5. Fulfilling Desiderata

In the preceding chapters, I developed my model-based account of diagnostic reasoning in psychiatry. In this chapter I want to let it do some work by showing that it not only meets the adequacy conditions to for an answer to the Methodological Question, as suggested by the end of the last chapter, but in addition fulfils the desiderata that I set out in the Introduction. These desiderata were that the proposed answer to the Methodological Question should:

1. provide a comprehensive account of the core aspect of the process of psychiatric diagnostic reasoning
2. present a proposal to us that is cognitively realistic, thus can take place in actual diagnostic efforts
3. make sense of the difference between misdiagnosis and diagnostic malpractice in psychiatry
4. explain the occurrence and resolution of diagnostic uncertainty in psychiatric clinical diagnostics
5. explain the phenomenon of good instinctual diagnosis and what is problematic about it
6. explain the occurrence and resolution of diagnostic disagreements over time within and between experts
7. provide guidance for thinking about how changes in psychopathology may be integrated with or change the methods of diagnostic reasoning.

These desiderata were proposed to be relevant to address in a proposal for answering the Methodological Question since they show that the proposal is a helpful guide either to attaining a basic grasp of psychiatric diagnostics itself, or to understanding more specific aspects of (and phenomena in the context of) diagnostic reasoning that are commonly encountered and thus useful to explain. Let us briefly recap the relevance of each of the desiderata.

A proposal for answering the Methodological Question should ideally provide a comprehensive account encompassing all aspects of the diagnostic process and leaving no central aspect unexplained. It should ensure that its proposal is within

the general capacities of a psychiatrist to be carried out as a realistic person-level cognitive process, and thus can be taken as a realistic method (i.e., a learned belief-forming procedure) that psychiatrists may pursue in their everyday diagnostic clinical work. The proposal should also enable us to understand the occurrence and resolution of diagnostic uncertainties and disagreements. Mistakes in diagnostics unfortunately occur and differentiating between mere misdiagnosis and actual malpractice is of high ethical and legal relevance. To get a hold on "diagnostic instincts" seems important since everyone who has ever worked in a clinical context will have seen experienced clinicians shooting diagnostic guesses from the hip who, more often than not, seem to be right, so that it is relevant to have a well-founded attitude towards how this form of diagnostics works and why it is (or is not) credible. Finally, to make sense of the possibility of integrating into diagnostic practice ongoing changes in our understanding of psychopathology, as well as to speculate as to what the future of diagnostics might mean for our current methos of diagnostic reasoning, is central to showing the theory's plausibility in terms of its responsiveness to change. It should be robust in that it allows us to explain how current diagnostic reasoning integrates minor changes, but sensitive enough to large-scale changes to diagnostics to be falsifiable, otherwise it would be too generic. In the following section, I will discuss how my answer to the Methodological Question enables us to meet all the desiderata listed above.

4.1 Comprehensiveness

For a proposal to address the Methodological Question in a comprehensive manner requires it two do two things. It requires the proposals descriptive suggestion of a method as part of the methodology to leave no relevant aspect of the diagnostic reasoning process unaddressed and to make sense of its different aspects with a reasonable degree of detail. To meet these two requirements is what would make the proposal comprehensive. Whether my own proposal, the model-based account of psychiatric diagnostic reasoning, meets the criterion of comprehensiveness depends on two things. First, it depends on whether one accepts my basic account of the process of clinical psychiatric diagnostics as the proper core procedure of contemporary diagnostic reasoning, as presented in the first chapter and via a more example-oriented treatment in the third chapter. Second, meeting this criterion depends on whether one accepts that the attempt to map my understanding of diagnostic modelling as laid out in the second chapter, plus my limited additional remarks about how the case formulation (as a composition of modelling outcomes) and the disorder diagnosis (as pattern recognition) maps onto the described process of clinical psychiatric diagnostics indeed explains the described diagnostic reasoning process on a sufficient level of detail. The reasons why I believe that my presentation of the

clinical diagnostic process is adequate were presented in the first chapter, and the considerations that make me think that the proposed method of modelling and pattern recognition maps onto psychiatric diagnostics, have just been laid out in Chapter 3, so I will simply reiterate my previous points here in a more abstract fashion.

The first aspect of ensuring that my proposal to address the Methodological Question meets the criterion of comprehensiveness involves checking that I provided my attempt to answer it with an adequate starting description of psychiatric diagnostics – in other words, a description that itself has an adequate scope and explores the process in relevant depth. To ensure that it has an adequate scope, as discussed in more detail in earlier chapters, I considered a recent edition of widely regarded psychiatric training literature that is intended to lay out the general core procedures of clinical psychiatric diagnostics, as well as recent guidelines of psychiatric expert societies. Focusing on those sources was meant to ensure the proper scope for what I consider to be the constitutive core procedures of proper, contemporary, clinical psychiatric diagnostics. While my approach to account for the overall diagnostic procedure in Chapter 1 did not delve into too much detail for specific cases but rather provided an overview, Chapter 3 provided several clinical examples in line with my general understanding in a more illustrative fashion. This more detailed presentation in Chapter 3, with the more general architecture from Chapter 1 in the background, provided a foundation on which I then attempted to demonstrate the mapping between my model-based proposal and pattern recognition in the diagnostic process.

