Precision Medicine: Measurements without concepts are blind; concepts without observations are empty.

DONALD stanley[1]

[1]Maine Medical Center - Scarborough Campus

May 5, 2020

Abstract

What is precision medicine? No standard definition is found. Sometimes described as High Definition Medicine which is: “the dynamic assessment , management, and understanding of an individual’s health measured at (or near) its most basic units.” The impetus for this venture was advanced and stimulated by biogenetics, the study of how genes and their products affect health but also contribute to disease or resistance to cure. Medicine is an epistemology: a way of knowing, perceiving, remembering, finding out, proving, inferring, wondering, reflecting, a conceptual knowing relying upon observations fitted to disease concepts.

1. Introduction

How should we conceive of and evaluate the process of medical diagnosis? Saying that it is a matter of inferring the correct disease from the available signs and symptoms is too sparse. Clinical diagnosis is an*experimental* science based on observation, hypotheses, and testing. It is a dynamic process that involves observation, diagnostic conjectures and testing, possibly leading to new or revised conjectures. For the clinician, it should always be a reflexive process subject to revision.

Consider for example the following scenario: A 54-year-old man with no previous history of chronic disease suffers sudden substernal chest pain and is rushed to an emergency room. His presenting symptoms also include tachycardia (abnormally rapid heart rate), shortness of breath and sweating. The challenge a clinician faces in cases like this is not just to evaluate the likelihood of different possible causes of these symptoms; she also has to select which hypotheses to actively consider in the first place, which to prioritize for further testing, which can be put aside for the time being and when to initiate treatment on the basis of a given hypothesis. Additionally, all of these decisions presuppose that the relevant hypotheses have been generated and introduced into the diagnostic inquiry. The clinician does not start out considering every possible cause of chest pain known to medicine; rather, she needs to decide when and how to generate new diagnostic hypotheses, as well as when to stop.

In this paper, we present a framework for understanding the different kinds of reasoning underlying medical diagnosis as it occurs in clinical practice. Our starting point is the observation that, in addition to evaluating the likelihood of candidate diagnostic hypotheses in light of the evidence, the process of medical diagnosis involves two distinct types of reasoning, namely: (i) reasoning concerned with generating new candidate hypotheses and (ii) reasoning about which hypotheses should be pursued, i.e. prioritized for testing and further consideration. That these forms of reasoning are crucial to understanding inquiry was argued by C. S. Peirce in his writings on the form of reasoning he called*abduction* . Following recent commentators (Upshur 1997; Stanley and Campos 2013, 2015; Chiffi and Zanotti 2015), we believe that Peirce’s mature account of abduction provides important lessons for understanding diagnostic reasoning. Specifically, we argue that recent Peirce scholarship, which construes abduction in terms of*strategic reasoning* , provides a promising framework for analyzing diagnostic reasoning.

Our aims in presenting this framework are primarily normative: we want to explicate the *reasons* which underlie diagnostic reasoning in realistic clinical situations, rather than necessarily describing the psychological processes clinicians go through in diagnosis. The best psychological description may often be that the clinician makes a quick, intuitive judgment, perhaps based on some unconscious heuristic. By contrast, our framework aims to explicate the factors which make such judgments reasonable in a concrete, clinical situation. Despite this normative scope, our aims in this paper are not prescriptive in the sense of recommending whether existing practices can or should be improved. Rather, our main aim is to be able to explicate diagnostic reasoning as it occurs in current practice.

A unified, normative framework for understanding clinical reasoning is currently lacking from the methodological literature. On the one hand, when hypothesis generation is addressed (e.g. Kassirer, Wong and Kopelman 2010, Ch. 13) it is mainly discussed from the perspective of cognitive psychology without an underlying normative framework. On the other hand, the probabilistic approach to clinical-decision-making currently popular in the medical literature—the so-called *threshold approach* —while normative, does not address the question of hypothesis generation. As we shall argue, because of the way hypothesis generation and reasoning about pursuit are intertwined, this neglect means that threshold models, in their current form, fail to capture all relevant reasons for pursuing a hypothesis.

Our discussion proceeds as follows. We start, in Section 2, by outlining our understanding of Peircean abduction and, Section 3, explaining how these ideas apply to medical diagnosis. In Section 4, we then use this framework to analyze a clinical case study. In Section 5 we return to our criticism of threshold models. Finally, in Section 6, we defend the strategic reasoning interpretation of abduction as a framework for analyzing diagnostic reasoning.

2. Peirce on Induction and Abduction

Abduction, on Peirce's mature view, differs in kind from induction. Induction, for Pierce, includes any use of empirical evidence to test or support a hypothesis. Abduction, by contrast, is reasoning which introduces hypotheses into inquiry in the first place: "abduction commits us to nothing. It merely causes a hypothesis to be set down upon our docket of cases to be tried" (Pierce 1932-58, §5.602).[1] Abduction, in other words, aims to identify hypotheses which are worth *pursuing* further (McKaughan 2008). As Peirce stresses, the pursuit-worthiness of a hypothesis cannot be reduced to how plausible or probable it is; in fact: "Sometimes the very fact that a hypothesis is improbable recommends it for provisional acceptance on probation [i.e. for further testing]" (Peirce 1932-58, §6.533). For instance, if a hypothesis can be reliably falsified and if it would be valuable (given the goals of inquiry) to rule it out conclusively, increasing its *improbability* can make it more pursuit-worthy (cf. Nyrup 2015).

While Peirce thus distinguishes abduction from induction, he is somewhat ambiguous as to what it means for abduction to "introduce" or "suggest" hypotheses. Does it mean *generating* hypotheses or merely to *select* already generated hypotheses for further pursuit?[2] Sometimes it is characterized in terms of generation. For instance: "Abduction consists in studying facts and *devising* a theory to explain them"; it is "the *process of forming* an explanatory hypothesis" (1932-58, §5.145, §5.171, emphases added). However, abduction is also described as "the operation of *adopting* an explanatory hypothesis" (§5.189, emphasis added), through the characteristic inference schema (*ibid.*):

> The surprising fact, C ¸ is observed;
>
> But if A were true, C would be a matter of course,
>
> Hence, there is reason to suspect that A is true.

