The inclusion of a strict learning criterion, across each disease category and at the completion of training, was included to ensure a high level of competency with the diagnostic rules. The order of learning was randomized, but the learning protocol was identical for each disorder to be learned.

Case practice Phase. The practice phase contained 12 prototypical case vignettes, each of which contained patient identity information (name, age, area of residence, employment, familial situation, etc) and four unique instantiations of the informational features found in the diagnostic rule. These cases were considered to be prototypical as all training cases contained all four features mentioned in the diagnostic rule, presented in their instantiated form. There were three practice cases for each of the four disorders, all presented in random order. Participants were required to assign diagnostic probabilities to each of the four diagnostic options, and report the diagnostic features found in the case in an open text field. After participants reported the diagnostic features they detected, they received feedback regarding the correct diagnosis, the diagnostically relevant features presented in the case vignette and their accompanying informational level labels. Feedback included providing participants with the correct diagnosis, and the highlighting of the instantiated features labeled with the language of the medical rule. The purpose of the practice phase was two fold – to help participants achieve some level of expertise in assigning

diagnostic probabilities, and to expose them to a variety of instantiations of diagnostically relevant features that would function as "familiar" instantiations in the test phase of this experiment.

Test phase. Participants were randomly assigned to either the immediate test group, who went directly from the case practice phase to the test phase, or the delay test group, where participants moved to the test phase following a 24-hour delay. Participants saw a total of 12 test cases, and the order of presentation was randomized. Each test case contained a total of four symptoms presented in instantiated form, and each case contained two, objectively equivalent, diagnostic options. Test cases contained two familiar instantiated features drawn from a case in the practice phase indicative of one disorder, and two novel feature instantiations indicative of another disorder, creating two equiprobable diagnoses. Participants were asked to assign diagnostic probabilities to each of the four possible disorders, and to report diagnostically relevant features. If participants were unaffected by the familiarity of the instantiated features, then an unbiased response of 50/50 is expected. If participants were influenced by the presence of familiar feature instantiations, then we would expect a deviation from a 50/50 split in diagnostic probabilities in favor of the diagnosis supported by the familiar feature instantiations.

Self-reported diagnostic strategy. Following the completion of the test phase, participants were asked to describe their diagnostic strategy using a continuous scale anchored at the extremes by a 'generalizing strategy only' and a 'counting strategy only'. Participants were told that a counting strategy would indicate that they were counting the symptoms and then 'doing the math' to determine diagnostic probabilities. If they were using a

counting strategy only, each symptom would receive equal weighting (as there was four features in each case), and no other information would influence their decision. A generalizing strategy was explained as assigning probabilities according to their 'gut instincts', without keeping track of the individual symptoms. A combined strategy could be 'feeling' that the case represented a particular disorder, and then seeking features that might support that hypothesis. Participants were asked to use a scroll bar to indicate which strategy, or combination of strategies, best described their approach to diagnosis on a visual analog scale. Participants were also given the opportunity to verbally explain their choice of strategy, and an experimenter verbally assessed the participants' strategy on debriefing.

Analysis

Diagnostic probabilities. For each participant, the probability assigned to the diagnosis supported by the familiar feature instantiations and the probability assigned to the diagnosis supported by the novel feature instantiations was averaged across twelve test cases, creating a mean diagnostic probability score for both the familiar and plausible alternative diagnostic categories. Mean diagnostic probabilities were analyzed using a mixed design ANOVA, where the within subject factor of interest was mean probability assigned to the two possible diagnoses (one supported by familiar feature instantiations and the other by novel feature instantiations), and the between subject factors of interest were the impact of delay (immediate versus a 24-hour delay), and self reported strategy (mainly counting, mainly generalizing, and mixed strategy).

Impact of familiar instantiated features on feature identification. During each test case, participants are asked to assign diagnostic probabilities to each of the four possible pseudopsychiatric disorders. Following the assignment of probabilities, participants are asked to report the diagnostically relevant features in the case. These responses were coded as: total number of features reported, the number of familiar feature instantiations reported, and the number of novel feature instantiations reported, and mean scores for each participant were created. Mean number of features identified was analyzed using a mixed design ANOVA, where the within subject factor of interest was number of features identified (mean number of familiar features reported compared to the mean number of novel features reported) and the between subject factor of interest was delay (immediate versus 24-hour delay).

Results

Diagnostic probabilities

Mean diagnostic probabilities assigned for both immediate and delay testing group are shown in Figure 4.1.

Insert Figure 4.1 about here

Participants assigned significantly higher diagnostic probability to the disorder supported by the familiar instantiated features than the diagnosis supported by the novel

feature instantiations (mean probability assigned to the diagnosis supported by the familiar feature instantiations = 52.2%, mean probability assigned to the diagnosis supported by the novel feature instantiations = 43.1% (F (1, 30) = 13.13, p = .001))[2]. There was no significant interaction between testing group (immediate vs. delayed testing) and the impact of familiar instantiated features (F (1, 30) = 3.90, p = .07).

Impact of familiar instantiated features on feature identification

Participants reported slightly more familiar features than novel instantiations (1.81 familiar feature instantiations compared to 1.70 novel feature instantiations), although this did not reach statistical significance (F (1, 30) = 3.22, p = 0.08). The number of features did not differ by delay (F (1, 30) = .025, p = .87), nor did the number of familiar features identified interact with delay (F (1, 30) = .59, p = .46).

Self-reported strategy

When rating strategy, participants were asked to move a toggle along a visual analog scale that was anchored using 'counting strategy only' and 'generalizing strategy only'. Participants saw only the written anchors, but for the purposes of analysis, an arbitrary -100 to +100 scale (where -100 referred to 'counting strategy only' and +100 refers to 'generalizing strategy only') was used to record the participants placement of the toggle along the visual analog scale. When collapsed across delay groups, mean reported strategy was -28.8, indicating that, on average, participants reported using more of a

[2] When a larger analysis was conducted, including the diagnostic probabilities assigned to the alternate (as defined by the diagnostic probabilities assigned to the two diagnostic categories not intentionally represented in the case vignette), there is still a significant difference in the mean probabilities assigned to the familiar feature instantiations and the other alternate diagnoses.

counting strategy than a generalizing strategy. Participant responses were coded as either a counting strategy (lower cutpoint = -100, upper cutpoint = -30, range = -100 to -45), a mixed strategy (lower cutpoint = -30, upper cutpoint = +20, range = -24 to +17), and mainly generalizing strategy (lower cutpoint = +20, upper cutpoint = +100, range = +25 to +100). The cut points were created to be consistent with the verbalized self-report of strategy during debriefing with the experimenter. When categorizing individual participant responses, 16 reported using mainly counting strategies, 8 reported using mainly generalizing strategies, and 12 reported using combined strategies. There was no significant interaction between self reported strategy (counting strategy, combined strategy, or generalizing strategy) and the influence of familiarity (F (2, 30) = 2.12, p = .137). This self-report measure will serve as an evaluation tool to validate the success of tacit manipulations in Experiment 2.

Discussion

These data demonstrate that familiarity with a unique, complex feature instantiation can influence diagnostic probability judgments. The effect of familiar instantiated features did not change following a 24-hour delay, indicating a lasting effect of prior instantiated features.

The study adds to the growing literature demonstrating the impact of prior instances, or instantiations (Brooks & Hannah, 2006; Brooks, Norman, & Allen, 1991; Hatala, Norman, & Brooks; Kulatunga-Moruzi, Brooks, & Norman, 2001; Young, Brooks, & Norman 2007), and demonstrates that the impact of similarity is not limited to

visual feature instantiations (Brooks & Hannah, 2006) or in overall similarity to whole previous cases (Brooks, Norman, & Allen, 1991). The results of this study indicate a role of similarity within the context of an authoritative rule, that familiar instantiations of a definitionally relevant feature are perhaps being over-weighted compared to the equally valid novel instantiations present in the case vignette. Perhaps this indicates that participants are stopping their search for features prematurely – they find familiar features more readily than novel features, are comfortable with their understanding of the informational label, and as a result, overweight them compared to the novel feature instantiations.

In many clinical (and experimental) situations, one must provide rationale, or support, for a diagnostic (or categorical) decision, particularly in the case of novices. This rationale or justification typically involves reporting the features that were present, or that led to a diagnosis. For example, one considers a patient to have a heart attack *because* he has shortness of breath, and pain radiating down his left arm. Often a minimum criteria diagnostic rule must be met – a patient *must* have three of the following symptoms in order to be diagnosed.

Most medical rules, with the exception of The Diagnostic and Statistical Manual of Mental Disorders (the DSM-IV (American Psychiatric Association, 2000)) used in psychiatry, do not have a specified number of features that must be detected in order to make a diagnosis. By report from experienced diagnosticians, the quality and clarity of the features that are present often outweighs the number of features. Possibly this is why specific numbers of features are rarely specified. Certainly we did not indicate to the

participants in this experiment a specific number in order to make a diagnosis. However, many of our subjects reported adopting a counting strategy. In principle, if a counting rule were adopted, then the familiarity (possibly the 'quality') of the features would have little effect. Perhaps, if we encourage participants to undertake a similar 'counting' strategy, they will be less influenced by familiar feature instantiations.

Experiment 2

A benefit of examining the role of instantiated and informational features within psychiatry is that authoritative counting rules have been systematized. The Diagnostic and Statistical Manual of Mental Disorders (the DSM-IV (American Psychiatric Association, 2000)), is the current gold standard for the diagnosis of mental illness, and is founded on the concept of explicit counting rules - a summation of features and the establishment of a minimum criterion for diagnosis underlies most clinical psychiatric diagnoses. With the adaptation of materials from psychiatry the use of, and reliance on, counting rules can be explicitly investigated and manipulated while maintaining strong ties to real-world categorization tasks. In this experiment, we will examine the role of a tacitly cued counting strategy on the impact of familiar feature instantiations. If participants are cued to 'count' symptoms in a way similar to the minimum criteria counting rules present in some areas of medicine (although not included in the instructions given to the participants in these experiments), this may potentially mediate the impact of familiar feature instantiations.

Methods

Participants

Thirty-six first year undergraduate students participated in the experiment in exchange for course credit. Eighteen students were tested immediately following training, and eighteen returned for testing following a 24-hour delay. This study was approved by the McMaster University Ethics Board, and written consent was obtained from all participants.

Procedure

The experimental design in Experiment 2 is identical to that described in Experiment 1, with the exception of the mode in which participants report the diagnostically relevant features in both the practice and test phases. In Experiment 1, participants reported as few or as many features as they wanted to; no minimum or maximum requirement being set. In addition to the absence of a criterion, participants reported what they believed to be the diagnostically relevant features in one large text field. In Experiment 2, participants reported the diagnostically relevant features in four small text fields, 3 of which had to contain features in order for the participant to continue to the next case. This manipulation was included in order to indirectly cue participants to search for four features, and to encourage a strategy of counting features. It is hypothesized that by attending to individual features, and objectively 'counting'

features, the role of similarity in the diagnostic decisions of novices may be reduced. All other aspects of the experiment were identical.

Analysis

Data were analyzed in a similar manner to that described in Experiment 1.

Results

Diagnostic Probabilities

Mean diagnostic probabilities assigned to the diagnosis supported by the familiar feature instantiations and the plausible alternate, for both testing groups, can be seen in Figure 4.2.