To ensure that my efforts to establish my proposal turn out to be a comprehensible account of psychiatric diagnostics, I went through all phases of the diagnostic process initially identified in Chapter 1 to map onto it all aspects of the method I had claimed take place. Thereby I outlined how we should understand the relevant facets of each stage of psychiatric diagnostics in light of the method I proposed. To briefly take one example, I opened my discussion with the first step of the diagnostic process, the screening phase. This phase is meant to enable the psychiatrist to recognise a patient's complaints based on previous assumptions about what is within the range of normal psycho-behavioural features, such that deviations of a patient from these assumed states might indicate the presence of a psychiatric symptom and are thus identified as complaints, which further down the road, in the in-depth evaluation, will be evaluated to decide whether they are indeed a psychiatric symptom or not. I illustrated this step in detail, moving from a generalising description of this step to concrete clinical examples. In my attempt to map diagnostic modelling onto psychiatric diagnostics, I proposed that this step in the diagnostic process and its different aspects is equivalent to the initial error-recognition step in diagnostic modelling. I argued that the background assumption of the psychiatrist to discover complaints equals the normative model based on which initial error recognition identifies *prima facie* errors, and that the complaints identified by the psychia-

trist equal these *prima facie* errors, being discovered using the normative model and later evaluated via the diagnostic procedures. I then discussed the realisation of this process in the concrete clinical examples I provided.

This exemplary step from my work shows that I described the clinical diagnostic process in general terms, to a degree of detail where the next best step to offer further detail was to provide concrete case examples. In other words, I described the clinical diagnostic process to the lowest still general level of detail in which I could describe it before transitioning to single cases. It therefore seems that the mapping of the method onto the process whose description is provided on this level of detail is as comprehensive as it can become before forfeiting its claim to allow us to discuss the diagnostic procedure in general. Hence the discussion of stages, aspects, and the functional connections between them in psychiatric diagnostics, and the fact that everything I claimed about diagnostics was mapped onto my application of the method of model-based diagnostics (just as I did in the brief excerpt of my efforts just discussed), together seem to justify the assessment of my answer to the Methodological Question as comprehensive.

4.2 Cognitive Realism

To ensure that an answer to the Methodological Question is not only in principle adequate to match the requirements to qualify as an answer to the Methodological Question, it should also be realistic – that is, be a procedure that could plausibly be carried out as a learned person-level procedure by real clinicians doing diagnostic work. Only then can it qualify as a method (i.e., a learned belief-forming procedure) that could be the actual cognitive work undertaken by clinicians. In other words, the proposal should be cognitively realistic.

To see whether my model-based account presents a realistic proposal, we need to ensure that it proposes a format of reasoning that seems to equal what commonsensically takes place in clinicians' minds when they think about their patients. Regarding the requirements on information-processing, the amount should not exceed what can plausibly be assumed to be within the capacity for cognitive load of diagnostic experts. In addition, since a method is (as discussed in the Introduction) a learned belief-forming procedure, it should be *prima facie* realistic that the way diagnostics take places according to the proposed method, and thus following the rules of the method, should be something that can plausibly be learned.

Let us begin with the format. The chief format of diagnostic reasoning that I am proposing is qualitative reasoning in the form of propositions that contain diagnostically relevant information. *Prima facie* this seems to fit well with what psychiatrists do. As I said earlier, clinicians do not calculate the diagnoses of their patients. Rather, when we look at conversations between clinicians speaking about patients,

or when, as discussed earlier, we look at diagnostic exercises or research involving diagnosing clinicians using think-aloud protocols, we usually find them engaging in diagnostic reasoning in terms of normal-language sentences, describing diagnostic requirements as well as information about the patient and deciding which of these propositions apply and what to infer from that. It thus seems that my account considering propositional models as information bearers and as vehicles of diagnostic reasoning matches well with what we find in clinical diagnostics, when it comes to describing the process on the personal level of the psychology of clinicians and their intentional efforts to evaluate patients.

When we think about the cognitive load associated with this proposal, it seems bearable. Of course, the psychiatrist does not have all potentially relevant propositions that might become relevant in the diagnostic process present in their working memory at the same time, but they are present in the background knowledge base resulting from the psychiatrist's education. When carrying out the screening procedure, for example, psychiatrists systematically explore the different aspects of the patient's life, bearing in mind the propositions of the aspect of the normative model that is being compared with the patient's psycho-behavioural functioning in this area. If the patient spontaneously reports complaints, the psychiatrist entertains the normative propositions relevant for the relevant aspect of the psycho-behavioural presentation of the patient and compares the complaint with the propositions. The same goes for the diagnostic propositional models. The psychiatrist never has all of them at the forefront of their mind all the time, but a recognised complaint will trigger the recall of potential diagnostic options that are all connected with diagnostic model structures whose content can be entertained and used to guide in-depth evaluation if needed. Furthermore, the inferences from present patterns of symptoms to an adequate diagnosis are (if carried out by a clinician who has learned the diagnostic manual) made not by calling to mind all disorders and their symptoms, but by recalling the adequate disorder diagnosis based on a certain set of previously identified symptoms present. Thus, cognitive load is managed by bringing only what is needed into the psychiatrist's immediate cognitive workspace. This management process is further supported by documenting (taking clinical notes on) steps of the diagnostic process to ensure that once made, inferences and their outcomes do not get lost.