In order to employ this inference schema (as Peirce himself notices) a reasoner must already have formulated the explanatory hypothesis A mentioned in the second step, seemingly contradicting the idea that abduction is the process through which A is generated.

This tension can be resolved by interpreting abduction as encompassing both the generation *and* the selection of hypotheses to be pursued.[3] The reason it makes sense to include both under the label 'abduction' is that they share a common goal, namely to identify hypotheses worth pursuing further. As Fann (1970, 41-43) argues, simply formulating new hypotheses is not difficult. Rather, since we cannot examine every conceivable hypothesis, the problem is to generate hypotheses it is worth spending time and resources considering further. Generative reasoning should thus be evaluated along the same dimensions as reasoning concerned with selecting hypotheses for pursuit. The above schema outlines one salient criterion for pursuit-worthiness, namely whether the proposed hypothesis accounts for the phenomena we seek to explain. But other criteria—e.g. testability and economy—are relevant too; the schema does not exhaust what Peirce meant by abduction (Pietarinen and Bellucci 2014, 355-6).

This interpretation also allows us to answer two objections, often raised against normative accounts of generative reasoning.[4] First, since we do not control which ideas occur to us, how can anything normative be said about hypothesis generation? The answer is that one can still choose *when* ,*whether* and *how* to generate new ideas, and these choices can be normatively evaluated in terms of how efficiently they generate pursuit-worthy hypotheses. For example, a rather ineffective strategy would be to flip through a medical lexicon, hoping to chance upon diseases with symptoms similar to the ones observed in the patient. Experienced physicians will (one hopes) be able to deploy better strategies for generating hypotheses. We discuss some of these in the clinical case in Section 4.

Second, why assume that there is any particular kind of reasoning (viz. "abduction") through which physicians generate hypotheses? The answer is that we do not assume this. For us, the term 'abduction' does not refer to any particular process of reasoning. Rather, the distinction between abduction and induction refers to a difference in goals: abduction aims to identify hypotheses worth pursuing further; induction aims to determine the likelihood of a hypothesis. Similarly, although the generation and selection of hypotheses for pursuit share a normative foundation, the concrete (e.g. verbal, mental, symbolic) reasoning processes involved in generating or formulating a hypothesis of course differ from those involved in choosing between already formulated hypotheses. Trying to overcome the problem of having *too few*hypotheses is still different from trying to overcome the problem of having *too many* . The two can still be distinguished.

A particularly useful, unified interpretation of abduction, which has been proposed in recent scholarship (Hintikka 1998; Paavola 2004; Pietarinen and Bellucci 2014), construes abduction in terms of*strategic reasoning* . On this interpretation, individual choices about hypothesis generation and pursuit should be evaluated in terms of whether they contribute to an overall investigative strategy. We will explain this interpretation in more detail and show how it helps make sense of our clinical case study in Section 6.

In summary, we draw the following lessons from these Peircean ideas. First, we distinguish three general types of reasoning in medical diagnosis: (1) generating plausible diagnostic hypotheses; (2) selecting and prioritizing these for pursuit, i.e. for further testing and consideration; (3) accepting or rejecting diagnoses, or more broadly, evaluating the how likely different diagnoses are in light of the available evidence. Second, we take (1) and (2) to share the same goal, viz. pursuit-worthiness. Third, thinking about diagnosis in terms of strategic reasoning provides a useful unified framework for evaluating the generation and selection of hypotheses. We will now describe in more detail how these lessons apply to medical diagnosis.

3. Generation and Pursuit-Worthiness in Medical Diagnosis

A typical diagnostic process begins when a patient arrives at a hospital or clinic and reports certain symptoms or ailments. Insofar as the situation allows it, the physician will start by interviewing the patient and performing a physical examination to gather information about the patient's state, how long they have experienced the symptoms and their broader medical history. Based on these, the physician tries to generate one or more possible explanations for the salient aspects of the case. For example, if a patient has uncontrollable hypertension (high blood pressure), the physician may conjecture that the patient has renal artery stenosis (narrowing of kidney arteries), since this would explain the signs.

Our use of the term 'generation' here should be understood in a broad sense. In most cases, medical diagnosis does not involve formulating completely novel hypotheses. Rather, it will primarily be a case of recalling already known conditions and realizing that they could potentially account for the salient signs and symptoms.[5] However, this is not a sharp distinction. When facing atypical or complex cases, physicians may have to combine their knowledge of possible diseases in novel ways to explain the condition of that specific patient.

While physicians will often be able to think of a large number of theoretically possible diagnoses, it is neither practically possible nor advisable to consider every single one. Physicians need to pick out a limited number of hypotheses to focus on. The set of diagnostic hypotheses actively considered at a given time is called the *differential diagnosis*.[6] There are good reasons why physicians need to limit themselves to a relatively narrow differential diagnosis. First, limitations of working memory preclude working on too many hypotheses at once (Sox, Higgins and Owens 2013, 9). Second, actively pursuing too many hypotheses can lead to potentially harmful over-testing (Richardson et al 1999, 1214-15). Third, in emergency situations there is no time to test every conceivable hypothesis. With a patient's health or life on the line, we need to be able to effectively, rapidly and *efficiently* determine the likeliest cause of their ailments. This requires wisely selecting a limited range of hypotheses to focus on.

These arguments are often applied to the *choice* of a differential diagnosis, but similar points apply already at the *generative* stage. Just as it is inadvisable to select too broad a differential diagnosis, physicians cannot—and should not—try to generate a list of every single possible explanation before selecting a differential diagnosis. As argued above, generating hypotheses and selecting them for pursuit are subject to the same normative considerations. Just as physicians need to make good choices about which hypotheses to include in their differential diagnoses and which of these to prioritize for testing, they must choose how to generate possible diagnoses, as well as when to *stop*.

On the grounds of what kinds of considerations, then, should these decisions be made? The most popular approach to the problem of choosing whether to test a given hypothesis in the medical literature is the so-called *threshold approach* (Pauker and Kassirer 1980; Djulbegovic et al 2015). This approach is based on decision-theoretic models which compare, e.g., a choice between: (i) applying treatment on the assumption that the hypothesis H is true; (ii) applying a test for H, and then only apply treatment if the test is positive; (iii) stop working on H, i.e. neither test nor treat. Given quantitative estimates of (a) the reliability of the test, (b) the likelihood of the salient consequences of treating and testing and (c) the utility of these consequences, one can derive thresholds for how probable the hypothesis needs to be in order for it to be most rational to test, treat or abandon the hypothesis.