Insert Figure 4.2 about here

Participants assigned significantly higher diagnostic probabilities to the diagnosis supported by the familiar instantiated features than the diagnosis supported by equally valid, but novel feature instantiations (mean diagnostic probability assigned to diagnosis supported by the familiar feature instantiations = 49.7%, mean diagnostic probability assigned to the diagnosis supported by the novel feature instantiations = 42.9%(F (1,30) = 6.97, p = .01)). There was no significant interaction between the impact of familiar

feature instantiations and delay period (whether participants were tested immediately following training, or following a 24-hour delay) (F (1, 30) = 0.01, p = .93).

Self-reported strategy

The mean of the reported strategies was -46, indicating that, on average, participants were reporting more of a counting strategy than reported in Experiment 1. When examining individual participant responses, 24 reported using mainly counting strategies (compared to 16 in Experiment 1), 6 reported using generalizing strategies (compared to 8 in Experiment 1), and 6 reported using combined strategies (compared to 12 in Experiment 1). The shift in proportion of participants who self-reported using counting strategies from Experiment 1 to Experiment 2 is evidence of the effectiveness of the experimental manipulation in Experiment 2. Significantly more participants reported using a counting strategy in Experiment 2 compared to the frequency of a counting strategy in Experiment 1 (proportion of participants reporting mainly a counting strategy in Experiment 1 = .44, proportion of participants reporting mainly a counting strategy in Experiment 2 = .67 (X^2 (1) = 3.60, p = .056)).

Impact of familiar instantiated features on feature identification.

Participants reported more of the familiar feature instantiations (mean of 1.94 features) than the novel feature instantiations (mean of 1.72 features) (F (1, 30) = 7.98, p =.008). The number of features did differ by delay (F (1, 30) = 7.38, p = .01), with participants reporting a larger total number of features on immediate testing (1.90 features) compared to delay testing (1.76 features), and the number of familiar features identified did not interact with delay (F (1, 30) = .01, p = .91).

Between-Experiment analysis

Assessing the effect of cueing a counting strategy on the impact of familiar feature instantiations requires the comparison of the reliance on familiar instantiated features between studies. Additionally, a between study comparison can investigate the role of cued counting strategies on the number and nature of features reported by the participants, providing additional data regarding the effectiveness of the cuing manipulation varied between experiments.

Diagnostic Probability. Impact of familiar instantiated features on diagnostic probabilities assigned was assessed by a post-hoc analysis conducted across Experiments 1 and 2. One might expect that when participants are cued to count diagnostic features, and self-report using this strategy more frequently, that they would be less susceptible to the biasing impact of familiar feature instantiations. A between-study analysis found no significant relation between experimental condition (cued to count as in Experiment 2, or not cued to count in Experiment 1) and the impact of familiar feature instantiations ($F(1,69) = .318, p = .57$).

Impact of familiar instantiated features on feature identification. Participants in Experiment 2 reported significantly more features in total (3.88 out of a total of four possible features) compared to participants in Experiment 1 (3.67 out of a total of four possible features) ($F(1,76) = 4.03, p = .04$), further supporting the cuing impact of four individual fields compared to one open response field, and a minimum response criterion. In addition, participants across both experiments reported more of the familiar feature instantiations (mean of 1.88 features) than the novel feature instantiations (mean of 1.71

features) (F (1, 74) = 11.46, p =.001). In the delay condition, the effect of familiar features is actually in the direction being larger in the group cued to count.

Discussion

Participants assigned significantly more diagnostic probability to the diagnosis supported by the familiar feature instantiations, even when cued to count individual features and to meet minimum diagnostic criteria. Post-experiment reports of strategies demonstrated that participants cued to count features (Experiment 2) were more likely to adopt a counting strategy than their peers (Experiment 1). Participants who were cued to count features report significantly more diagnostic features than those not cued to count features. However, the presence of a cue to count features did not affect the biasing impact of the familiar feature instantiations. Even when participants were paying closer attention to the diagnostic features, and were encouraged to count these diagnostic features, participants were influenced by familiarity.

In addition to higher diagnostic probability assigned to the diagnosis supported by the familiar feature instantiations, participants across both studies were more likely to report familiar feature instantiations as diagnostically relevant than equally valid novel feature instantiations. This could indicate that participants were potentially having difficulty 'translating' the novel instantiation to the informational feature in the language of the diagnostic rule, or noticing the novel instantiations at all. However, Young, Brooks, & Norman (2007) demonstrated that even perfect translators (individuals who consistently and accurately identified all of the diagnostic features in a case vignette, and who could correctly identify the corresponding informational level feature) were

influenced by familiar feature instantiations to the same magnitude as the rest of the study population. When we combine that result with the differential reporting of familiar and novel feature instantiations found in this study, it is likely that familiarity is functioning not only as an influence on the individual weighting a particular feature, as described by previous research (Young, Brooks, & Norman, 2007) but may suggest that familiarity can influence the likelihood of a feature being detected at all.

General Discussion

The studies presented in this paper demonstrate the influence of familiarity on the diagnostic decisions of novices. Participants assigned significantly more diagnostic probability to the disorder supported by familiar feature instantiations. The same pattern of data was seen following a 24-hour delay, and when participants were cued to count the features present in the case vignette. In addition, across both studies, participants were more likely to report the familiar feature instantiations than the equally valid, but novel, feature instantiations.

These results support the role of familiarity of instantiated features in categorization (Brooks & Hannah, 2006), and support a growing body of literature on the role of familiarity on diagnostic decisions in medicine (Brooks & Hannah, 2006; Brooks, Norman, & Allen, 1991; Hatala, Norman, & Brooks, 1999; Kulatunga-Moruzi, Brooks, & Norman, 2001; Young, Brooks, & Norman 2007). However, the purpose of the present design is to isolate the role of instantiations in the use of known, relevant features. Similarity on the feature level – similarity between current and previously experienced

instantiations of informational features – may enable the use of analytic and abstract knowledge. General language use of terms (e.g. "papules" understood as "solid bumps on the skin," as defined in medical dictionaries) allows the use of rules by beginners, but in general must be expanded upon to allow expert performance in applying a rule containing the terms. This separate features design has allowed the demonstration of a preference for relying on familiar instantiations of individual features in categorizing whole cases. Novel manifestations of features are suspect for both learners and experts. Markedly novel manifestations could lead the beginner to hope that their general language knowledge is sufficient to deal with the current case, and may make the expert wonder whether a different disease process is determining or interacting with the suspected disease. That is, even when rules are used, as they evidently are in these novel test cases, the role of similarity is not eliminated and may have an important role in increasing confidence in the diagnosis.

These considerations are also relevant when the learner is trying to induce an explicit rule for natural categories. Given the variety of feature appearance in most natural materials, the learner may try to use the predictive value of some verbal descriptors and simultaneously learn the categorical value of particular instantiations. If the tentative descriptors do not yield sufficient predictive power, then the learner may try different descriptors of the known instantiated features to produce a more accurate rule. Such variability is missing in many artificial concept formation tasks, where the materials have no variation in appearance of individual manifestations of a particular informational feature, so that the feature descriptors are obvious and easily named. Under these

circumstances the re-descriptive or translational process central to the present studies need not occur. Regardless of the other values of such restricted materials, they do not allow a type of learning that is obvious with materials that have more natural variation in appearance. Thus, either starting with explicit prior knowledge or trying to induce explicit rules regarding natural categories (as seen in many inductive concept learning studies), requires experiments and models that represent features at both the information and instantiated level.

It is critical to consider the results of these experiments in the context of learning in a complicated domain. Even with extended practice, extensive feedback, and all-prototype training, individuals are not treating equivalent novel and familiar instantiated features equally. In order to competently use a diagnostic rule, one must know the informational features of the rule, be able to recognize the instantiations of those informational (or diagnostic) features, and continue to rely on previously seen manifestations of features.

In psychiatry, there is an explicit and deliberate attempt to moderate the ambiguity between diagnostic categories through the development and implementation of strict diagnostic rules as presented in the DSM-IV (American Psychiatric Association, 2000). These diagnostic rules are summative in that a minimum number of clinical features must be present to reach a diagnosis. In addition, the DSM-IV is structured so that each clinical feature is of similar weight (i.e. there is no feature that is 'more schizophrenic' than another), and is constructed to be atheoretical, and contain no causal relationships. Research conducted by Kim and Ahn (2002) demonstrate that both experts

and novices have causal theories regarding different forms of psychopathology, and these different theories affect the weighting of individual features present within a case vignette. Features that are considered more central to an individual's causal theory are weighted more heavily, remembered better, and more frequently misremembered as being present in a case than features that are considered to be more causally peripheral. Kim and Ahn (2002) explicitly manipulated the presence of informational features in case vignettes according to their judged causal centrality, however did not investigate the role of instantiated features in addition to causal centrality. However, they do clearly demonstrate similar differential use of features within a diagnostic rule that has been constructed to ensure each feature is equally probative. The experiments presented in this paper show a similar re-weighting of objectively equal features, even when participants are implicitly cued with a counting rule (as is encouraged, if not directly specified) by the DSM IV, and examined the mediating role of these counting rules on the performance of novice diagnosticians. The role of explicit counting rules was examined and discussed in Experiment 2, however further research is needed to complete our understanding on the use, and potential misuse, of explicit counting rules.

We acknowledge that there are some limitations to the research presented in this paper. The use of indirectly cued counting rules does not perfectly generalize to the use of strict counting rules in medicine, but this remains a future direction for research. Additionally, the materials for these studies were specifically developed to reduce or minimize between and within category ambiguity, which may also reduce the generalizability of these findings to the 'real world' of medical diagnosis. However, we

remain confident that the data presented in this paper indicate a consistent role of familiarity, and of counting rules on the diagnostic decisions of novices.

Medical decision-making and the processes of diagnosis have received much attention in the medical literature (e.g. Bordage & Lemieux, 1991; Croskerry, 2002; Groves, O'Rourke, & Alexander, 2003; Gruppen, Wolf, & Billi, 1991; Norman, Young, & Brooks, 2007; Redelmeier, 2005), in conjunction with concern over medical errors (e.g. Cosby, 2003; Croskerry, 2003; Graber, Gordon, & Franklin, 2002; Kempainen, Migeon, & Wolf, 2003; Wears & Nameth, 2007). Medicine provides not only a domain of rich stimulus variability to examine complex categorization, but the practical implications of furthering our understanding of clinical reasoning are also clear (e.g. the development of physician support systems to reduce medical error (Berner, Maisiak, Heuderbert, & Young, 2003; Berner, Webster, Shugerman, Jackson, Algina, Baker, Ball, Cobbs *et al*, 1994; Friedman, Elstein, Wolf, Murphy, Franz, Heckerling *et al*, 1999; Hynniman, Conrad, Urch, Rudnick, & Parker, 1970; Means, Derewicz, & Lamy, 1975)). From the results presented and discussed in this paper, we propose that good knowledge of the diagnostic rules is insufficient to be able to categorize diagnostically relevant features. Additionally, we demonstrate a clear role for similarity and familiarity *within* the application of a formal diagnostic rule. Finally, these results speak to the importance of distinguishing between the informational and instantiated levels of features, and the role of familiarity in research on categorization and medical diagnosis. Without recognizing and incorporating the naturalistic variability in feature presentation into our laboratory studies, we may be encouraging an incomplete account of categorization and

decision making. These considerations are especially important for adapted teaching methods to ensure competence and expertise in both understanding and implementing diagnostic rules.

References

Allen, S. W. & Brooks L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3-19.

American Psychiatric Association. (2000). Diagnostic and Statistical Manual for Mental Disorders: Fourth Edition, Text Revision. *American Psychiatric Association Press,* Washington.