Finally, the overall intentional person-level procedure of diagnostics that is carried out in this way also appears to be something that can be learned and that thus qualifies as a method. Nobody is born a diagnostic expert. Psychiatrists acquire their psychopathological and general medical knowledge base through their studies and clinical experiences and learn how to use it in a diagnostic process by consulting training literature and gathering clinical practice in which they are supervised in carrying out the stepwise process. They are taught what information about the patient may indicate which psychiatric or medical problem, what further information

is needed to assess these options, and how they can generate this information in contact with the patient. All this and the further steps of the overall diagnostic procedure are taught to psychiatrists, which is possible because they can be told what to consider and which actions to take and not take as part of the diagnostic process. Because they can express what they had in mind when they attempted to provide a diagnosis and what their reasoning was in considering option (a) rather than option (b), they can be corrected in their reasoning and action, and so come closer to embodying the proper method of psychiatric diagnostics. Although nobody tells psychiatrists about normative models, propositional diagnostic models, or *prima facie* errors as part of their education, and they thus do not learn the method on a theoretical level, they do learn to carry out the diagnostic procedures such that by following these procedures they indeed follow the standards of the method of proper clinical diagnostic reasoning.

In sum, it seems that my proposal of the model-based account manifests all aspects of cognitive realism. It requires a plausibly manageable format and cognitive load from clinicians, and it appears that the method used is something that can be learned as part of clinical training. Thus, the desideratum of cognitive realism is fulfilled.

4.3 Misdiagnosis and Diagnostic Malpractice

Medical diagnosis is fallible. A diagnosis given to a patient by a diagnostic expert in any field of medicine can be wrong. The reasons why a wrong diagnosis can be made are numerous, from accidental documentation mistakes to mixing up test results, and from lack of scrutiny in examining a radiographic assessment to a blood test that against all the odds repeatedly yields false negatives. Some reasons why diagnostics may fail (such as mixing up results) can occur across many fields of medicine, while others (such as the failure to spot something important in a radiographic assessment) are more specific to certain medical disciplines. But independent of the medical discipline we are looking at, we may initially distinguish two general types of wrong diagnosis. I will label the first type *misdiagnosis* and the second type *diagnostic malpractice*. If a wrong diagnosis is a result misdiagnosis, the diagnosis was provided in accordance with the standards of diagnostic procedures and reasoning but the resulting diagnostic conclusion eventually turns out to be false. A wrong diagnosis resulting from malpractice, on the other hand, is one that results from a procedure of diagnostic reasoning that was not pursued in accordance with the standards of diagnostic reasoning.[1]

1 There are some complexities related to the notions of misdiagnosis and diagnostic malpractice. Misdiagnosis seems to be conceptually more closely linked to wrong diagnosis than to

To keep these two sources of error conceptually distinct and to know how to identify them is important for normative reasons. If someone follows the correct diagnostic procedures providing the standards of good diagnostics, arriving at a wrong diagnose is upsetting, but intuitively it seems that such an outcome is not the personal fault of the diagnostic expert. Imagine that the gold standard for diagnosing depression were a saliva test with a 0.1% false positive and false negative rate. If the diagnostic expert uses the test correctly, and the result is positive although the patient (as it turns out later) is not depressed, it seems that this is not the fault of the expert (who did as well as he could), but a risk inherent to the testing procedure. In cases of this kind, the diagnosing clinician would not be at fault or responsible for the wrong diagnosis or its immediate consequences. If, on the other hand, the wrong diagnostic result is attributable to mistakes made by the diagnostic expert in the diagnostic process that is under their control, things look different. In such a case, the clinician would arguably be at fault and responsible because they could have prevented the wrong diagnosis by following the standards of their profession.

Beyond just knowing who to blame, being able to differentiate between malpractice and misdiagnosis is important for legal reasons because malpractice, in contrast to misdiagnosis, is a legally relevant error that might grant patients the right to receive financial compensation and might cost a malpractising clinician their licence. Identifying such cases is also important for generating statistics on where and how often malpractice occurs, as well as for assessing the need for educational or administrative programs to prevent malpractice.[2]

---

malpractice. If someone is misdiagnoses, the diagnosis will necessarily be false. If someone receives a diagnosis via malpractice, this diagnosis might nonetheless be right by accident. However, even if a malpractising clinician is lucky and provides the right diagnosis, this would be considered problematic because they are not practising according to medical standards, which – independent of the outcome of their practice – is an issue, since there is an agreement to practise according to such standards in order to ensure quality care. So even if malpractice leads to the right result, there is reason to criticise the malpractising clinician. In the following, I will focus on malpractice with a wrong diagnostic outcome since these are the instances in which identifying and differentiating between malpractice and misdiagnosis will be of most relevance, at least legally, due to the (potential) cause of harm.

2 This understanding of malpractice is generally in line with the way it is treated in common law jurisdictions. Although details of the law differ significantly between different countries, in general, liability for malpractice in medical professions is given if there is a failure to show a fair, reasonable, and competent degree of skill, measured by the standards of the profession, and/or there is a violation of ethical standards (Giesen, 1988). A difference between most understandings of malpractice in law and my understanding is that there is often an additional *harm condition*. Only if the behaviour of the clinician caused significant harm to the patient will it qualify as malpractice. Although this may be a reasonable approach for the purpose of lawsuits for practical reasons (e.g., saving court resources, determining compensation), I think it is unreasonable to accept this consequentialist condition when we are discussing the nature of malpractice. The fact that the clinician enjoyed the moral luck that their be-

A theory of psychiatric diagnostic reasoning should provide the resources to make sense of this distinction between malpractice and misdiagnosis and provide guidance on how to identify malpractice in the context of psychiatric diagnostics. In the following, I will discuss how the model-based account does this. Let us start with misdiagnosing.