Threshold models highlight a number of factors that should be weighed against each other in clinical decision-making, including: How reliable are the available tests? How safe/harmful are the tests? How dangerous would the disease be, if missed? How effective is the available treatment? How safe/harmful is the treatment in itself? Briefly put, on this approach, physicians have to consider whether their confidence in H is high enough for the potential benefits of treating the disease (if H is true) to outweigh the potential harms of treating or testing unnecessarily (if H is false).

While these factors are indeed important, we want to highlight a further type of consideration, which can be called *strategic considerations*, that go beyond the direct consequences of tests and treatments for the health of the patient. As Peirce (1938-1952, §7.220) points out, the pursuit-worthiness of a hypothesis also depends on what we might learn from pursuing the hypothesis even if it turns out to be false. Testing a hypothesis can have important downstream effects for later stages of inquiry, in addition to merely confirming or disconfirming the tested hypothesis.[7] For instance, an imaging study which fails to detect renal artery stenosis may also show that the adjacent adrenal gland is enlarged, thus instead suggesting pheochromocytoma (a tumor of the adrenal gland) as the cause of hypertension. At other times, it can be worth trying to rule out a potential diagnosis simply to make the diagnostic space more manageable, i.e. to pre-emptively prune back possibilities that might otherwise become relevant later on. If testing can be done reliably and without risk of harm, it can be worth trying to rule out even fairly unlikely hypotheses early on. Examples of this could include serologic testing for Lyme disease, fat aspiration for amyloidosis and ferritin levels for myocardial iron.

Strategic considerations involve reasoning about how pursuing a specific hypothesis can influence later stages of inquiry, including future generation of hypotheses. It is this dynamic and intertwining relationship between hypothesis generation and selection for pursuit which threshold models, in their current form, fail to capture. Before making this argument, however, we want to provide a concrete illustration of our framework, by way of analyzing a detailed clinical case.

4. Clinical Case: Chest pain

In the following, we distinguish between the description of clinical details (in *italics*) and our commentary.[8]

Scene: At home in the Northeastern U.S.A., at 08:00, a 54-year-old man is walking down the stairs to breakfast. He suffers sudden substernal chest pain radiating to his left shoulder and back. No previous history of chest pain. No previous chronic disease. His spouse immediately calls the local emergency number .

Emergency medical technologists (EMT) arrive in a quarter of an hour. Based on hospital protocol, an intravenous line is inserted in patient's right arm; he is administered morphine sulphate 10 mg, aspirin, beta blocker, supplemental oxygen by mask; an electrocardiogram (ECG) tracing is radioed to the local emergency room while he is in the ambulance. Sublingual nitroglycerin was given with minimal relief of pain. He receives nasal 100% oxygen but is not intubated. His respirations are more than 30 per minute and shallow. His skin is cool and clammy. Blood pressure: 110/78. The substernal pain is slightly relieved with medications and rest. EMT calls emergency triage nurse at nearest community hospital regarding middle-aged white male complaining of severe chest pain. He is breathing rapidly and perspiring.

The patient arrives in the emergency room and is seen immediately by a triage nurse. He complains of severe chest pain when descending the stairs to the kitchen; the pain persists . Nurse inquires if he has had a previous a history of chest pain. "No," the patient answers. She assesses his vital signs: heart tracing on electrocardiogram, respiratory rate, and temperature. She searches for any previous medical record in the computerized system to share with a physician. She is following protocol for chest pain and patient estimates pain at 7/10. He receives an additional 10 mg morphine sulphate that eases his pain. On auscultation a soft, decrescendo diastolic murmur is heard over the precordium.[9] *Respirations are labored. Tachycardia is evident, 110 bpm.*

Initial diagnosis: Physician arrives in the acute side of the emergency room. She decides that the likeliest diagnosis is acute coronary syndrome (ACS), i.e. a sudden restriction of blood flow from the coronary arteries into the heart, leading to cardiac ischemia (oxygen deprivation to the heart) and subsequent myocardial injury (death of heart cells). Based on the history and physical examination she orders two laboratory tests: ECG and serum cardiac enzymes.

Commentary: How does the clinician reach her initial diagnosis? The first step is to generate one or more diagnoses capable of explaining the most salient signs and symptoms. She knows that shortness of breath and sweating are common symptoms of ACS, so following something like Peirce's schema, she concludes that there is reason to suspect this diagnosis: if ACS were the case, it would explain the chest pain and most other symptoms. She also knows (from prevalence studies and clinical experience) that ACS is the most common cause of chest pain in men in their fifties in this part of the country.

At this stage, rather than systematically generating a wider list of potential diagnoses, she immediately orders two tests. Her reasons can be reconstructed as follows: (i) ACS is the most common cause of the chief symptom (chest pain); (ii) it can cause severe damage and is life-threatening if left untreated; (iii) the ordered tests are a rapid and effective way of confirming the hypothesis: if the ECG shows patterns characteristic of myocardial damage and the blood test shows elevated levels of the enzymes an ischemic heart muscle would release, this would be very strong evidence in favor of the hypothesis. The decision *not to generate* further hypotheses before taking action is based on the same kind of reasons as would justify *selecting already generated* hypotheses for further consideration.

Negative results: The laboratory and ECG results are negative: the cardiac enzymes test did not show elevated levels of the relevant enzymes (c-troponin). The electrocardiogram shows a rapid heart rate (120 bpm) but *none of the characteristics of heart disease (no elevation in the S-T segment, neither T wave inversion nor new Q wave occurrence). Both results rule against cardiac ischemia and thus against the diagnosis of ACS.*

Despite this lack of evidence, she does not simply dismiss her initial diagnosis. Although she faces conflicting evidence, she maintains her clinical suspicion of myocardial injury caused by ACS. She orders the tests repeated in two hours and has the patient monitored closely for any signs of worsening condition. Because of persistent pain he receives additional morphine. Meanwhile the clinician considers alternative diagnoses which could mimic the symptoms of ACS.