Bordage, G. & Lemieux M. (1991). Semantic structures and diagnostic thinking. *Academic Medicine, 66,* s70-72.

Brooks, L. R. & Hannah, S.D. (2006) Instantiated features and the use of "Rules". *Journal of Experimental Psychology: General, 132,* 133-151.

Brooks, L. R., LeBlanc, V. R. & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychological Science, 11*, 112-117.

Brooks, L.R., Norman, G.R. & Allen. S.W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278-87.

Cosby, K. S. (2003). A framework for classifying factors that contribute to error in the Emergency Department. *Annals of Emergency Medicine, 42,* 815.

Croskerry, P. (2002). Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Academic Emergency Medicine, 9,* 1184-1204.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78,* 16.

Dore, K., Weaver, B. & Norman, G. (2005). Does a rose by any other name smell as

sweet? The importance of variation in case presentation to diagnostic reasoning. Talk presented at the meeting for Research in Medical Education, Washington DC.

Graber, M., Gordon, R. & Franklin N. (2002). Reducing diagnostic error in medicine: what is the goal? *Academic Medicine, 77,* 981-992.

Groves, M., O'Rourke, P. & Alexander, H. (2003). Clinical Reasoning: the relative contribution of identification, interpretation, and hypothesis errors to misdiagnosis. *Medical Teacher, 25,* 621-625.

Gruppen, L. D., Wolf, F. M. & Billi, J. E. (1991). Information gathering and integration as sources of error in diagnostic decision making. *Medical Decision Making, 11,* 223-239.

Hatala R., Norman G. R. & Brooks, L. R. (1999). Research basic to medical education: influence of a single example on subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine, 11,* 110-117.

Hynniman, C. E., Conrad, W. F., Urch, W. A., Rudnick, B. R. & Parker, P. F. (1970). A comparison of medication errors under the University of Kentucky unit dose system and traditional drug distribution system in four hospitals. *American Journal of Hospital Pharmacology, 27,* 802-814.

Kempainen, R. R., Migeon, M. B. & Wolf, F. M. (2003). Understanding our mistakes: a primer on errors in clinical reasoning. *Medical Teacher, 25,* 177-181.

Keil F. C. (1991). The emergence of theoretical beliefs as constraints on concepts. In S. Carey & R. Gelman (Eds.) *The epigenesis of mind: Essays on biology and cogntion* (237-256). Lawrence Erlbaum Associates; Philadelphia, PA.

Kim, N. S. & Ahn, W. (2002). Clinical psychologists' theory based representations of mental disorders predicts their diagnostic reasoning and memory. *Journal of Experimental Psychology: General 131,* 451-476.

Kulatunga-Moruzi,C., Brooks, L.R. & Norman, G.R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine, 13*, 110-6.

Lacroix, G. L., Giuere, G. & Larochelle, S. (2005). The origin of exemplar effects in rule-driven categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 272-288.

Means, B. J., Derewicz, H. J. & Lamy, P. P. (1975). Medication errors in a multidose and a computer based unit dose drug distribution system. *American Journal of Hospital Pharmacology, 32,* 186-191.

Norman, G. R., Young, M. E. & Brooks, L. R. (2007). Non-analytic models of clinical reasoning: the role of experience. *Medical Education, 41,* 1140-1145.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioural and Brain Sciences, 28,* 1-49.

Redelmeier, D. A. (2005). The cognitive psychology of missed diagnosis. *Annals of Internal Medicine, 142,* 115-120.

Regehr, G. & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: the priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92-114.

Schmidt, H. G., Norman, G. R. & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: theory and implication. *Academic Medicine, 65,* 611-21.

Thibaut, J. – P. & Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 1403-1415.

Wears, R. L. & Nemeth, C.P. (2007). Replacing hindsight with insight: towards better understanding of diagnostic failures. *Annals of Emergency Medicine, 49,* 206-209.

Woods, N. N., Brooks, L. R. & Norman, G. R. (2005). The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Medical Education, 39,* 107-112.

Young, M. E., Brooks, L. R. & Norman, G. R. (2007). Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Medical Education, 41,* 1146-1151.

Zarin, D.A. & Earls, F. (1993). Diagnostic decision making in Psychiatry. *American Journal of Psychiatry, 150*, 197-206.

Table 4.1

Table 4.1: Example of Diagnostic criteria for Mania (one of the four modified diagnostic categories), and one example of the informational features in their instantiated form.

Diagnostic criteria	Instantiations of diagnostic rules
Increased energy	She says that she becomes really 'hyper' and that she can't ever sit still
Decreased need for sleep	She often works up to 80 hours a week, with only short naps between shifts.
Inflated self esteem	She believes that she will be accepted for the space program, even though she has not passed high school.
More talkative	Her mother reports she has recently been spending half of her life on the phone and chatting on MSN.

Figure Captions

Figure 4.1: Mean diagnostic percentages assigned to the diagnosis supported by the familiar instantiated features and the diagnosis supported by the novel instantiated features. Both the immediate and delay test (following a 24-hour delay) groups are depicted, and error bars represent the standard error of the mean.

Figure 4.2: Mean diagnostic percentages reported for participants cued to count symptoms in Experiment 2. Mean probabilities are displayed for the diagnosis supported by the familiar feature instantiations and the diagnosis supported by the novel feature instantiations. Data for both the immediate and delay test groups are shown, and error bars represent the standard error of the mean.

Figure 4.1

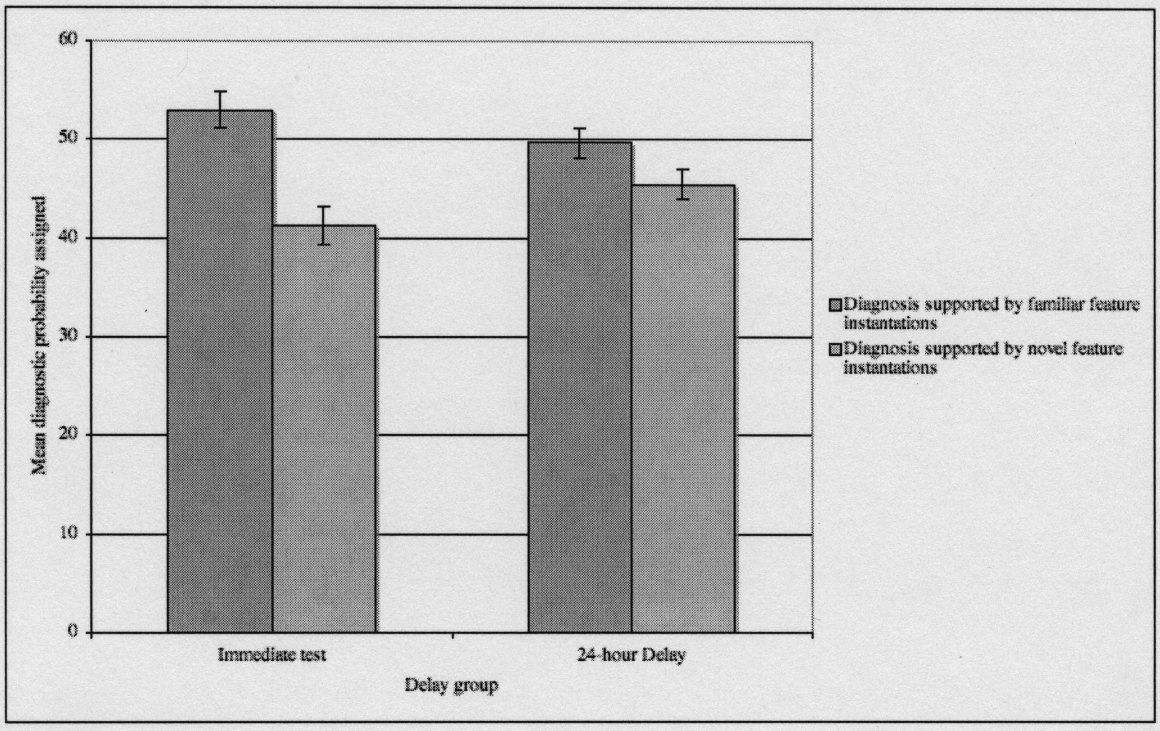
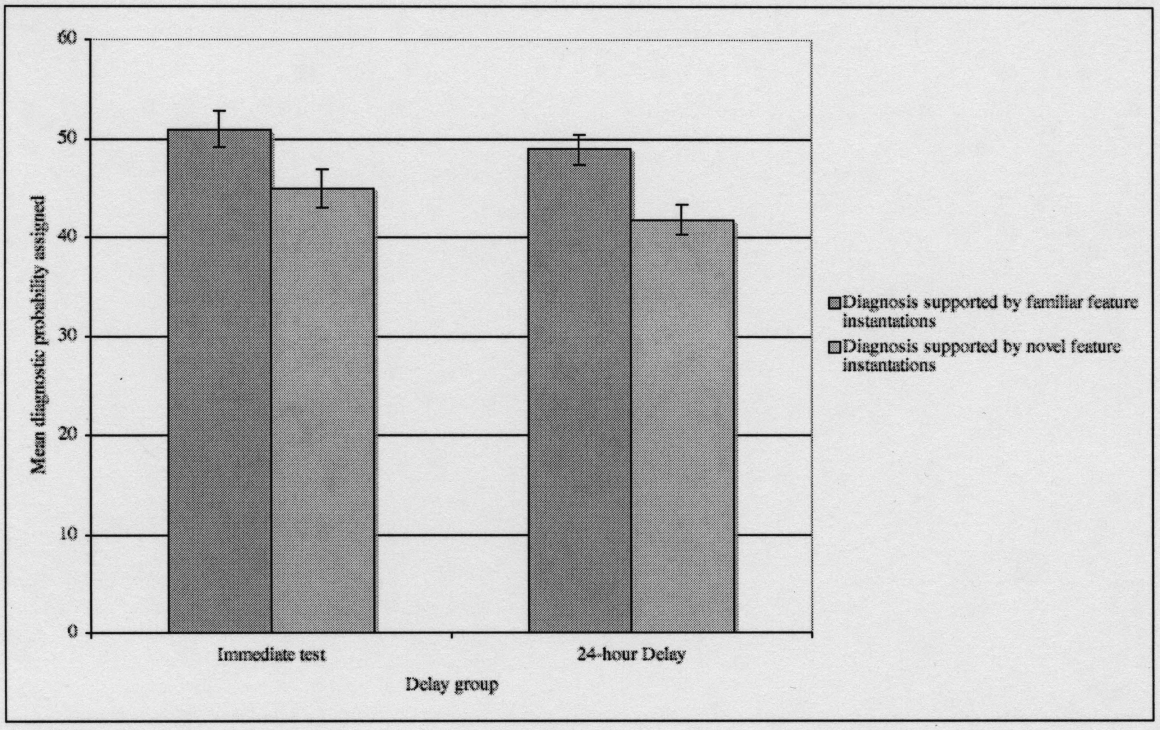

60
50
40
30
20
10
0
Mean diagnostic probability assigned
Immediate test
24-hour Delay
Delay group
Diagnosis supported by familiar feature instantations
Diagnosis supported by novel feature instantations

Figure 4.2


60
50
40
30
20
10
0
Mean diagnostic probability assigned
Immediate test
24-hour Delay
Delay group
Diagnosis supported by familiar feature instantations
Diagnosis supported by novel feature instantations

Chapter 5:

Impact of a single case: the influence of familiar non-diagnostic information on the diagnostic decisions of novices

This paper is in preparation for submission to: Medical Education.