In short, misdiagnosis happens if the clinician follows best practice of diagnostic reasoning and nonetheless ends up providing a wrong diagnosis. How may misdiagnosis occur, according to the model-based account? Let us look at the diagnostic process as understood in the model-based account to try to spot the places where error leading to wrong diagnosis may occur, even if good practice has been conscientiously pursued. As we may recall from previous chapters, to carry out a proper diagnostic procedure the psychiatrist will have listened to the spontaneous complaints of the patient and systematically evaluated their psychopathological status. After so doing, the psychiatrist will have considered the different potential models of psychopathological, other medical, or non-medical conditions the patient may present accounting for their complaints. Then, by interviewing, testing, and examining the patient, they will gather the information that is relevant to evaluating the models of these conditions against the patient's presentation. Once the information has been gathered, the best-fitting (and sufficiently well-fitting) models for the present complaints will be selected, one or more diagnoses will be attributed to the patient based on the classification rules of the manual being used, and a case formulation will be provided. Assuming that all these steps are carried out adequately by the psychiatrist, there are two remaining loopholes that may promote wrong diagnosis. Both relate to the problem of insufficient information as the basis of the diagnostic reasoning procedure.

The first reason for misdiagnosis is *diagnostic uncertainty* resulting from ambivalence between multiple diagnostic options, because the information is insufficient to make a clear decision, potentially leading to a wrong diagnostic conclusion. As the topic of diagnostic uncertainty *qua* ambivalence is important in itself, I will explore it in detail in 4.2. When exploring the topic of diagnostic uncertainty later, I will say more about its contribution to misdiagnosis. For now, let us focus on the second potential source of misdiagnosis, which is the *lack of relevant information*.

---

haviour had no negative consequences for the patient does not seem make their behaviour less problematic and unprofessional considering what should be expected of a clinician. To make an intuitive comparison: whether a driver engaged in speeding should be determined not by the consequences of them speeding, like hitting someone or not (although this might be relevant in court), but by what constitutes speeding and whether the driver did what we consider to be speeding. If you disagree, this is no problem; nothing really depends on this preference of mine. If you do disagree, you could just add in the harmfulness condition on top, and the rest of my explanation in terms of the model-based account would not change.

Lack of relevant diagnostic information might come about in many ways. Patients might intentionally misinform or hold back information from the diagnostic expert, or they might misremember or have forgotten things when asked about them. They might have performed intentionally badly in cognitive tests, or just have been unmotivated to cooperate and therefore not performed well. Or they might simply misunderstand the instructions or questions but appear so confident and competent that the clinician had no reason to think that there was a problem.

Imagine a patient showing the objective complaint of reluctant speech behaviour. As discussed in the last chapter, such speech behaviour may point towards the psychiatric symptom of *alogia* and so is of interest to the psychiatrist. As we also discussed in the last chapter, besides being alogia, reluctant speech might occur as a medical symptom in the context of a traumatic brain injury, or the patient's speech behaviour might result from the patient's intention to be uncooperative. Let's say that the patient intended to be uncooperative – specifically, to make the psychiatrist think they had a traumatic brain injury. If the psychiatrist interviewed the patient to gather information in order to evaluate the models for the respective diagnostic options, the patient could simply pretend to be unable to give longer answers if required and could say that he has not always been like this, which would be supported by relatives and friends of the patient because he indeed is not normally like this. This would then exclude the model for the diagnosis of motivated monosyllabism. Also, he would easily be able to pass the cognitive tests evaluating the presence of alogia discussed in the last chapter. Finally, the patient might then claim to have stumbled over a chair today, hit his head, briefly lost consciousness, and has the feeling that he lost some time afterwards. He may claim that he felt disoriented for a minute after this and was feeling sick. Maybe this patient planning the fraud even hit himself with a stick, hard enough to have a bump on his head to support the illusion. Although a CT scan provided for the patient would not show any lesions, the rest of the story and the overall evidence would perfectly fit the case of a traumatic brain injury, and not every traumatic brain injury necessarily shows up as a lesion in a CT scan of the brain. In conclusion, the psychiatrist would likely and wrongly conclude that the complaint of the patient's reluctant speech results from a traumatic brain injury. This wrong conclusion, however, would be a misdiagnosis rather than malpractice, because at this point the psychiatrist invested reasonable effort and carried out the required diagnostic procedures to gather the diagnostically relevant information, but arrived at a wrong conclusion based on an informational bias. This bias did not result from the psychiatrist doing anything that would go against good diagnostic practice guidelines, and so we would usually not consider him to be at fault for having arrived at this wrong conclusion. So much for misdiagnosis for now; we will return to it in 4.3. Now let us turn to what would constitute a case of wrong diagnosis *qua* malpractice.

As in the case of misdiagnosis, let me point out what may go wrong in the diagnostic process as presented by the model-based account in the case of malpractice. While misdiagnosis occurs when all steps are carried out correctly but there is a residual uncertainty or misleading diagnostic information that leads to wrong diagnostic conclusions, malpractice occurs if the psychiatrist makes significant mistakes in the procedure of diagnostic reasoning. Again, this procedure consists in listening to the spontaneous complaints of the patient and systematically evaluating their psychopathological status; considering the various potential models of psychopathological, other medical, or non-medical conditions; testing and examining the patient for information relevant to evaluating these models against the patient's presentation; selecting the best-fitting models for the present complaints; providing a formulation based on the selected models; and providing one or more diagnosis based on the classification rules of the manual in use and the symptoms identified in the case formulation. In any of these steps, the psychiatrist could make mistakes leading to a wrong diagnosis, constituting a case of malpractice. Here are some examples. Psychiatrists might not spend enough time listening to their patients' complaints, or might incompletely assess their mental status, which then leads them to fail to consider all relevant models and therefore to end up not evaluating all relevant complaints. They might make mistakes in selecting a best-fitting model for patients' complaints, because they do not invest enough effort in thinking about which model is best supported by the information gathered about the patient. Or they might not pay close enough attention to the diagnostic criteria of disorder diagnosis and provide an unjustified diagnosis. In all these cases, the psychiatrist would be at fault for the wrong diagnosis and the harm that might take place in consequence of a wrong diagnosis produced by malpractice, because they did not fulfil their diagnostic responsibility at the level of the diagnostic procedure.