Commentary: Why does the clinician maintain her initial suspicion? While the negative results are sufficient to make her hold off treatment based on the initial diagnosis, she wants to avoid prematurely turning away from the commonest disease, especially since not treating it could have potentially very negative effects for the health of the patient. The case does not fit the textbook picture of cardiac injury caused by ischemia. However, she knows both from her own experience and epidemiological studies that atypical disease presentations are not uncommon: the pain may be referred to the jaw instead of chest, the T-waves in ECG may not develop early, etc.

Due to this uncertainty, she initiates two lines of action. Firstly, although she does not have a single and simple hypothesis about what could cause the atypical presentation, she decides to repeat the tests. The fact that ACS is the most common cause of chest pain and the life-threatening character of the diagnosis makes it reasonable to repeat the tests. Secondly, her lowered confidence in the ACS diagnosis is reflected in the fact that she decides to start systematically generating alternative diagnoses in order to select a differential diagnosis.

Generation and selection of differential diagnosis: The clinician thinks of alternative diagnoses. She asks herself a number of questions to guide her thinking: "what else could explain acute chest pain?", "which are the most common causes?", "what would be the most serious disease to miss?" "what could I test effectively and quickly?", "which conditions are effectively treatable?" She can think of a wide range of possibilities. For instance, she briefly considers Chagas disease, but decides this is too rare in the U.S.A. to merit immediate consideration. Another possibility would be referred pain e.g. from acute pancreatitis or another abdominal organ. In certain clinical circumstances this can be confirmed by re-examination after delaying an hour or more. In the present case, there is no time to delay, as the pain seems to be continuous even with morphine analgesia. She ultimately decides to focus on four alternative hypotheses in addition to acute coronary syndrome:

1. *Acute pulmonary embolism (a blockage of a lung artery, e.g. by a blood clot).*
2. *Acute aortic syndromes, (different kinds of damage to the aorta, the main artery leading blood out of the heart and into the body).*
3. *Pericarditis (inflammation to the pericardium, the sac surrounding the heart).*
4. *Gastrointestinal reflux disease.*

Commentary: At this stage, the clinician starts systematically generating possible explanations. She probably has already thought of alternative possibilities, but she now tries to explicitly elicit and search her memory and clinical experience using a series of interrogatives. The case presents a puzzling picture, so the clinician decides initially to cast her net widely. She asks herself which other diagnosis could explain the symptoms. Her first concern is to make sure she has not forgotten to think of a potentially dangerous condition. However, she needs to quickly limit herself to a short list of actionable diagnoses. Trying to actively consider and rule out all disease candidates is not practically possible, especially when the patient is unstable. Thus, the physician asks herself further questions to limit and focus her search, prioritizing conditions that (i) are common for this type of patient, (ii) would be dangerous if left untreated and (iii) allows of effective testing and treatment.

In this case the clinician chooses to generate hypotheses based on her own experience and background knowledge. Sometimes the generation of hypotheses happens almost automatically: the clinician recognizes a known pattern and immediately recalls the most common and important diagnoses. In more atypical cases, it can be necessary to employ a more structured or directed form of thinking. Other possibilities would have been to use one of the existing artificial intelligence programs designed to assist medical diagnosis (e.g. Isabel, DXplain) or request a colleague to review the data and offer a second opinion. All of these would classify as different strategies for hypothesis generation.

Current computer programs are based on the prevalence and weighting of signs and symptoms linked to diseases. They indicate which conditions are most the common causes of the symptoms entered into the program and red-flag potentially life-threatening diagnoses. The cardinal arrangement of the diagnoses is based on the experience of the writers of the program and the epidemiology of diseases commonly encountered in the indicated age group, sex and geographical region. Using a computer program can be helpful for reminding a clinician of rare but dangerous conditions. Many experienced physicians however consider them of limited usefulness.[10]

First, they tend to generate a fairly long list, which is not particularly helpful in an emergency situation. Trying to work through an extensive list of possibilities is not a feasible strategy especially when the patient is unstable. It may also subject the patient to unnecessary and potentially harmful over-testing. Second, physicians do not know how the program assigns weights to the different signs and symptoms and to the prevalence of the disease. The computer program is based on geographically common epidemiologic findings in specific diseases and populations, but this population level information cannot be translated directly to the individual case. Experienced physicians will be attuned to the concrete clinical setting (how stable is the patient, what are the urgent problems), details of the case (e.g. medical history, country of origin, foreign travel, use of drugs, smoking) and the patient's response to therapy (e.g. pain relief and normalization of heart rate, breathing). These facts have to be interpreted and the physician has to judge whether or not the findings are properly perceived and integrated into the diagnostic picture. So, physicians will in any case need to draw on their experience and insight to interpret the results.

An important part of this is the ability to recognize the important signs and symptoms, i.e. to recognize *that* a certain fact, or combination of facts, is important information which needs to be taken into account. For instance, had a patient travelled to sub-Saharan Africa, this might suggest anemia as a result of malaria; sedentary life style would place coronary ischemic heart disease higher in the list of diagnoses; family history of diabetes in mother and grandmother would interpret weight gain, hypertension, and dyslipidemia into the spectrum of metabolic syndromes. Of course, neither a computer nor a human reasoner can take a given piece of evidence into account before it is recognized *as* evidence. As with hypothesis generation, trying to take into account every single piece of information is not feasible, nor necessarily very efficient. Perspicacious observation is here an intricate part of the reasoning process itself. As Peirce noted, "abductive inference shades into perceptual judgment without any sharp line of demarcation between them" (1932-1958, §5.181). These computer programs are, at best, an aid to, rather than a replacement for, clinical judgement and experience.

To have a colleague review the data, whose difference in experience, training and background knowledge may lead him to think of other possibilities, can also be a way to ensure that important diagnoses are not overlooked. At other times, it grows out of the physician being confused by the picture or even dissatisfaction with her own thought process. Although this strategy for generating hypotheses cannot guarantee to be as exhaustive as a computer program, drawing on the judgement of a colleague has the advantage of being better attuned to the concrete clinical situation, and so being more likely to generate suggestions that are plausible and useful in context.