Impact of a single case: the influence of familiar non-diagnostic information on the diagnostic decisions of novices

Meredith E. Young[1], Lee R. Brooks[1], Geoffrey R. Norman[2]

[1]Department of Psychology, Neuroscience and Behaviour, McMaster University

[2]Department of Clinical Epidemiology and Biostatistics, McMaster University

Abstract

OBJECTIVES: Previous research has demonstrated the role of familiar symptom descriptions, and of entire case similarity, but here we explore the influence of familiarity of non-diagnostic patient information on the diagnostic decisions of novices. If an instance model has strong explanatory power within clinical reasoning, we should see an influence of familiar patient identity information (e.g. name and age) on later cases matched on similar identifying characteristics even though such information is objectively irrelevant.

METHODS: In this experiment, 36 participants (undergraduate psychology students) were trained to competence on four simplified psychiatric diagnoses, and allowed to practice their diagnostic skills on 12 prototypical case vignettes, with feedback provided. One third of participants were tested immediately, one third following a 24-hour delay, and one third following a one-week delay. Test cases were created to be equiprobable, with both possible diagnoses supported by two novel symptom descriptions. However, one diagnosis was also supported by identity information similar to a patient seen in the training phase. A deviation from a 50/50 split in probability demonstrates a reliance on the familiar, non-diagnostic information.

RESULTS: Participants assigned significantly higher diagnostic probability to the diagnosis supported by the familiar patient information (52.6%) than to the plausible alternate diagnosis (38.9%). Participants also reported more clinically relevant features to support the cued diagnosis than to the plausible alternate diagnosis. The influence of

familiar patient identity information was consistent across delay periods, and cannot be well accounted for by forgetting of the diagnostic rules.

CONCLUSIONS: Participants were clearly relying on familiar patient identity information, and it influenced their diagnostic conclusions. These results support an exemplar or instance model of reasoning which is not limited by whole case similarity, or similarity of diagnostic information.

Impact of a single case: the influence of familiar non-diagnostic information on the diagnostic decisions of novices

When acquiring expertise in a complicated domain such as medicine it is generally accepted that a student needs both formal knowledge and experience. We would not trust a clinician who never attended class, but we would also be wary of one who had never been in a clinic with patients. There is something to be gained from each type of learning – both formal rules and experience. In order to explain the benefits of both of these types of knowledge, clinical reasoning has been studied from multiple perspectives, identifying many cognitive strategies that may be called upon in different situations. Evidence has accrued for the use of basic science knowledge (Woods, Brooks, & Norman, 2005), consideration of base rates or Bayesian reasoning (e.g. Elieson & Papa, 1994), and a strong role for similarity to previously seen examples (see Norman, Young, & Brooks, 2007 for a review). This paper will further examine the role of similarity to previously seen examples, specifically investigating the influence of familiar non-diagnostic information on the diagnostic decisions of novices.

The use of similarity-based judgments has been demonstrated in several domains of medicine, and in several roles. Whole case similarity, and visual case similarity, has been demonstrated in expert and intermediate physicians in the case of dermatology (Brooks, Norman, & Allen, 1991; Kulatunga-Moruzi, Brooks, & Norman, 2004). This demonstration of the reliance on similarity in a relatively expert population supports the notion of exemplars in decision making – that we maintain information about previous cases in memory, and use these memories to aid our current decision making. Distinct

from whole case similarity, we can also see the reliance on similarity of individual features embedded within the context of a strong diagnostic rule; novice diagnosticians rely more heavily (Dore, Weaver, & Norman, 2005; Young, Brooks, & Norman, 2007), or differentially report (Young, Norman, & Brooks, *unpublished manuscript*), familiar manifestations of clinical features. More specifically, participants assign more diagnostic weight to a symptom description that they have encountered before, and are more likely to diagnose the fictional patient with the diagnosis supported by the familiar feature description (Dore, Weaver, & Norman, 2005; Young, Brooks, & Norman, 2007). This functional role of similarity may be what helps an early learner to understand how to apply a diagnostic rule, and may be a very adaptive use of previous experience.

In a larger theoretical context drawn from cognitive psychology, an exemplar model generally proposes that prior experience is maintained in memory in the form of processing episodes and is rapidly recalled when a similar case is encountered again (Brooks, 1978; 1990). The exemplar (or instance) model, proposes that current decisions are aided by a rapid pattern matching to previously encountered cases, and this similarity match can occur on any stimulus dimension; little distinction is made between encoding critically relevant and irrelevant information (Brooks 1978; 1990). However, anyone applying a diagnostic rule makes a clear distinction between what is relevant and what is not – it is the function of a rule to point to what is important in a diagnostic case. If an exemplar model is indeed informative in clinical reasoning, then non-diagnostic information should be able to elicit a similar previous experience, despite the fact that the retrieval cues are objectively irrelevant to the diagnosis. Previous research has

demonstrated that familiar patient identity information (e.g. 52 year-old banker) can bias decisions in medical residents doing ECG interpretation – when familiar clinical scenario information supports the correct diagnosis performance across all levels of expertise is increased, but when familiar clinical scenario information supports an incorrect diagnosis, performance suffers (Hatala, Norman, & Brooks, 1999a; 1999b).

In examining literature from cognitive psychology, the role of non-diagnostic information within non-medical decision tasks is mixed. With the increased presence of non-diagnostic information, some decisions are rendered more extreme (a classic example of this is the representativeness heuristic where the presence of non-diagnostic information can increase the likelihood of a stereotypical judgment of an individual – e.g. Tversky & Kahneman, 1974), and some are rendered less extreme (the dilution effect where the presence of non-diagnostic information can decrease the likelihood of a stereotypical judgment of an individual – e.g. Nisbett, Zukier, & Lemley, 1981) than without the non-diagnostic information. However, a more recent investigation into the role of irrelevant information seems to be able to explain these mixed results; if the non-diagnostic information can be interpreted (albeit it incorrectly) as relevant to the decision at hand, it results in more extreme judgments, if it can be interpreted as irrelevant, it results in less extreme judgments (LaBella & Koehler, 2004).

The research presented here was conducted to investigate the role of familiar non-diagnostic information, presented in the form of familiar patient identity, on the diagnostic decisions of novices. As patient identity is rarely diagnostically informative, we should not expect to see reliance on a similar previous case when matched on an

irrelevant dimension. However, previous research indicates that non-diagnostic information can influence the diagnostic decisions of young clinicians (Hatala, Norman, & Brooks, 1999a; 1999b). When examining research in medical education, it is possible that previous research investigating the role of familiar patient identity in ECG interpretation may have inadvertently used what may be stereotypical profiles of cardiac patients (e.g. relatively young males in high stress 'mental' jobs – e.g. Clarke, 1992) to bias participant responses (Hatala, Norman, & Brooks, 1999a; 1999b). One might then conclude that the impact of patient identity could be driven by a representativeness heuristic (Tversky & Kahneman, 1974), rather than a reliance on a previously seen case. Further, Hatala *et al* (1999a: 1999b) examined a specially educated population including medical students, medical residents, and established physicians. It could be possible that the biasing impact of familiar patient identity information is limited to populations with some specialized medical knowledge, suggesting that the exemplar model of reasoning may come about as expertise develops rather than a fundamental component of decision making. Previous research has demonstrated support for exemplar based reasoning within novice populations (Dore, Weaver, & Norman, 2005; Young, Brooks, & Norman, 2007) but these demonstrations have thus far been limited to clinically relevant information, which may reflect a learning strategy for recognizing manifestations of a diagnostic rule, rather than an automatic memory retrieval of an entire previously encountered case. Finally, the studies conducted by Hatala *et al* (1999a; 1999b) used nearly identical age and occupation descriptions (matched patients had identical jobs and ages); so the generalizatility of the findings to everyday clinical contexts may be limited.

To further earlier investigations, we carefully manipulated aspects of patient information to maintain similarity but not identity, to ensure that these data were not relevant to the diagnosis, and did not conform to any popular stereotypes or base rate probabilities regarding the specific disorders used (e.g. if a disorder was popularly believed to be more present among males, female patient identities were used). We hypothesized that, because of the role of non-analytic, exemplar based reasoning, the clinically irrelevant, or non-diagnostic information would bias participants towards the diagnosis that supports that sense of familiarity. More specifically, when faced with equiprobable diagnoses, we hypothesized that participants would assign significantly higher diagnostic probability to the disorder supported by familiar patient characteristics (age, name, marital status, employment, etc) consistent with a previous training case.

Methods

Participants

Thirty-six students enrolled in a first year undergraduate psychology course participated for course credit. Informed consent was obtained from all participants, and this study was approved by the McMaster University Research Ethics Board.

Stimuli

Four pseudo-psychiatric disorders were created for the purposes of this research. Similar methods and materials have been used before (see Young, Brooks, & Norman, 2007) to examine the role of familiarity within the application and use of a diagnostic rule. The disorders were drawn from disparate areas of psychiatry and included mania,

obsessive-compulsive disorder, schizophrenia and paranoid personality disorder. The rules were simplified versions of the rules found in the Diagnostic and Statistical Manual (American Psychiatric Association, 2000; see Table 5.1 for modified medical rules used in this experiment) to ease learning. Participants were told that these disorders were fictional, and were not representative of the real world disorders of the same name, so to rely only on the material presented in the experiment. This instruction was included to ensure that participants would be less likely to rely on medical stereotypes, if any were inadvertently included in the patient description information.

Insert Table 5.1 about here

Each disorder was characterized by four unique, non-overlapping, characteristic symptoms. During practice and test phases of this experiment, each symptom was presented using a patient voice, or unique symptom description (e.g. the rule feature 'concern over fidelity of their spouse' would be presented as 'she has been worried ever since her husband started a new job with an old girlfriend, so she has been secretly checking his email regularly'). Each symptom description was unique and participants received feedback during training regarding the appropriate 'translation' of the symptom description. During practice and test phases of the experiment, stimuli were presented as case vignettes, with personal identifying information (e.g. name, age, type of employment, familial situation) and unique symptom descriptions. Participants were

asked to report their diagnosis in the form of percentages assigned to each of the four diagnostic options. Participants were instructed to distribute the diagnostic percentages as they saw fit, with the only restriction being that they must sum to 100%. Following the presentation of a written case vignette, and the assignment of diagnostic percentages, the case was removed and participants were asked to report the clinically relevant features that were present in the case.

During the practice phase of the experiment, participants were shown prototypical written case vignettes. Cases were defined as prototypical as they contained all four features of the diagnostic rule for that disorder, presented as unique symptom descriptions. Each case vignette during the test phase included two symptom descriptions supporting one diagnosis, and two symptom descriptions supporting an alternate diagnosis, creating a case with two equiprobable diagnoses. Additionally, each case contained a similar patient description to one seen in practice. Similarity was defined along 5 dimensions, and the values for each dimension were similar, but never identical. Similar patients had a name starting with the same phoneme, were within 5 years in age (but were always within the same decade), had similar but never identical professions, had similar but never identical familial situations including martial status and number of children, and lived in similar areas surrounding McMaster University. An example of similar, but not identical, patients can be seen in Table 5.2.

Insert Table 5.2 about here

For example, if a participant first met Allen as an individual diagnosed with schizophrenia during training, Albert's case vignette would include two symptom descriptions that would support a diagnosis of schizophrenia, and two symptom descriptions that would support another diagnosis, for example Obsessive Compulsive Disorder. If participants were influenced by the familiarity of the patient identity information, then we expect a deviation from a 50/50 split in diagnostic probabilities in favor of the diagnosis supported by the familiar patient identity.