Taking this approach to misdiagnosis and malpractice, what does it do to help us identify and distinguish between them? Imagine an instance in which a patient has received a diagnosis that has later been judged to be wrong, and that this patient has received treatment based on this diagnosis that was harmful – for instance, because of side-effects of medication that she would not have been prescribed if her initial diagnosis had been correct. Now the patient is pressing malpractice charges against the practitioner. For someone to decide whether the wrong diagnosis of the patient resulted from malpractice, rendering the clinician at fault, or was a misdiagnosis that is not the fault of the clinician, someone investigating the case would have to answer a question deriving from the most general understanding of malpractice and misdiagnosis, as presented in the first paragraph of this section: did the wrong diagnosis result from the practitioner not carrying out the diagnostic procedure with thoroughness, or because the diagnosis was based on wrong or incomplete information, or on information that led to diagnostic ambivalence in which the wrong choice appeared plausible? The interpretation of the difference between malprac-

tice and misdiagnosis in light of the model-based account to diagnostic reasoning provides an approach to answering this question in principle.

If there is sufficient information available about the diagnostic process that was carried out and the diagnostic considerations made by the diagnostic expert (e.g., in the form of documentation, notes, the case formulation, and (honest) reports), someone investigating the charge of malpractice may look at this information to evaluate whether it indicates that the clinician followed the model-based diagnostic reasoning step by step in the way outlined earlier and presented in detail in the preceding chapters. If not, this would suggest that the clinician engaged in malpractice. If no malpractice took place, the only other option is that the wrong diagnosis is classified as a misdiagnosis. If, however, the investigation comes to the conclusion that somewhere in the diagnostic process malpractice took place and led to the wrong diagnostic outcome, the clinician will be responsible for the wrong diagnosis and the consequences of actions that were taken or not taken based on it.[3] In this way, the model-based account helps us to differentiate and identify instances of misdiagnosis and diagnostic malpractice.

3 It could be the case that although some aspect of the diagnostic process qualifies as malpractice, correctly carrying out the diagnostic procedure would have made no difference. In other words, the same wrong conclusion would have been drawn even if no malpractice had taken place. This might happen, for example, because in another part of the diagnostic process an important piece of information was not accessible to the clinician even though everything was done right in this part of the diagnostic procedure, while the part of the diagnostic process that was carried out wrongly would not have provided information or conclusions that would have made a difference. For example, it might be that the clinician did not carry out a proper mental status examination but did not miss anything relevant to the wrongly made or potential correct diagnosis because of this. It was a patient's lie later in the interview that led to the wrong evaluation of a complaint as some particular symptom and in the end to a wrong overall diagnosis – as, for example, in the case of the patient faking the TBI. In this case, malpractice took place but this malpractice would not be the cause of the harm to the patient. This again may have different legal consequences and depending on our moral stance might also make moral differences. Malpractice took place nonetheless. And again, the model-based understanding provides the resources for deciding whether the malpractice is responsible for a potential harmful outcome. It can help us evaluate where in the process specific diagnostic decisions have been made in the context of the evaluation of diagnostic models against diagnostic information, and so can tell us which step in the process was relevant to which conclusion. If, given the analysis of the diagnostic process that took place, no lack of information, misused models, or inferential mistakes resulting from the malpractice in this case seems to be responsible for the wrong diagnostic choice, the wrong diagnosis would be a misdiagnosis even though there was also malpractice involved in the overall diagnostic procedure.

4.4 Diagnostic Uncertainty through Ambivalence

Another phenomenon well known in clinical contexts is diagnostic uncertainty and the attempts to overcome it. While it is sometimes easy to determine what the diagnosis of a patient should be, this is not always the case. There are occasions on which psychiatrists are uncertain about diagnostic decisions because what they have learned about the patient seems to allow for several potential diagnostic conclusions, so that additional effort is necessary to carve out which among the plausible diagnostic options might be the best. And even then, finding a certain answer might not always be possible. How uncertainties in diagnostics arise, and how they might successfully or unsuccessfully be resolved, will be the focus of discussion in this section. In addition, I will say a few words about how, despite great effort, a failure to resolve uncertainty might lead a psychiatrist to draw a wrong diagnostic conclusion, and why such cases are misdiagnosis rather than malpractice. This discussion will supplement the previous work in 4.3.

For psychiatric diagnostics we must consider two levels of uncertainty: the level of syndromal diagnosis and the level of symptoms. On the syndromal level, clinicians may be uncertain whether they should attribute a certain mental disorder diagnosis (X) to a patient or not, whether they should attribute one or another diagnosis (X or Y or …) to a patient, or whether they should attribute more than one diagnosis (X and Y and …) to a patient. Although this level of uncertainty often occurs, it is philosophically relatively uninteresting from the perspective of the model-based account, because how this decision must be made in accordance with best practice is solved by the major diagnostic manual in use, and if it were not solved by the manual, there would be no right or wrong way to do it.