The experience and training of the clinician, or her colleagues, play a crucial role in several respects in the selection of a differential diagnosis. First, the clinician has to make a wise choice of which strategy for generating new suggestions strikes the right balance between expediency, exhaustiveness and quality in the concrete situations. Although one can discuss general considerations of advantages and disadvantages of different strategies, as above, the choice ultimately has to rely on the judgement of the physician. Second, it is the training, experience and background knowledge of the clinician that allows her to recognize patterns and recall possible diagnoses. Finally, in choosing which diagnostic hypothesis to focus on the clinician needs to judge which diagnoses are most likely in the concrete case and decide how to weigh this, e.g., against the seriousness and urgency of the disease, its testability and its treatability.

Prioritize hypothesis for testing: Cardiac enzymes after two hours are borderline elevated, with c-Troponin is 98.5% of the normal range. The clinician ponders if perhaps the origin of the troponin elevation is from the epicardium or the pleura, rather than the heart muscle, reassessing the hypotheses of enzyme origin and cause of chest pain. She requests a cardiologist consultation. Meanwhile, she is at the bedside. She next considers pulmonary embolism and orders a chest-computed tomogram (CT-scan) with contrast media to search for the embolism (the blockage).

Commentary: The clinician now decides to request a second opinion from a specialist colleague. Meanwhile she prioritizes the pulmonary embolism hypothesis for testing. There are several good reasons for this. First, she currently considers pulmonary embolism one of the most likely diagnoses. Second, pulmonary embolism and acute aortic syndromes are both emergency conditions and would require immediate treatment. Pericarditis is also a very serious condition but less urgent, whereas gastrointestinal reflux disease is not an immediate threat. Third, a CT-scan is a highly reliable way to detect embolism. Fourth, the chest CT-scan might also show a widened mediastinum (the area containing the heart and the pericardium), a possible sign of pericarditis. It would also show the thoracic aorta (the part of aorta situated in the chest-region), a possible site of any aortic syndrome. This last point is an example of how a test can have other epistemic consequences in addition to testing hypotheses directly, in this case by potentially providing clues for future hypothesis generation. In sum, the CT-scan would be a reliable test of the most likely and serious condition while potentially also providing information relevant for the two other most serious conditions.

Further puzzling results: The results of the CT-scan adjusted to an early phase of contrast injection are reported as negative for pulmonary embolism, but the ascending aorta is reported to be prominent, measuring 4.3 cm in diameter (normal ascendingaorta is 3.63 to 3.91 cm). The patient still complains of chest pain but feels relieved by increased dose of morphine, and breathing is improved somewhat by continuous 100% oxygen therapy.

Consulting cardiologist arrives. He reviews the history and testing and is still convinced the patient has cardiac ischemia. Given the negative results of the CT-scan he judges that the picture is atypical but consistent with ischemia, probably caused by coronary artery disease. He also considers the other available results. Although the CT-scan has ruled out pulmonary embolism, the prominent aortic valve shadow is worrisome. He requests a transthoracic echocardiogram (TTE).

Commentary: Since the cardiologist takes pulmonary embolism to be ruled out by the CT-scan, he still considers cardiac ischemia most likely. However, the shadow on the aortic root is puzzling and he decides this merits further investigation.

Conclusion of scenario: The TTE shows a widened mediastinum and that the aortic root is dilated to 4.5 cm. The patient's condition is unchanged. Because of the degree of clinical pain, the cardiologist and emergency room clinician decide that immediate coronary artery intervention (stent or bypass) is necessary. They consult with the nearest cardiac surgical unit for immediate transfer and transport helicopter arrives. The cardiologist accompanies the patient to the surgical unit.

Upon arrival, the cardiologist is still concerned about the diagnosis. He decides to review the inflight recordings and TTE together with the other available clues. He returns to the patient, listens for the diastolic murmur reported at initial examination, notes that the pain was non-stress induced and considers the dilated aortic root. He tries to think of an alternative diagnosis which could integrate all of these clues, thinking through different possible aortic conditions, and realizes that a dissecting thoracic aortic aneurysm[11] could explain all of these symptoms: the dilated root is part of the aneurysm and a small flap in the beginning tear could produce the murmur. The patient is taken directly to surgery where ascending (type A) aortic dissection is repaired. He was discharged home after one week.

Commentary: Given the state of the patient, the physicians are forced to act even though the evidence remains puzzling. The cardiologist does not consider the diagnosis of cardiac ischemia particularly likely due to the lack of expected observations (minimal pain relief from morphine, non-diagnostic cardiac enzymes and ECG). But he currently lacks a plausible alternative.

He chooses a strategy for generating an alternative diagnosis, deciding to review all of the available clues, including ones that initially were not considered salient (the soft murmur), to guide his search for alternative diagnoses. Like the emergency room clinician, given the negative result for embolism, he considers the most serious remaining possibility an acute aortic syndrome, also suggested by the dilated aortic root. Relying on his background knowledge, he directs his attention towards possible aortic syndromes and quickly thinks of a possibility—a dissecting aneurysm—capable of explaining the symptoms. Once he has in mind this newly generated hypothesis he immediately recognises that it would be able to explain all of the otherwise puzzling evidence. On this basis, he judges it more likely than ACS and decides to adopt it as a basis for the surgical intervention. While one cannot guarantee such judgements to always be correct, given the urgency of the situation, it was in this case a reasonable and, fortunately, successful approach.

5. Limitations of Current Accounts of Medical Diagnosis

Throughout the preceding case study, we have discussed a number of decisions made by the physicians regarding the generation and pursuit of diagnostic hypotheses. In the follow two sections we, first, highlight some limitations of the two primary frameworks used for discussing diagnostic reasoning in the medical literature: (i) the normative, probabilistic framework associated with the threshold approach and (ii) descriptive frameworks based on cognitive psychology. Second, we then offer our constructive proposal: to conceptualize the process of diagnosis in terms of strategic reasoning.