Procedure

This experiment was programmed and presented using RunTime Revolution 2.1 (RunTime Revolution Ltd. Edinburgh, Scotland) and presented on a computer screen. The experiment included a learning phase, a practice phase and a test phase. Twelve participants were tested immediately following training, twelve 24-hours after training, and twelve one week after training.

Learning phase. The learning protocol was identical for each disorder, but order of learning was randomized across participants. Participants were given a list of the four symptoms characteristic of the disorder to study. Participants were asked to correctly identify the four features characteristic of the disorder from the full list of 16 symptoms. If participants were unable to do this, they were asked to review the symptom list and re-take the quiz. If they could correctly identify the appropriate features in the diagnostic rule, they saw an initial prototypical case vignette and were asked to identify the features present in the patient vignette. This protocol was repeated for each of the four pseudopsychiatric disorders. Before participants moved to the practice phase, they were

again required to identify the four correct features for each disorder. A pass criterion of 15/16 was set, and if participants were unable to meet this criterion they returned to the beginning of the learning phase. If participants passed this strict learning criterion, included to ensure competency, they progressed to the practice phase of the experiment.

Practice phase. The practice phase included 12 prototypical case vignettes, presented in random order. Each case contained patient identity information (name, age, area of residence, employment and familial situation, etc) and four unique symptom descriptions indicative of a single diagnosis. There were three practice cases for each of the four disorders, and participants were asked to assign diagnostic probabilities to each of the four possible diagnoses and report the relevant features in the case vignette. After participants reported the clinically relevant features in the case, they received feedback regarding not only the correct diagnosis, but on the relevant features that were present in the case. The practice phase in this experiment filled two purposes: firstly, it allowed participants to hone their diagnostic abilities of the pseudo-psychiatric disorders, but secondly it was designed to expose participants to a variety of unique patients that functioned as familiar patients in the test phase.

Test phase. Participants were randomly assigned to either the immediate test group (who went directly from the practice phase to the test phase), the 24 hour delay group (who returned to the lab after 24-hours after the practice phase to complete the test phase), or the one week delay group (who returned to the lab 1 week after the practice phase to complete the test phase). A total of 12 unique case vignettes were presented, and order of case presentation was randomized.

Each test case was constructed to contain two equally probable diagnoses – two novel ways of describing clinical symptoms would support the same disorder as the familiar patient identity, and two novel symptom descriptions would support a competing diagnosis. After participants assigned diagnostic probabilities, the case vignette was removed from the screen, and they were asked to report the clinically relevant information that was in the case.

Post-test assessment. Following the test phase of the experiment, participants were tested on their memory for the diagnostic rule. They were asked to generate the four features characteristic of each disorder in a free recall task. This addition to the experimental protocol was only implemented for those participants tested following a 24-hour or 1-week delay. It was added for these groups of participants to ensure that they could indeed recall the diagnostic rule following a substantial delay period. However, a similar study was conducted previously and an independent group of participants (n = 19) scored approximately 95% accuracy for the post-test assessment of memory for the diagnostic rule when testing took place immediately after the practice phase.

Analysis

Diagnostic probabilities. Mean diagnostic probabilities for the disorder supported by the familiar patient identity and the plausible alternate diagnosis were computed. Diagnostic probabilities assigned were analyzed using a mixed design analysis of variance (ANOVA) where mean diagnostic probability assigned (mean diagnostic probability assigned the diagnosis supported by the familiar patient identity) was the

within subject comparison of interest, and the between subjects factor was delay (immediate test, 24-hour delay, 1-week delay).

Analysis of clinically relevant features identified. Participants were asked to report the clinically relevant features present in each case following the assignment of diagnostic probabilities. Participants' responses were coded by an independent research assistant blinded to the individual biasing manipulation. The features reported by participants were coded as: feature identified in support of the diagnosis supported by the familiar patient identity, feature identified in support of the plausible alternate diagnosis, and other. The number of features identified congruent with the biasing manipulation and the plausible alternate were averaged across the 12 cases to create the mean number of each type of feature identified by the participants. To examine whether participants were more likely to report features congruent with the familiar patient identity than the equiprobable alternate diagnosis, a mixed design ANOVA was conducted, where the type of diagnosis supported by the features identified (features identified in favor of the diagnosis supported by the familiar patient identity versus number of features identified in support of the plausible alternate diagnosis) was the within-subjects factor of interest, and the delay (immediate testing, 24-hour delay, or 1 week delay testing) was the between subjects factor of interest.

Results

Diagnostic probabilities

Mean diagnostic probabilities assigned for the immediate test group, 24-hour delay and 1-week delay groups can be seen in Figure 1.

Insert Figure 5.1 about here

Participants assigned significantly higher diagnostic probabilities to the disorder supported by the familiar patient identity (mean diagnostic probability assigned to disorder supported by familiar patient identity = 52.6 %, SEM = .69; mean diagnostic probability assigned to the plausible alternate diagnosis = 38.9 %, SEM = 1.14. F (1, 36) = 71.40, $p < .001$). A test group by familiarity interaction was also found, with the strength of the familiarity effect increasing across delay period, as seen in Figure 1 (F (2, 36) = 3.40, $p = .04$).

Identification of clinically relevant features

Participants reported significantly more features for the diagnosis supported by the familiar patient identity than the plausible alternate (mean number of features identified for the diagnosis congruent with the familiar patient identity = 1.87, mean number of features identified for the plausible alternate diagnosis = 1.72, F (1, 34) = 6.55, $p = .015$). The number of features identified for the diagnosis supported by the familiar patient identity or the plausible alternate diagnosis did not interact with delay (F (2, 34) = .56, $p = .64$).

Post-test recall for the diagnostic rules

Participant responses to the free recall of diagnostic rules was coded and performance was compared across delay periods (24-hour and 1-week). No significant difference in performance was found between the two delay conditions (mean performance for 24-hour delay = 88.5% correct, mean performance for 1-week delay = 87.5% correct, $t(22) = 0.24$, $p = 0.81$). A comparison population of undergraduate students (n = 24) were tested immediately following similar training for the purposes of another experiment, and performance on post-test recall for the diagnostic rule was approximately 95% accuracy. The comparison population did perform significantly better than the delay population ($t(29) = 2.22$, $p = 0.02$).

Discussion

Novice diagnosticians assigned significantly more diagnostic probability to the disorder supported by the familiar patient identity than the equally valid alternate diagnosis. In essence, if a case vignette included a patient who had a similar name, similar occupation, similar age, and similar familial situation to one encountered before, participants were more likely to consider them to have a similar diagnosis. This pattern of response was consistent across participants who were tested immediately following training, a 24-hour delay, or a one-week delay, and the reliance on non-diagnostic information appears to increase across delay periods. The increased impact of similar non-diagnostic information is unlikely to be accounted for by participants' forgetting the diagnostic rule - as performance on post-test assessment, although it decreased with

delay, remained near 90%. Additionally, as each feature mentioned in the rules had equal probability of occurring in support of the diagnosis cued by the familiar patient identity and the plausible alternate diagnosis, small decrements in post-test recall for the diagnostic rule cannot fully account for the influence of familiar patient identity across delay periods.

The data presented here further support for an exemplar theory of clinical reasoning. The irrelevant patient information used in this study was non-stereotypical, so the pattern of results found here are unlikely to be explained by the use of a representativeness heuristic (Tversky & Kahneman, 1974). The fact that the effect of exemplars was based entirely on irrelevant and non-stereotypical factors, such as patient name, strongly supports the claim of a qualitatively different thinking process that is not restrained by diagnostically relevant information. The data presented here suggest that diagnostically irrelevant information can function as a cue to retrieve a similar previous case, and therefore influence current decision making. This research also indicates that the influence of a single case can be relatively long lasting, at least up to one week, suggesting that exemplar based effects previously demonstrated are not necessarily transient (Hatala, Norman, & Brooks, 1999a; 1999b; Norman, Young, Brooks, 2007 for a review). Further, care in designing the similar but not identical patient identity information allows for an experimental demonstration of diagnostically relevant information and an initial documentation of what can be counted as 'similar' in a carefully controlled experimental setting.

Participants also reported significantly more symptoms in support of the disorder associated with the familiar patient identity than the plausible alternate diagnosis, although the numerical difference was small. The increased reporting of features that support the diagnosis cued by the familiar patient identity information may indicate an interactive influence of diagnostic and non-diagnostic information, or more analytical (rule based) and more exemplar-like (familiarity based) processes. While this remains speculative, it may suggest that participants are rapidly recalling a previous instance, gaining a sense of fluency (e.g. Kelley & Jacoby, 1989) or familiarity for a particular diagnosis, and relying on supporting information for that diagnosis – in essence a possible demonstration of confirmation bias (e.g. Bower, 1978). Although this conclusion remains speculative, the data presented here support the influence of non-diagnostic information on the reporting of clinically relevant features.

This study is not without limitations. In order to study single case effects, and the role of specific previous experience, one must know what participants have encountered before. With clinicians, this range of experience can be idiosyncratic, which is a unique challenge to this type of research. As a solution to this problem we can use novices, teach them simple medical rules, and have control over their particular set of previous case encounters. The disadvantage, however, is that by studying exemplar effects in novices, we can only infer that the phenomena demonstrated in this paper are seen in more expert populations. However, a growing body of literature suggests that exemplar-based effects can be seen in more expert populations (Allen, Norman, & Brooks, 1992; Brooks, Norman, & Allen, 1991; Hatala, Norman, & Brooks, 1999a; 1999b; Kulatunga-Moruzi,

Brooks, & Norman, 2004), and are sometimes more prevalent with increased levels of expertise (Hatala, Norman, & Brooks, 1999a). Additionally, static case presentation is not perfectly analogous to clinical encounters, but recent research may indicate that exemplar based effects of non-diagnostic information can be found in more dynamic patient-physician encounters (Young, Brooks, Hanson, Bossio, & Norman, *unpublished manuscript*). However, this remains a direction for future research.

This research has demonstrated the role of non-diagnostic, or diagnostically irrelevant, information on the diagnostic decisions of novices using a task that has strong rules and decision criteria, and so should be an 'analytic' task. Additionally, participants in this study were more likely to report diagnostic features that supported the familiar patient identity than the alternate diagnosis present in the case vignette – a demonstration that non-diagnostic information influenced the likelihood that a diagnostic feature is mentioned, or perhaps even detected. Previous research has demonstrated a role of similarity within the application of a diagnostic rule (Dore *et al,* 2005; Young, Brooks, & Norman, 2007), and in combination with the current finding, it is possible that the extent to which analytic and non-analytic processes operate independently is less clear. Instead, they may represent interdependent or interacting processes, rather than exclusionary or opposing mechanisms (Pothos, 2005). If an interactive account is appropriate, it could explain why the impact of similarity is observed not only in the influence of non-diagnostic information, but the application of diagnostic rules, and perhaps in detecting diagnostically relevant features. While this study cannot speak directly to the time course or nature of the interdependent or interactive relationship between rule-based and

exemplar-based reasoning, the research presented here suggests that matching to prior exemplars (here constructed to be similar on non-diagnostic information) can influence the probability of reporting clinically relevant features (rule-based information), and the influence remains following a relatively long delay period. Finally, this research provides evidence of exemplar or instance-based reasoning in novices, and suggests that reliance on exemplars is not dependent on large amounts of expertise, but rather may reflect a more general model of clinical reasoning and decision making.

References

Allen, S. W., Norman, G. R. & Brooks, L. R. (1992). Experimental studies of learning dermatological diagnosis: The impact of examples. *Teaching and Learning in Medicine, 4*, 35-44.