In general, a diagnostic evaluation produces evidence of a sufficient standard to allow us to infer the presence of symptoms and so to provide a diagnosis whose list of diagnostic requirements most closely matches the patient's presentation, maximising the number of psychopathological relevant features addressed by one diagnosis. Whether a subset of the diagnostic features already employed to provide this diagnosis is allowed to be used again to justify another diagnosis is case-dependent. The DSM-5 (APA, 2013, pp. 155f.), for example, does not intend clinicians to reuse symptoms used to diagnose a major depression to additionally diagnose a patient with moderate and mild depression. However, it does allow clinicians to reuse them to additionally diagnose patients with dysthymia (ibid., p. 168), which would be what is usually called a double depression. The DSM-5 does support diagnosing agoraphobia (ibid., p. 218) on top of a panic disorder (ibid., pp. 208f), but not panic disorder if panic attacks occur in response to social situations (i.e., social anxiety) (ibid., p. 209). Manuals also offer many diagnostic options to account for leftover symptoms that are insufficient to support an independent diagnosis. The DSM-5 (ibid., pp. 160f.), for example, allows us to specify that a major depression diagnosis

is accompanied by anxiety features that in themselves do not suffice for an anxiety disorder diagnosis, by adding the specifier "with anxious distress" to the diagnosis (ibid., p. 161). And finally, for certain disorders that are clearly approximated in terms of present symptoms but not fully met by the diagnostic findings, there are diagnostic categories that allow clinicians to classify these as well. For example, according to DSM-5, cases in which several depressive symptoms are present but no constellation is observed that would allow for any formal diagnosis of depression, the clinician is supposed to diagnose "other specified depressive disorder", which is a "presentation whose symptoms [are] characteristic of a depressive disorder that causeses clinical significant stress […] but do not meet the full criteria for any of the disorders" (ibid, p. 165). Whatever critique we might wish to make of the major diagnostic manuals DSM or ICD from the perspective of the model-based account – which, remember, is not an attempt to criticise diagnostic practices but rather an effort to make them intelligible – it does not seem that if well applied, these manuals leave the diagnostic expert who is aware of the symptoms of their patients in the dark about what diagnostic decisions they have to make. However, the "who is aware of the symptoms" qualifier brings us to the philosophically more interesting instances of diagnostic uncertainty from the model-based perspective: uncertainty regarding what symptom to attribute.

Diagnostic uncertainty regarding symptoms can occur in various patterns if it is not unequivocal which symptom value an initial complaint should be assigned after the patient has gone through the diagnostic process. The psychiatrist might be uncertain as to whether a complaint should be evaluated as one psychiatric symptom or another, or as a medical problem or a non-medically relevant issue instead. Such uncertainty often occurs in clinical contexts and may force the clinician to think harder or do additional diagnostic work to reach a solution, which sometimes but not always works. Uncertainty may persist as to whether a patient's complaint clearly qualifies as a psychiatric symptom or is a psychological complaint of non-clinical value. How exactly we can understand the occurrence of such uncertainty and the ways in which it may be resolved? Here is how the model-based approach can account for it.

If we consider the above-described diagnostic uncertainty regarding symptoms via the modelling account, it appears there are three possibilities for how it may arise:

i) None of the models set up for an initially recognised complaint matches the patient's well enough to be accepted. As a result, the psychiatrist has no unambiguous basis on which to make any judgement for or against evaluating the complaint to be a psychiatric symptom, a medical complaint, or a non-medical issue.

ii) More than one model for a complaint from amongst those set up based on knowledge from the domain of psychiatry (e.g., models that would render the complaint psychiatric symptom (a) or (b)) fits the patient's condition sufficiently well to be accepted. As a result, the psychiatrist has no unambiguous basis on which to make a diagnostic judgement regarding the initial complaint.
iii) More than one model for a complaint from amongst those set up based on knowledge from a range of domains (i.e., psychiatry versus other medical or non-medical fields) fits the patient condition sufficiently well to be accepted. As a result, the psychiatrist has no unambiguous basis on which to make a judgement for or against evaluating the complaint to be a psychiatric symptom.[4]

In all these cases, the decisions regarding the psychiatric symptom value of a complaint cannot simply be looked up. If we have only the complaint as the prior, there is no straightforward formal way to derive the correct evaluation in the way we can do it if we are on the level of disorder diagnostics, already equipped with a set of symptoms that we can take as priors to decide which disorder(s) to diagnose. How, then, do we overcome such a situation? The psychiatrist has several options. Some of these options are attempts to deal with the uncertainty by forms of further theorising and evaluation, while others present pragmatic solutions. I will discuss in turn the three instances of uncertainty and how they can be addressed by such means.

The first type of uncertainty, resulting from no diagnostic model suiting the patient's presentation sufficiently well according to the fidelity criteria assumed for the tested models, is the most severe case of diagnostic uncertainty. Think of an example of a patient reporting anxiety. On close evaluation, it turns out that this patient does not show any signs of the typical cognitive style and somatic reactions of anxiety that would allow the psychiatrist to identify their anxiety as a psychiatric problem. The patient has also had no recent experiences that would render his currently high anxiety level understandable. He has taken no medication and has no physical condition that might induce such reactions. The severity of such cases lies in the problem that there are no theoretical resources that seem to provide a theoretically justified diagnosis, because the complaint matches no diagnostic models whose application would justify the inference to any diagnostic conclusion regarding a complaint. The psychiatrist just has no way to say what is going on here, and ideally this would also become clear in the psychiatric case formulation.