The probabilistic framework is the most popular normative framework employed in the methodological discussions of diagnosis in the medical literature, especially among proponents of evidence-based medicine. This approach is typically summarized as follows (e.g. Richardson and Wilson 2015): First, the physician identifies a plausible differential diagnosis for the patient and assigns an initial prior (or "pretest") probability to each of the hypotheses in the differential diagnosis. Second, the clinician compares the initial probabilities of the hypotheses to the probability thresholds, as determined by the decision-theoretic models of threshold approach, in order to decide whether to test or treat for the disease. Third, as test-results become available, the clinician should use Bayes' Theorem together with information about test reliability to update their probabilities.

While this probabilistic framework can highlight important lessons for clinical reasoning,[12] it does not provide a general framework for explicating clinical reasoning; the probabilistic approach presents an idealized, simplified picture of clinical decision-making which leaves out many important aspects of the process of diagnosis. In our case study, factors that eventually led to successful diagnoses included: (i) decisions by the emergency room clinician and later by the cardiologist about when to generate more diagnoses for consideration; (ii) choosing effective and efficient strategies for generating relevant hypotheses; (iii) recognizing whether the generated hypotheses can explain the salient symptoms; (iv) recognizing the importance of subtle clues, such as the dilated aortic root or the diastolic murmur, that may initially appear puzzling or unimportant, as well as knowing which features (most of them unmentioned in our description of the case) to ignore; (v) the strategic choice of test (the CT-scan) which could reveal important information for further inquiry even if it failed to confirm the hypothesis tested.

This last point is crucial. The decision-theoretic models of the threshold approach are limited to considering the direct benefits and harms of testing or treating. They do not take into account the kinds of downstream consequences highlighted in Section 3. However, these considerations proved crucial to the successful resolution of our case: the CT-scan produced the crucial clue that eventually led the cardiologist on the right track. Considerations of this type are difficult to represent directly in the probabilistic framework, since it is difficult to assign meaningful probabilities or utilities to these unknown unknowns. What is the probability that a given test will produce a valuable clue for a diagnosis we have not thought of yet? What is the utility of treating this as-yet-unknown disease? Successful diagnosis depends, in part, on recognizing and considering these possibilities. Of course, one can always add a term into the decision-theoretic calculus to represent the weight these considerations are given relative to the direct consequences of testing/treating. But this does not represent the reasoning that leads physicians to give them that weight.

Finally, probabilistic models start from the assumption that one has already formulated a diagnostic hypothesis. In its current form, it only addresses the question of whether the hypotheses generated satisfy the goal of being pursuit-worthy. To the extent that it succeeds in the latter, it at best represents the *aim* of generative reasoning, rather than describing this reasoning in itself.

When hypothesis generation is discussed in the medical literature, it is done primarily within the framework of cognitive psychology. For instance, while Kassirer, Wong and Kopelman (2010, Ch. 13), discuss hypothesis generation in several case studies, their focus is on which structures of memory allow (or prevent) physicians from recalling the correct diagnosis—e.g. perhaps the physician's memory is structured in condition-action pairs, one of which state (say) that IF an adult has a high serum cholesterol value, THEN consider the possibility of hypothyroidism (*ibid.* , 75)? While much can no doubt be learned about generative reasoning from cognitive psychology, these analyzes currently lack a guiding normative framework. Correct diagnosis of course requires physicians to have structures of memory which allow them to recall the correct diagnosis, but as a normative account it amounts to asking "how can we recall the right diagnosis?" As we have argued, the relevant question is rather: which strategies for hypothesis generation allow physicians to generate a manageable set of hypotheses that are most important to consider at the given stage of inquiry?

6. Diagnosis as Strategic Reasoning

The framework we want to outline in this section does not aim to rival the probabilistic approach in formal rigor. Rather, we want to outline a more flexible, general framework for thinking about diagnostic reasoning, drawing on Hintikka's (1998) suggestion that abduction can be understood as in terms of *strategic reasoning* . Hintikka's proposal is based on an analogy between game-theoretic reasoning and scientific inquiry. Knowing how to play a game such as chess involves at least two kinds of knowledge. The *definitory rules* tells us what kinds of moves are allowed and what the consequences are of those moves. The *strategic principles* , by contrast, tell us what would be a wise or an unwise move in a given situation, i.e. whether the move is likely to help us achieve the goals of the game.

In medical diagnosis, the definitory rules involve knowledge of how of a given diagnostic hypothesis should be evaluated in light of current evidence, what the available tests are, how the hypothesis should be re-evaluated in light of different possible test results and what the potential consequences are for the health of the patient. Based on this, the clinician has to adopt a *diagnostic strategy* . By this we mean an overall strategy for how to generate hypotheses, select a differential diagnosis and prioritize hypotheses for testing. As Hintikka emphasizes (1998, 513), one usually has to evaluate entire strategies, rather than individual steps. This is because it is often only possible to evaluate the importance of potential consequences—e.g. providing clues for hypothesis generation—within the context of a broader strategy.

We propose to see diagnostic reasoning as two-tiered. Individual moves (ordering a given test, choosing to stop generating new hypothesis, etc.) are justified in terms of whether they contribute to an overall strategy. The crucial choice then concerns which strategy to pursue. Diagnostic strategies can be thought of as analogous to the strategies a seasoned chess player might employ.[13] Choosing a chess strategy depends, in part, on what kind of opponent you think you are up against together with knowledge of the definitory rules. Similarly, the choice of a diagnostic strategy will be informed by what kinds of diseases the physician thinks are most likely. In some very simply cases it may be possible to represent this in an explicit decision-theoretic model. In this sense the threshold approach is not incompatible with the broader framework proposed here. However, for the reasons given above, in many cases a physician's reasoning about what the best strategy is in the given clinical context will not be adequately captured by any generally applicable formal model. This does not mean that we cannot say something about what kinds of considerations are involved in this type of reasoning.

In our case study, we identify three crucial choices of strategy. The first is the initial choice to pursue the

ACS hypothesis before systematically generating new hypotheses. The emergency room clinician is trying to achieve a quick resolution, to spare the patient from potential harm of leaving the condition untreated and from unnecessary testing. This choice is in part justified by what diagnosis she thinks is most likely. In the chess analogy, a player may push for a quick checkmate because she thinks the opponent is likely not to recognize a certain trap. Similarly, the clinician recalls the most common cause of the presenting symptoms (ACS) and knows of tests which, if positive, would quickly and conclusively verify the hypothesis. This strategy in turn justifies not systematically generating further hypotheses. *Given* the strategy of achieving a quick resolution, it was reasonable to stop generating new hypotheses once she had identified what was—given the evidence—the most likely cause and realized that this could be quickly and reliably tested. Spending more time generating hypotheses in this context would have been unnecessary.