American Psychiatric Association. (2000). Diagnostic and Statistical Manual for Mental Disorders: Fourth Edition, Text Revision. *American Psychiatric Association Press,* Washington.

Bower, G. H. (1978). Contacts of cognitive psychology with social learning theory. *Cognitive Therapy and Research, 2*, 123-146.

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In: Rosch & Lloyd, eds. Cognition and Categorization. *Lawrence Erlbaum*, Hillsdale, NJ.

Brooks, L. R. (1990). Concept formation and particularizing learning. In: Hanson, ed. Information, Language and Cognition. *UBC Press*, Vancouver.

Brooks, L.R., Norman, G.R. & Allen. S.W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278-87.

Clarke, J. N. (1992). Cancer, Heart Disease, and AIDS: What do the media tell us about these diseases? *Health Communication, 4*, 105-120.

Dore, K., Weaver, B. & Norman, G. (2005). Does a rose by any other name smell as sweet? The importance of variation in case presentation to diagnostic reasoning. Talk presented at the meeting for Research in Medical Education, Washington

DC.

Elieson, S.W. & Pappa, F.J. (1994). The effects of various knowledge formats on diagnostic reasoning. *Academic Medicine, 69*, 81-3.

Hatala R., Norman G. R. & Brooks, L. R. (1999a). Research basic to medical education: influence of a single example on subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine, 11,* 110-117.

Hatala R., Norman G. R. & Brooks, L. R. (1999b). Impact of clinical scenario on accuracy of electrocardiogram interpretation. *Journal of General Internal Medicine, 14,* 126-129.

Kelley, C.M. & Jacoby, L.L. (1996). Memory attributions: Remembering, knowing and feeling of knowing. In L.M. Reder (Ed.), Implicit memory and metacognition (pp.287-307). Mahwah, NJ: Erlbaum.

Kulatunga-Moruzi, C., Brooks, L.R. & Norman, G.R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine, 13*, 110-6.

LaBella, C. & Koehler, D.J. (2004). Dilution and confirmation of probability judgments based on nondiagnostic evidence. *Memory & Cognition, 32,* 1076-1089.

Nisbett, R. E., Zukier, H., Lemley, R.E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology, 13,* 248-277.

Norman, G. R., Young, M. E. & Brooks, L. R. (2007). Non-analytic models of clinical reasoning: the role of experience. *Medical Education, 41,* 1140-1145.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioural and Brain Sciences, 28,* 1-49.

Tversky, A., Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124-1131.

Woods, N. N., Brooks, L. R. & Norman, G. R. (2005). The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Medical Education, 39,* 107-112.

Young, M. E., Brooks, L.R., Hanson, M., Bossio, J., & Norman, G.R. (2009). Non-analytic reasoning in medicine: Evidence from clinical encounters with standardized patients. *Unpublished manuscript.*

Young, M. E., Brooks, L. R. & Norman, G. R. (2007). Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Medical Education, 41,* 1146-1151.

Young, M.E., Norman, G.R., Brooks, L.R. (2009). Preference for familiar instantiations: Using rules in medicine. *Unpublished manuscript.*

Table 5.1

Table 5.1: Adapted medical diagnoses and diagnostic rules used in this study.

Diagnostic Category	Diagnostic rule
Mania	Decreased need for sleep
	Increased energy
	Inflated self esteem
	More talkative
Schizophrenia	Hallucinations
	Delusions
	Disorganized speech
	Disorganized behaviour
Paranoid Personality Disorder	Concerns over fidelity of their spouse
	Distrustful of others
	Reads hidden meaning into daily events
	Reluctant to confide in others
Obsessive Compulsive Disorder	Repetitive behaviours
	Recurring thoughts
	Difficulty ignoring thoughts
	Interferes with day to day life

Table 5.2

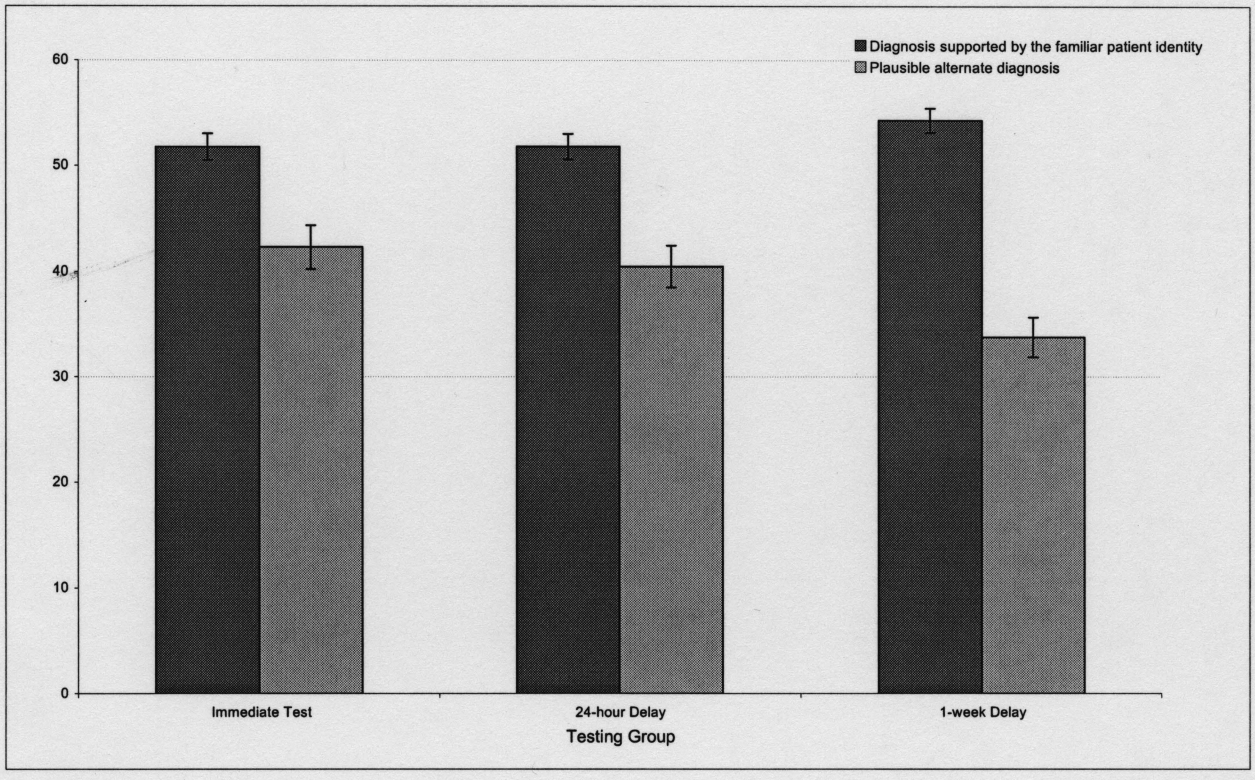
Table 5.2: Sample of familiar patient identity pairs. Patients were constructed to have similar, but never identical ages, names, familial situations, professions and areas of residence.

Name	**Age**	**Occupation**	**Familial situation**	**Other (including area of residence)**
Allen	65	Accountant nearing retirement	Four children, each have a minimum of 2 children each	Caregiver for his wife who has Alzheimer's disease
Albert	64	Retired financial consultant	Two children, each with a minimum of three children each	Caregiver for his sister who has Parkinson's disease

Figure Captions

Figure 5.1: Mean diagnostic probability assigned to diagnosis supported by the familiar patient identifier, mean diagnostic probability assigned to the plausible alternate diagnosis, and mean diagnostic probability assigned to the other two diagnostic categories. Data is presented across testing group (immediate test, 24-hour delay, or one-week delay). Error bars represent standard error of the mean.

Figure 5.1:


Diagnosis supported by the familiar patient identity
Plausible alternate diagnosis
60
50
40
30
20
10
0
Immediate Test
24-hour Delay
1-week Delay
Testing Group

Chapter 6:

General Discussion and Conclusion

General Discussion

Medical diagnosis is a complex decision that relies on many different information sources and cognitive strategies, with large consequences for error. Previous research has proposed the use of what can broadly be considered analytic strategies for clinical reasoning, including the use of defining features (Medin, 1989; Regehr & Norman, 1996), Bayesian probabilities (Weinstein, Fineberg, Elstein, Frazier, Neuhauser, Neutra, & McNeill, 1980), regression approaches to diagnosis and reliance on clinical features (Freidman, Elstein, Wolf, Murphy, Franz, Heckerling *et al*, 1999; Wigton, 1988), and semantic axes (Bordage, 2007). These models generally propose that the process of diagnosis is accessible to conscious control, rely on decision criteria or rules, are deliberate, and relatively methodical. Other models propose that many aspects of clinical reasoning occur quickly, without reliance on rules, are outside of conscious control, and are based on the rapid unconscious matching of a current problem to a previously seen case retrieved from memory (proposed by Brooks, 1978; Brooks, 1990; reviewed in Norman, Young, & Brooks, 2007). This perspective proposes that individuals can identify objects without an effortful search for and identification of features. Instead, individuals can rapidly recognize objects through a similarity matching between the current object and a set of individual examples (or instances), collected throughout their maturation. Support for this perspective can be found in both cognitive psychology (e.g. Brooks & Hannah, 2006; Homa, Sterling & Trepel, 1981), and in medicine (Brooks, Norman, & Allen, 1991; Hatala, Norman & Brooks, 1999).

Support for an instance model has been established as one process in medical diagnosis (Brooks, Norman, & Allen, 1991; Hatala, Norman, & Brooks, 1999). However, medicine is a domain in which clear rules for membership exist – there are signs, symptoms, 'normal' ranges of clinical values, and mechanisms to explain the underlying relationship between symptoms that exist as definitional properties of an illness. The specification of critical or defining features dictates the difference between what should be relevant to a clinical diagnosis and what is irrelevant, or should not influence a clinical decision. Where previous research has demonstrated the role of familiarity at the level of similarity to a whole case (Brooks, Norman, & Allen, 1991; Kulatunga-Moruzi, Brooks, & Norman, 2001), a main theme within this dissertation is examining the role of similarity within the application of a diagnostic rule. More specifically, the research presented in this dissertation demonstrates the influence of familiar featural information, as distinct from whole case similarity, and the biasing impact of diagnostically irrelevant information. The influence of similarity was examined both in terms of differential diagnostic outcomes, and also in the context of the identification and weighting of clinically relevant features. Overall, this program of research set out three main questions: 1) Can familiarity influence reliance on features?, 2) Can familiarity influence diagnostic decisions in the context of a rule?, and 3) Can familiar non-diagnostic information influence decision making?.

Can familiarity influence reliance on features?

The research presented in Chapter 3 was an initial experimental demonstration of the reliance on familiar instantiations of features relative to unique manifestations of the features in the diagnostic rule. Novice diagnosticians assigned significantly more diagnostic probability to the disorder supported by the familiar feature instantiations, and demonstrated an increased weighting of familiar features above equally valid novel feature instantiations. The influence of familiar feature instantiations remained following a 24-hour delay, and the heavy reliance on familiar symptom descriptions remained when examining a subpopulation of participants who could consistently and accurately identify the clinically relevant features present in the written case vignette. The research in this chapter suggests that participants appear to be heavily weighting familiar features, rather than an inability to recognize novel feature manifestations, indicating that diagnosis is more complex than simply the ability to valid criterial features.

Can familiarity influence diagnostic decisions in the context of a rule?