4 What about the option of multiple medical but non-psychiatric models, or wholly non-medical models, fitting equally well? While this option exists, I will not discuss it here, as in these cases it is to be assumed that the complaint is not a psychiatric symptom and further diagnostic efforts would either be a matter for another medical profession (where multiple non-psychiatric medical options fit) or be of no medical interest at all (where multiple non-medical models fit).

Pragmatically speaking, a psychiatrist may nonetheless support the evaluation of a complaint as a symptom or a medical problem initially suggested by the complaint. In such a case, the clinician would end up making what has been called a *suspicion diagnosis*. A suspicion diagnosis may be understood as the diagnostic proposal that is the most plausible option given all diagnostic evidence but that is still not sufficiently certain to fully endorse it. It is supported by pragmatic considerations regarding the cost/benefit calculus of treating a patient according to this diagnosis versus another diagnosis versus refraining from providing any diagnosis and not treating the patient at all.

To give an example, it might be the case that a patient meets all but one criterion sufficient for a major depressive disorder (MDD) and displays a complaint that, if it were a symptom. would allow for this diagnosis. However, no model evaluated suggests that the complaint be considered a symptom. Further, imagine that there is a certain intervention that, based on treatment guidelines, is intended to be provided only to MDD patients, but there is a good chance that this intervention might help the considered-close-to-MDD patient, because there is some evidence that it may help reduce symptoms in other depressed but not MDD patients. In such situations, psychiatrists take the path of what has been discussed in the literature as "workarounds" (Whooley, 2010): they diagnose as if the complaint were a symptom. While everyone working in clinical practice will be familiar with such patterns of practical reasoning, the question of course arises as to whether these patterns of reasoning are rational and ethically permissible considering the overall practical purpose of psychiatry to help patients, or whether other considerations (e.g., the risk of biasing epidemiological studies based on clinical data, not meeting general standards of evidence-based practice) speaks against such practice. I will remain agnostic regarding this normative question.[5] To come back to our anxiety example, the psychiatrist may for pragmatic reasons decide to consider the initial complaint of anxiety as a psychiatric symptom for the practical purpose that this might allow for a diagnosis that could be used to justify therapeutic or pharmacological treatment, so that there is at least a chance of improving the patient's condition.

The second and third type of uncertainty occur if there are several models of a psychiatric complaint that match the patient's presentation sufficiently well, while at least one of these model, if chosen, would render the complaint a psychiatric

5 The pragmatic reasoning process feeding into suspicion-diagnostic conclusions is a kind of clinical reasoning rather than diagnostic reasoning. The interaction of this clinical reasoning with theoretical diagnostic reasoning evaluating the initial plausibility of diagnostic conclusions purely on the basis of diagnostics is an interesting and clinically relevant topic. However, delving into the logic of pragmatic reasoning in clinical diagnostics would require a new line of investigation and is thus beyond the scope of my project, which focuses on epistemic (i.e., diagnostic not clinical) reasoning. I will therefore not discuss the topic of how exactly suspicion diagnosis is provided and justified, but only outline its structure and purpose.

symptom. Going back to the anxiety example, a patient reporting the complaint of anxiety might present in the in-depth evaluation such that a model evaluating the anxiety as a psychiatric symptom – by assuming a model of anxiety's typical cognitive style (including attentional bias, memory bias, and interpretation bias) – applies sufficiently well. At the same time, a model that assumes the anxiety to be a normal psychological reaction in light of a model assuming a combination of environmental factors to increase stress in the patient, making their anxiety response normal, also fits the patient sufficiently well. That is, it appears justified to assume the patient's complaint to be a psychiatric symptom as well as a normal psychological reaction.

To resolve uncertainty in this instance, two approaches seem to be available. For a theoretical solution, the matching models may be compared in terms of how good their match is with the targeted complaint of the patient. If it turns out that one model matches the patient's presentation better than the other model, even though both models seem to be in principle applicable, it appears rational to choose the best-fitting model to make a diagnostic decision as to how to classify the patient's complaint. If, for example, two propositional models target the same complaint and from each model enough central propositions apply to the patient's presentation that in principle both models seem to match the patient's presentations, the diagnostic expert will go for the model that contains more diagnostic propositions that match with the patient's presentation – that is, the model that is a better fit. Of course, the judgement of "better fit" again has its complexities. Typical goodness-of-fit models that can be used in mathematical modelling to quantify how well a model matches with observations of the modelled system, producing a numerical value that allows for a decision between models, do not seem straightforwardly applicable given that we are dealing with qualitative models. Rather, it appears useful to ask what fraction of the total number of the propositions that the models consist in, beyond those sufficient to make a well-fitting candidate, are met.[6]

If this procedure does not lead to a conclusion favouring one model over another, because again both models seem to apply equally well, uncertainty is residual. Then the clinician must either refrain from drawing a diagnostic conclusion regarding the

6 Here, another weak point of psychiatric modelling (beyond its potential vagueness due to its qualitative format) surfaces. Since the models used to identify psychiatric symptoms are constitutive models, they do not necessarily entail any claims about specific causal relationships or aetiologies of the phenomena they attempt to model. They only identify constituents that must be present to attribute a symptom. The problem with this account is that if the constituents of more than one type of model apply equally well (or at least indistinguishably similarly well), to decide between them becomes impossible. What could solve this problem would be evaluating which potentially constitutive features are also causally responsible for the patient's presentation. An option that is not at its disposal of psychiatric modeling as it stands. Coming up with reliable causal models that would allow us to evaluate psychiatric symptoms would be beneficial in this regard.