Unfortunately, many opponents—whether chess players or diseases—will not be defeated by such a direct strategy. When the initial tests fail to confirm the ACS diagnosis, putting the clinician in a more uncertain situation, she adopts a new strategy. Since she no longer has a clear view of what the likeliest diagnosis is, her priority shifts to ruling out the most serious threats, hoping in the process to discover—or better: create—a "strategic opening", that is, a clue which could lead to the correct diagnosis. Adopting this strategy, she attempts to systematically recall the most dangerous alternative causes of chest pain. She chooses a test which is both highly reliable for ruling out such alternative (pulmonary embolism) and might enable future moves, by producing clues for further hypothesis generation. Her decision to request a cardiologist consultation at this stage also makes sense in light of this strategy, since his expertise would (i) complement her ability to think of relevant hypotheses and (ii) enable him recognize the relevance of any emerging clues.

Finally, at the concluding stage of the case, the cardiologist adopts a strategy for generating hypotheses that is focused on the salient clue—a dilated aortic root—brought out by the CT-scan and the echocardiogram, as well as the puzzling diastolic murmur. Due to the worsening state of the patient there is no time for further testing of the hypotheses thus generated. Whether to retain the ACS diagnosis or to adopt a newly generated hypothesis has to depend on his clinical judgement of which hypothesis best 'fits' the clinical picture. He therefore chooses a strategy of thinking quickly through a range of hypotheses, counting on his experience to allow him to recognize the correct hypothesis when he 'sees' it. Given his specialization as a cardiologist, and the severe constraints of the situation, considering possible aortic syndromes with a focus on explaining the dilated root and the murmur was a reasonable—and as it turned out successful—way of implementing this strategy.

7. Conclusion

In this paper, we have outlined a general framework for analyzing diagnostic reasoning, based on C. S. Peirce's notion of abduction. We have distinguished between reasoning concerned with generating, pursuing and accepting/rejecting diagnostic hypotheses. Through our case study we have highlighted the crucial role played by clinical experience and judgement at each of these stages. Finally, we have critically evaluated currently existing frameworks for conceptualizing diagnostic reasoning, and proposed that diagnosis can be fruitfully thought of in terms of strategic reasoning.

As illustrated in Section 6, the latter framework allows us to naturally describe the kinds of reasons that led to successful diagnosis in our case study. We do not claim this to be prescriptive: while we have explicated reasons which in the concrete situation made the strategies adopted by the physicians *reasonable* , we do not claim that these represent the *best possible* strategies. However, we believe that our framework can contribute to a better normative understanding of diagnostic reasoning as it occurs in existing clinical practice. By allowing us to identify and discuss the prescriptive limitations of different diagnostic strategies, it provides a basis for evaluating proposed improvements of clinical practice and for teaching diagnostic reasoning.

Notes

[1] Thus, as recent scholarship has emphasised (Paavola 2006, McKaughan 2008, Campos 2011), abduction should not be identified with the so-called Inference to the Best Explanation.

[2] The point has been raised by Frankfurt (1958) and Kapitan (1997). McKaughan (2008) defends the pursuit-worthiness interpretation against generative interpretations. Schaffner (1993) similarly criticizes Hanson (1958) for obscuring the distinction between the generation and preliminary evaluation of hypotheses.

[3] Interpretations along these lines have been proposed by Fann (1970), Curd (1980) and Psillos (2011).

[4] Similar objections were raised against Hanson (1958) and others who defended abduction as a "logic of discovery." See Nickles (1980) for several contributions to this debate and Schaffner (1985) for ones focused on medical diagnosis.

[5] Stanley and Campos (2013, 306) call the former of these "creative abduction" and the latter "habitual abduction."

[6] Sometimes 'differential diagnosis' is instead used to refer to the process or method of considering and distinguishing different diagnostic hypotheses (e.g. Sox, Higgins and Owens 2013, Ch. 2). We will here use the term only to refer to a set of competing hypotheses, rather than the process of generating or selecting between these.

[7] A similar point also applies to treatment: even if a given treatment fails to alleviate a patient's symptoms, it may still provide valuable clues for further investigations. The line between treatment and testing is not always a sharp one.

[8] The following case is developed on the basis of the clinical experience of one of the authors. While we do not make any claims as to how statistically common scenarios of this type are, we regard it as sufficiently typical to illustrate the ability of our framework to explicate patterns of clinical reasoning.

[9] 'Decrescendo diastolic murmur' means that a sound of decreasing intensity occurs during diastole, i.e. period of the cardiac cycle where the heart relaxes and refills with blood.

[10] For a recent survey of currently available programs, see Bond et al (2012). Philosophers in the 1980s debated whether computer programs could in principle replace all aspects of diagnostic reasoning (e.g. Schaffner 1985, Wartofsky 1986). We here focus on how useful currently existing programs are for the task of generating hypotheses.

[11] An aneurysm is an abnormal widening of a blood vessel. This can cause weakness in the wall of the vessel. A dissection is a rupture of the blood vessel where blood flows into the layers of the wall of the vessel, forcing them apart.

[12] For instance, threshold models highlight the importance of weighing initial probability against the potential benefits and harms of testing or treating. Similarly, Bayes' Theorem can highlight importance lessons about probabilistic reasoning. Thus, they may provide useful analytic frameworks for teaching clinical reasoning (Sox, Higgins and Owens 2013). We are less optimistic about proposals to reform clinical practice to conform more closely to probabilistic models, proposed e.g. by Richardson (2007); see Marewski and Gigenrenzer's (2012) critique of information-greedy procedures for clinical decision-making.

[13] Kassirer, Wong and Kopelman, (2013, 46) also mention aanalogies between expertise in chess and in medical diagnosis.