Chapter 4 presented two empirical studies continuing to examine the role of similarity within the application of a diagnostic rule. The first experiment presented in Chapter 4 clearly replicates the findings in Chapter 3, demonstrating the continued role of similarity of diagnostically relevant features, even when the transition from instantiated to informational level was made more difficult. Further, the proportion of participants who reported using a feature counting strategy was manipulated by having participants report clinically relevant features in either one large text field (Experiment 1), or four smaller text fields with a requirement of reporting a minimum three features (Experiment

2). Even when more participants reported using a feature counting strategy for diagnosis, the influence of familiarity remained, and remained consistent across immediate and 24-hour delay testing. Finally, participants reported more clinically relevant features that were familiar than equally valid novel features, and this differential reporting was seen more frequently when participants were indirectly cued to adopt a counting strategy (Experiment 2).

Can familiar non-diagnostic information influence decision making?

Distinct from studies presented thus far, the research presented in Chapter 5 suggests that non-diagnostic information can influence the diagnostic decisions of novices. Additionally, the role of familiar patient identity information appears to increase across delay periods (up to 1 week), and can influence the likelihood that clinically relevant features are mentioned in support of a diagnosis. More specifically, participants reported significantly more diagnostically relevant features in support of the diagnosis congruent with the familiar patient identity information than the plausible alternate diagnosis. Further, patient identity information was carefully constructed to be diagnostically irrelevant, and so the research presented in this chapter suggests a match between a current and previous exemplar can be made based on a diagnostically irrelevant stimulus dimension.

Implications

The results presented in Chapters 3 and 4 have important theoretical and practical significance. First, these studies indicate a strong role of similarity in the diagnostic

decisions of novices. More specifically, participants assigned significantly more diagnostic probability to the disorder supported by familiar instantiations or symptom descriptions than to the equally valid novel feature descriptions. This is significant because it demonstrates a clear influence of similarity within the use of a strong analytic procedure, in this case the use of medical rules. The role of similarity within a diagnostic rule blurs the boundary of similarity and rule based processes that have previously been argued to be separate, and competing processes (e.g. Croskerry, 2000; Elstein, Shulman, & Sprafka, 1990).

The ability to translate the variety of individual symptom descriptions to their informational level is a difficult task faced by learners. A good rule functions as an indication of what aspects of an instance to attend to, and learning must encompass discovering what the language in the rule means – learning the unique perceptual variability meant to be described or indicated by a good medical rule. Similarity or familiarity can function as an adaptive strategy for learners; it makes sense to rely on instantiations of features that you have seen before, particularly in the context of uncertainty. This use of familiarity within a formal diagnostic rule, can be seen in individuals who can perfectly translate from the instantiated to the informational level (Chapter 3), and can be seen in those indirectly cued to adopt a strategy to count features (Chapter 4). Further, in both Chapters 3 and 4 the influence of familiar feature instantiations is seen to last a minimum of 24 hours.

The research presented in the initial chapters in this thesis demonstrated two slightly different influences of familiarity. In Chapter 3, participants who were identified

as perfect translators – students who consistently correctly reported and identified the informational level features present in the case vignette – showed an influence of similarity of the same magnitude of their peers. This suggests that the familiar symptom descriptions are being weighted more heavily than the equally valid novel symptom descriptions within the same case. The differential weighting of clinical features is not a new hypothesis (Freidman, Elstein, Wolf, Murphy, Franz, Heckerling *et al*, 1999; Wigton, 1988), however familiarity as a contributing factor for re-weighting a clinical feature has rarely been investigated (Dore, Weaver, & Norman, 2005).

In Chapter 4, a second mediating role of familiarity was demonstrated, as familiar features were more frequently reported in support of a diagnosis than equally valid novel symptom descriptions. Feature instantiations in this experiment were created to closely mimic a patient voice. To this end, all symptoms were created to reflect descriptions of symptoms that patients might generate using lay language. This manipulation was included in order to increase the generalizability of the finding to clinical reasoning – patients rarely use anything close to the medical rules in their own descriptions of symptoms (Pahal & Li, 2006). The first experiment in Chapter 4 demonstrated a clear reliance on familiarity when participants were asked to diagnose case vignettes and report clinically relevant features, a replication and extension of the research reported in Chapter 3. However, medicine often has clear criteria for diagnosis in the form of diagnostic rules. When participants were reminded of the counting rule, and encouraged to count the features present in a written case vignette, a reliance on familiar feature descriptions remained. Further, participants report more diagnostic features when cued to

count, but preferentially report familiar symptoms. Combining the results from the first two empirical chapters in this thesis, it appears as though familiarity may influence the relative importance of features (Chapter 3), and whether features get reported, or noticed, at all (Chapter 4).

Research presented in the first section of this dissertation adds to research supporting a similarity based account of clinical reasoning (proposed by Brooks, 1978; Brooks, 1990; reviewed in Norman, Young, & Brooks, 2007) and of the differential treatment of familiar features (Brooks & Hannah, 2006; Dore, Weaver, & Norman, 2005). The unique contributions of this research include the demonstration of a reliance on familiarity within written case protocols, indicating that familiarity is not limited to whole case similarity (Brooks, Norman, & Allen, 1991; Kulatunga-Moruzi, Brooks, & Norman, 2001), strictly synonymous terms (Dore, Weaver, & Norman, 2005), or visual case presentation (Brooks, Norman, & Allen, 1991; Kulatunga-Moruzi, Brooks, & Norman, 2001). Rather, research presented here suggest a role for familiarity within a strict analytic rule, indicating that similarity and rule base processes are not perfectly distinguishable competing processes (e.g. Evans, 2008; Pothos, 2005), and instead indicating a potentially adaptive role for similarity, particularly in the learner.

The research presented in Chapter 5 provides further evidence for a similarity based model of medical decision making. Chapter 5 demonstrates a clear reliance on diagnostically irrelevant information, specifically the role of patient identity information. It is important to reiterate that the non-diagnostic information used in this study was

carefully constructed to avoid the possibility of participants relying on a representativeness heuristic (Tversky & Kahneman, 1974). Instead, patient identity information was created to avoid medical stereotypes, and under no circumstances did it reflect age, gender, or racial stereotypes in medicine.

Participants in Chapter 5 assigned significantly more diagnostic probability to the disorder supported by the familiar patient identity, reported significantly more clinically relevant symptoms in support of the familiar diagnosis, and this effect was not transient, lasting up to one week following training. The research presented in Chapter 5 continues to support a similarity based model of decision making, and demonstrates that cueing of a diagnostic hypothesis can occur both for clinically relevant (Chapters 2 and 3) and clinically irrelevant information (Chapters 4). However, all data presented up to this point involved learners and static, simple medical cases. This is a far cry from everyday clinical decision making, and addresses little in terms of the development of expertise and expert performance across the development and licensing of physicians.

Limitations and future directions

While research using medicine as a domain of study has many advantages including: opportunity for diverse and rich stimuli, the ability to orthogonalize formal knowledge and specific previous experience, and the ability to study the development of expertise, it is a domain that is not without challenges. The examination of the development of expertise in medicine requires the study of highly qualified and difficult

to reach expert populations. As such, the research formally presented within this thesis investigates the role of specific previous experience within novice diagnosticians. The use of novice diagnosticians, with static written case presentations does little to inform our understanding of the specific influences in dynamic patient encounters with more expert populations. As such, the expansion of this research program into more interactive patient physician encounters with more expert populations remains a direction for future research.

Preliminary data from a current pilot study is encouraging. In that study, medical students and residents were asked to interact with and diagnose a standardized patient (an actor trained to portray a particular illness) in the context of an Objective Structured Clinical Examination (OSCE, similar to a bell-ringer exam). The participants first encountered a clear diagnostic case (for example, dementia) with a particular standardized patient, but were later asked to interview a second standardized patient who was age, gender, name, and appearance matched to the previously seen standardized patient, but is presented with symptoms that could be interpreted in more than one way (for example, dementia and depression), creating an ambiguous case. If participants rely on the similarity to the previously encountered case, then participants should more frequently diagnose the patient with a similar diagnosis to the matched previous case. This study was originally designed to include only medical students, but during implementation we have also been including a slightly more expert population of psychiatric residents.

When cued with familiar non-diagnostic information, preliminary results suggest that psychiatry residents might be more likely to diagnose an ambiguous diagnostic case with the same psychiatric condition as a previously seen case that contained similar non-diagnostic information. No such pattern was seen in the clinical clerks. Consistent with previous research, more expert diagnosticians demonstrated a reliance on familiar non-diagnostic information (Hatala, Norman, & Brooks, 1999), or stated more generally - similarity to a previously seen case (in medicine: Eva & Cunnington, 2006; Kulatunga-Moruzi, Brooks, & Norman, 2004, in experts in other domains: e.g. Hutton & Klein, 1999). Secondly, a main effect of cueing condition was found when examining the responses of the expert evaluators – participants were rated as collecting less diagnostic information about the cued diagnosis than the plausible alternate diagnosis. Further, this did not interact significantly with expertise level of the participant, indicating that regardless of expertise, participants were collecting less diagnostic information for the cued diagnosis than the plausible alternate.

Research has demonstrated that a match between a current and previous case can be based on the similarity between instances at the level of the whole case. Further, upon qualitative analysis, it is interesting to note that the participants, regardless of level of expertise, were most likely to ask questions that did not differentiate between the two most likely diagnoses.

With the transition to live patient encounters, some experimental control is lost. We cannot guarantee that each clerk and resident was exposed to the same amount of cueing, or diagnostically relevant information, as this would vary with the direction of the

diagnostic interview. Finally, this study demonstrates a clear role of non-diagnostic information within dynamic patient encounters, but through the use of simulated patients and controlled diagnostic cases, we still remain one step away from live, real, clinical encounters.

This pilot study has the potential to extend the finding of similarity-based reasoning from visually dominant specialties like dermatology and electrocardiology to psychiatry, where clinical information is most frequently based on verbal descriptions. Furthermore, the current study seems to demonstrate non-analytic reasoning in actual patient encounters, rather than static case presentations. This suggests that the influence of non-analytic reasoning may function within clinical decision making in a dynamic way, not simply functioning as a 'first' answer within the context of a static image or case. Instead, it may be that quick matching to previously encountered cases may be what brings a diagnostic hypothesis to mind, which is then followed by a more systematic analytical phase. While these particular conclusions remain speculative at best, the interactive nature of similarity based and formal knowledge based processes remain a important direction for future research.

Evidence in categorization and medicine has demonstrated a strong role for personal (Kim & Ahn 2002) and causal theories in reasoning (Murphy & Medin 1985; see also Gentner & Brem, 1999; Keil, 1991; Lakoff, 1987; Medin & Watternmaker, 1997; Woods, Brooks, & Norman, 2005). While these empirical demonstrations have shown a clear role for causal and personal theories, studies from medicine indicate that causal mechanisms are most frequently used in order to explicitly reason through difficult

clinical problems (Norman, Trott, Brooks, & Smith, 1994; Verkoeijen, Rikers, Schmidt, van der Weil, & Koomna, 2004), or create coherence among symptoms to aid novices (Woods, Brooks, & Norman, 2005). A clear direction for future research will be examining the role of causal models in the interpretation, reliance, and use of clinically relevant features. It is possible that a strong causal model, as may be present in some clinical phenomena, could aid in the detection and interpretation of clinically relevant features, as is seen with a suggested diagnosis (LeBlanc, Brooks, & Norman, 2002; LeBlanc, Norman, & Brooks, 2001). However, even when features are selected as causally relevant, familiar features may be relied on more than novel but equally causally relevant features.

Finally, research presented in this thesis generally supports an adaptive use of similarity within novices. Clinically relevant features can present in a multitude of ways, and similarity to previously seen examples may aid a learner to recognize symptoms that are not well captured by the language of the diagnostic rule (i.e. violacious colour cues you to attend to a colour dimension, but provides little information regarding the boundary conditions of what could be considered violacious). While this adaptive use of similarity is supported by research presented here, an explicit examination of the benefits of similarity within instruction would be beneficial to not only the clinical reasoning literature, but the transfer and medical education literatures more generally. Also, we currently have little data examining whether instructional manipulations can alter how students approach diagnostic rules, or if instructional manipulations can make the representation or use of features more or less susceptible to familiarity, or perhaps even if

mental representations of features, or their subsequent 'translation', can remain relatively flexible. Again, these expansions of this research program remain a possible future direction for research.

Conclusions

Three main goals were proposed for this research program, and were integrated within each empirical question. For this dissertation, medicine was used as a rich and dynamic categorization task. The grouping and treatment of patients within a diagnostic category allows a strong analogy to theories of categorization drawn from cognitive psychology. Through the examination of clinical reasoning from a cognitive psychology perspective, the research presented within this thesis strongly support a similarity based account of clinical reasoning, and suggest a role for similarity within both the application of a strong diagnostic rule, and in the influence of non-diagnostic information.

The use of psychiatry as a medical domain of study provided a rich opportunity to examine the role of familiarity in the use, interpretation, and reporting of clinically relevant features. Further, psychiatric symptoms are most frequently reported verbally, and so a demonstration of the role of familiarity within the context of written case vignettes has expanded the role of familiarity from visually salient domains such as dermatology (Brooks, Norman, & Allen, 1991; Kulatunga-Moruzi, Brooks, & Norman, 2001) to include other more verbally dominated domains. Psychiatry is known as a difficult domain of medicine, and has a history of ambiguity in symptom identification. The research presented in this thesis suggests that familiarity can influence the use of

clinically relevant and clinically irrelevant information in the diagnostic decisions of novices. Finally, the ability to represent features at the instantiated and informational level, the ability to manipulate difficulty of translation, the ability to investigate the role of ambiguity, causal and personal theories, and expand a set of stimuli to examine expert performance clearly demonstrates the value and versatility of using medicine as a domain to examine complex decision making. The ability to examine all of these phenomena using a very similar stimulus set may seem like a small point, but cognitive psychology tends to be dominated by specific paradigms and experimental protocols designed to carefully examine one phenomenon. The ability to use similar stimuli and protocols across a variety of cognitive processes allows for research into each of these processes independently, and hopefully also examine the coordination of multiple reasoning strategies and influences.

In conclusion, this research program has demonstrated the adaptive role of similarity within the application of a diagnostic rule. This is separate from whole-case similarity that is typically referred to by an exemplar-based model (as suggested by Brooks, 1978; Brooks, 1990, and supported by Brooks, Norman, & Allen, 1991; Estes, 1976; Hatala, Norman & Brooks, 1999; Hintzman, 1986; Minda & Smith, 2001; Medin, 1989; Medin, Altom, Edelson & Freko, 1982; Medin & Schaffer, 1978; Nosofsky, 1989; Young, Brooks, & Norman, 2007), but still suggests a similarity-based process of clinical reasoning. Additionally, research presented here demonstrates that familiar symptom manifestations are frequently weighted more heavily, and are reported more frequently than equally valid novel manifestations. Further, this research program has demonstrated

the influence of familiar clinically irrelevant information, and suggest that familiarity can influence the probability that a feature is mentioned in support of a diagnosis, and may possibly indicate that similarity to a previously seen case could provide early diagnostic hypotheses.

Previously, it has been suggested that analytic and non-analytic reasoning are differentiable and distinct modes of processing (see Pothos, 2005 for a review), and within the clinical reasoning literature, non-analytic reasoning has been suggested to be error-prone and inferior (e.g. Croskerry, 2002; Lamond & Thompson, 2000). However, more recent research has demonstrated the possibility of combined strategies (Ark, Brooks, & Eva, 2006; 2007), and the role of similarity within a strict diagnostic rule (Young, Brooks, & Norman, 2007; Dore *et al*, 2005; McLaughlin, Heemskerk, Herman, Ainslie, Rikers, & Schmidt, 2008). The possibility of combined approaches, and the demonstration of similarity working within a diagnostic rules suggest that these two processes may not be as separable as previously thought. The research presented here may suggest an interdependent role for analytic and non-analytic processes. Within the presented studies, participants were faced with the difficult task of diagnosing patient cases in which more than one diagnostic answer was possible. It is possible that the similarity to a previous case, here manipulated by familiar non-diagnostic information, may function to cue a diagnostic hypothesis. By bringing a hypothesis to mind, which may be followed by the search for supporting information, previous experience is neither 'good' nor 'bad', but rather simply a starting point within the diagnostic process. While this suggestion remains speculative at best, it demonstrates the possibility of inter-

dependent roles for non-analytic and analytic processes, and the suggestion that non-analytic processing is adaptive, and perhaps even unavoidable.

References

(Chapters 1 and 6)

American Psychiatric Association. (2000). Diagnostic and Statistical Manual for Mental Disorders: Fourth Edition, Text Revision. *American Psychiatric Association Press,* Washington.

Ark, T. K., Brooks, L. R., & Eva, K.W. (2006). Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Academic Medicine, 81,* 405-409.

Ark, T. K., Brooks, L. R., & Eva, K.W. (2007). The benefits of flexibility; the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Medical Education, 41,* 281-287.

Bordage, G. Prototypes and semantic qualifiers: from past to present. *Medical Education, 41*, 737-745.

Brooks L. R. (1978). Non-analytic concept formation and memory for instances. In: Rosch & Lloyd, eds. Cognition and Categorization. Hillsdale, NJ: Lawrence Erlbaum.

Brooks L. R. (1990). Concept formation and particularizing learning. In: Hanson, ed. Information, Language and Cognition. UBC Press, Vancouver.

Brooks, L. R., & Hannah, S.D. (2006) Instantiated features and the use of "Rules". *Journal of Experimental Psychology: General, 132,* 133-151.

Brooks, L. R., LeBlanc, V. R. & Norman, G. R. (2000). On the difficulty of noticing

obvious features in patient appearance. *Psychological Science, 11*, 112-117.

Brooks, L.R., Norman, G.R., & Allen. S.W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278-87.

Croskerry, P. (2000). The cognitive imperative: Thinking about how we think. *Academic Emergency Medicine, 7,* 1223-1231.

Dore, K., Weaver, B., & Norman, G. (2005). Does a rose by any other name smell as sweet? The importance of variation in case presentation to diagnostic reasoning. Talk presented at the meeting for Research in Medical Education, Washington DC.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). An analysis of clinical reasoning. Harvard University Press; Cambridge.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83,* 37-64.

Eva, K. W., & Cunnington, J. P. W. (2006). The difficulty with experience: Does practice increase susceptibility to premature closure? *Journal of Continuing Education in Health Professions, 26,* 192-198.

Evans, J. S. B. T. (2008). Dual-Processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255-78.

Feltovich, P.J., & Barrows, H.S. (1984). Issues of generality in medical problem solving. In: Schmidt HG, De Volder ML, eds. Tutorials in Problem-Based Learning. Assess/Maastricht: Van Gorcum; 1280142.

Freidman, C. P., Elstein, A. S., Wolf, F. M., Murphy, G. C., Franz, T. M., Heckerling, P. S., *et al.* Enhancement of clinicians' diagnostic reasoning by computer-based consultation: A multi-site study of 2 systems. *Journal of the American Medical Association, 282*, 1851-1856.

Gentner, D., & Brem, S. K. (1999). Is snow really like a shovel? Distinguishing similarity from thematic relatedness. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty-first annual conference of the cognitive science society* (pp. 179–184). Mahwah, NJ: Erlbaum.

Hatala, R., Norman G. R., & Brooks, L. R. (1999). Research basic to medical education: influence of a single example on subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine, 11,* 110-117.

Hintzman, D. L. (1986). Schema-abstraction in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 418–439.

Hutton, R. J. B., & Klein, G. (1999). Expert decision making. *Systems Engineering, 2*, 32-45.

Keil F. C. (1991). The emergence of theoretical beliefs as constraints on concepts. In S. Carey & R. Gelman (Eds.) *The epigenesis of mind: Essays on biology and cognition* (237-256). Lawrence Erlbaum Associates; Philadelphia, PA.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory based representations of mental disorders predicts their diagnostic reasoning and memory. *Journal of Experimental Psychology: General 131,* 451-476.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine, 13*, 110-6.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lamond, D., & Thompson, C. (2000). Intuition and analysis in decision making and choice. *Journal of Nursing Scholarship, 32,* 411-414.

LeBlanc, V. R., Brooks, L. R., Norman, G. R. (2002). Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Academic Medicine, 77,* s67-s69.

LeBlanc, V. R., Norman, G. R., Brooks, L. R. (2001). Effect of a diagnostic suggestion on diagnostic accuracy and clinical feature identification. *Academic Medicine, 76,* s18-s20.

McLaughlin, K., Heemskerk, L., Herman, R., Ainslie, M., Rikers, R. M., & Schmidt, R. M. J. P. (2008). Initial diagnostic hypotheses bias analytic information processing in non-visual domains. *Medical Education, 5,* 496-502.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8,* 37-50.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Wattenmaker, W. D. (1997). Category cohesiveness, theories, and cognitive archeology. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge, UK: Cambridge University Press.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27,* 775-99.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Norman, G. R., Trott, A. L., Brooks, L. R., & Smith, E. K. M. (1994). Cognitive differences in clinical reasoning related to post-graduate training. *Teaching and Learning in Medicine, 6,* 114-20.

Norman, G. R., Young, M. E., & Brooks, L. R. (2007). Non-analytic models of clinical reasoning: the role of experience. *Medical Education, 41,* 1140-1145.

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension

stimuli: A choice model analysis. *Journal of Experimental Psychology: Perception and Psychophysics*, *38*, 415–432.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Journal of Experimental Psychology: Perception and Psychophysics*, *45*, 279–290.

Pahal, J. S., & Li, H. Z. (2006). The dynamics of resident-patient communication: Data from Canada. *Communication & Medicine, 3*, 161-170.

Pathos, E. M. (2005). The rules versus similarity distinction. *Behavioural and Brain Sciences, 28,* 1-49.

Regehr ,G., & Norman, G. R. (1996). Issues in cognitive psychology: implications for professional education. *Academic Medicine, 71*, 988-1001.

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.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124-1131.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., Schmidt, H. G., van der Weil, M. W. J., & Koomna, J. P. (2004). Case presentation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Medical Education, 38*, 617-27.

Weinstein, M. C., Fineberg, H. V., Elstein, A. S., Frazier, H. S., Neuhauser, D., Neutra, R. R., & McNeil, B.J. (1980). Clinical Decision Analysis. Phillidelphia: WB Saunders.

Wigton, R. S. (1988). Use of linear models to analyze physicians' decisions. *Medical Decision Making, 8*, 241-252.

Woods, N. N., Brooks, L. R., & Norman, G. R. (2005). The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Medical Education, 39,* 107-112.

Young, M. E., Brooks, L. R., & Norman, G. R. (2007). Found in translation: the impact of familiar symptom descriptions on diagnosis in novices. *Medical Education, 41,* 1146-1151.