Acknowledgments

We would like to thank Simon Barker, Sami Paavola, Nancy Cartwright, Julian Reiss and Sarah Wieten for their comments on earlier drafts, as well as the Library Staff at Maine Medical Center, Georgia Wark, Dina McKelvy and Heather Kemp, for help with the bibliography. The paper has been presented at the 2016 Meeting of the Nordic Network for Philosophy of Science and to the CHESS research group at Durham University. The discussion at both events was very helpful. The paper is fully joint work.

References

Bond, W. F., L. M. Schwartz, K. R. Weaver, D. Levick, M. Giuliano, and M. L. Graber. 2012. Differential diagnosis generators: An evaluation of currently available computer programs. *Journal of General Internal Medicine* 27: 213-9.

Campos, D. G. 2011. On the distinction between Peirce's abduction and Lipton's inference to the best explanation. *Synthese* , 180: 419-42.

Chiffi, D. and R. Zanotti. 2015. Medical and nursing diagnosis: A critical comparison. *Journal of Evaluation in Clinical Practice*21: 1-6.

Curd, M. 1980. The logic of discovery: An analysis of three approaches. In *Scientific discovery, logic, and rationality,* ed. Thomas Nickles, 201-219. Dordrecht and Boston: Reidel.

Darwin, C. 1958. *The autobiography of Charles Darwin 1809-1882. With the original omissions restored.* Edited by N. Barlow. London: Collins.

Djulbegovic, B., J. van den Ende, R. M. Hamm, T. Mayrhofer, I Hozo, and S. G. Pauker. 2015. When is it rational to order a diagnostic test, or prescribe treatment: The threshold model as an explanation of practice variation. *European Journal of Clinical Investigation* 45: 485-493.

Fann, K. T. 1970. *Peirce's theory of abduction* . The Hague: Martinus Nijhof.

Frankfurt, H. 1958. Peirce's notion of abduction, *Journal of Philosophy* 55: 593-597.

Hanson, N. R. 1958. The Logic of Discovery. *Journal of Philosophy*55: 1073-1089.

Hintakka, J. 1998. What is abduction? The fundamental problem of contemporary epistemology. *Transactions of the Charles S. Peirce Society* . 34: 503-533.

Kapitan, T. 1997. Peirce and the structure of abductive inference. In:*Studies in the logic of Charles Sanders Peirce* , N. Hauser, D. D. Roberts & J. Van Evra (eds.). Bloomington and Indianapolis: Indiana University Press. 477-96.

Kassirer, J. P., J. B. Wong, and R. I. Kopelman. 2010. *Learning clinical reasoning* (2^{nd} ed.). Philadelphia: Lippincott Williams & Wilkins.

Marewski, J. N. and G. Gigenrenzer. 2012. Heuristic Decision Making in Medicine. *Dialogues in Neuroscience* 14: 77-89.

McKaugan, D. J. 2008. From ugly duckling to swan: C. S. Peirce, abduction, and the pursuit of scientific theories. *Transactions of the Charles S. Peirce Society* 44: 446-468.

Nickles, T., (ed.) 1980. *Scientific discovery, logic, and rationality* (Boston Studies in the Philosophy of Science, Vol. 56). Dordrecht and Boston: Reidel.

Nyrup, R. 2015. How explanatory reasoning justifies pursuit: A Peircean view of IBE. *Philosophy of Science* 82: 749-760.

Paavola, S. 2004. Abduction as a logic and methodology of discovery: The importance of strategies. *Foundations of Science* 9: 267-283.

Paavola, S. 2006. Hansonian and Harmanian abduction as models of discovery. *International Studies in the Philosophy of Science* . 20: 93-108.

Pauker, S. G. and J. P. Kassirer. 1980. The threshold approach to clinical decision making. *The New England Journal of Medicine*302: 1109-1117.

Peirce, C. S. 1932–58. *Collected papers of Charles Sanders Peirce* . 8 vols. Edited by P. Weiss, C. Hartshorne, and A. Burks. Cambridge, MA: Harvard University Press.

Pietarinen, A.-V. & F. Bellucci. 2014. New light on Peirce's conceptions of retroduction, deduction and scientific reasoning. *International Studies in the Philosophy of Science* . 28: 353-373.

Psillos, S. 2011. An explorer upon untrodden ground: Peirce on abduction. In *Handbook of the history of logic, Volume 10: Inductive logic* , D. M. Gabbay, S. Hartmann, and J. Woods (eds.). Oxford and Amsterdam: North-Holland. 117-151.

Richardson, W. S. 2007. We should overcome the barriers to evidence-based clinical diagnosis! *Journal of Clinical Epidemiology* 60: 217-227.

Richardson, W. Scott, M. C. Wilson, G. H. Guyatt, D. J. Cook, and J. Nishikawa. 1999. Users' guides to the medical literature XV. How to use an article about disease probability for differential diagnosis. *JAMA* 281: 1214-1219.

Richardson, W. S. and M. C. Wilson. 2015. The Process of Diagnosis. In *Users' guides to the medical literature: Essentials of evidence-based clinical practice* (3^{rd} ed.), G. H. Guyatt, D. Rennie, M. O. Meade, and D. J. Cook (eds.), New York: McGraw Hill. 211-222.

Schaffner, K., (ed.) 1985. *Logic of discovery and diagnosis in medicine* . Berkeley & Los Angeles: University of California Press.

Schaffner, K. 1993. *Discovery and explanation in biology and medicine* . Chicago: University of Chicago Press.

Stanley, D. E. and D. G. Campos. 2013. The logic of medical diagnosis. *Perspectives in Biology and Medicine* , 56: 300-15.

Stanley, D. E. and D. G. Campos. 2016. Selecting clinical diagnoses: Logical strategies informed by experience. *Journal of Evaluation in Clinical Practice* 22: 588-597

Upshur, R. 1997. Certainty, probability and abduction: Why we should look to C. S. Peirce rather than Gödel for a theory of clinical reasoning. *Journal of Evaluation in Clinical Practice* 3: 201-206.

Sox, H. C., M. C. Higgins, and D. K. Owens, 2013. *Medical decision making* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Wartofsky, W. M. 1986. Clinical judgement, expert programs, and cognitive style: A counter-essay in the logic of diagnosis. *Journal of Medicine and Philosophy* 11: 81-92.