complaint or opt for the pragmatic solution strategy, assuming an evaluation without fully endorsing it in order to support a suspicion diagnosis as described above. However, it seems that in this context a suspicion diagnosis, although still not unequivocally supported by evidence, would be epistemically stronger, because there is at least some evidence base that in principle would be sufficient to support the diagnosis, rather than no evidence speaking for it. The pragmatic decision could therefore be made with a higher base level of confidence and perhaps with fewer alternatives that are equably plausible compared with cases where no model seems to match the complaint, and where all models are similarly (un)likely. As a result, however, the diagnosis of the symptom may be wrong, and its suspicion-diagnostic support may allow for a syndromal diagnosis that is wrong. Yet after all the diagnostic steps have been carried out correctly, arriving at such a diagnosis for pragmatic reasons, such as allowing for a most plausible and least harmful treatment that might potentially improve the patient's condition, is in line with the pragmatic aims of psychiatry to cure and care for patients. And if the conclusion turned out to be wrong, this would make it a misdiagnosis rather than a case of malpractice. In this way, I have also outlined the missing way to arrive at misdiagnosis, as promised in the previous section. Next, let us turn to the topic of instinctual diagnosis.

4.5 "Instinctual" Diagnosis

If one works in clinical context, say a psychiatric hospital, a story like the following will perhaps be familiar. A senior physician is coming to see a new patient who just got admitted to the psychiatric unit. She enters the room and exchanges only a few words with the patient. She then leaves the room and says to her colleagues something like "I suspect the patient has an XYZ diagnosis". And it turns out after more detailed diagnostic procedures that the senior physician was right. It seems that she has a special diagnostic "instinct". How can we explain how such often reliable instinct works, what its epistemic benefits and downsides are, and why we apparently want the actual diagnosis to be made according to formal standards even if we have a clinician with great intuition around? The model-based approach provides us with a story that allows for a plausible approach to all these questions.

Let us go back to the situation of the short encounter between a clinician and a client from which such an instinctual diagnosis might result. What is going on here? Plausibly, in a short encounter with a patient, the psychiatrist will at best be able to become aware from observation or incomplete evaluation of a limited number of complaints of the patient. Although no full picture of the patient's complaints can be claimed, since no complete screening has been conducted, the physician will at least have gathered some information about the most salient complaints of the patient, though not the necessary information to evaluate them properly for their symptom

value. In other words, the clinician has conducted an incomplete first step in the proper diagnostic process. What is she doing with the information to arrive at a diagnostic conclusion? The spotted complaints are treated *as if* they would have turned out to be psychiatric symptoms. The psychiatrist has a list of potentially present list of symptoms to hand and can think through the limited number of disorders that would match with this pattern, proposing that the patient will perhaps suffer from the disorder(s) matching the assumed symptoms that are most likely present, possibly for a subset of the clinical population that the patient falls into on first glance (e.g., as regards sex or age).

Although such quick likelihood assessment may generate a first hypothesis as to what might be the patient's disorder that may turn out to be correct, this approach to diagnosis often has the problem that it is not comprehensive or supported by evidence. In diagnosing a patient, we expect diagnosis to be supported by the best available evidence that can be collected with reasonable effort to determine what the patient's problem may be, so that they can be offered the most beneficial treatment for their condition and we can avoid harming them by offering wrong treatment or withholding better treatment options from them. In this case, there is a fair chance that we will do exactly this, since we cannot know whether any of the complaints would indeed be evaluated as psychiatric symptoms if properly assessed. A complaint may not be the symptom of relevance and may therefore mislead the diagnostic guess. There is also a risk that it is such a symptom but that this symptom is not part of the most likely psychiatric syndrome, or that the pronounced symptom is present but not enough other diagnostic criteria are met in addition to it to diagnose the suspected condition. Also, complaints that were not picked up on by a short encounter will not be considered in the diagnostic guess, and these might have pointed towards highly relevant symptoms that would have led to a different diagnostic conclusion. Hence, basing one's diagnosis on a short encounter and a diagnostic guess seems to harbour a significant epistemic risk of being wrong. As being wrong in this case would mean being wrong because of a lack of proper diagnostic procedures, taking this risk and ending up with a wrong conclusion would indeed mean having engaged in malpractice, which is why usually "instinctual diagnosis", although it provides some guidance for a clinician to think about what might be wrong with their client, is not accepted as a proper approach to diagnosing patients.

In the above case, we assumed that the diagnostic guess was the most rational possible based on the best knowledge of the likelihood of symptoms and disorders in certain reference populations of patients, under the assumption that every spotted complaint would be a psychiatric symptom. Another problem arises if we bear in mind that humans, especially when they think rapidly, are anything but perfect rational machines. In rapid diagnostic decisions, humans tend to unintentionally apply heuristics that bias their decisions (Tversky and Kahneman, 1974). Heuristics that are important in diagnostic contexts appear to be, for example, the availabil-

ity heuristic (which leads us to judge how frequent or probable something is based on how easily we can bring to mind an example of a state of affairs, leading us to mistake actual availability for actual frequency) and the representativeness heuristic (in which we assume that someone belongs to a category because they seem to match the stereotype of this category) (Tversky and Kahneman, 1981). Both are found to be widely present in expert judgements, including in the diagnostic judgements of medical and psychiatric experts (e.g. Elstein, 1999; Garb, 1996; Koehler, Brenner, and Griffin, 2002; Ægisdóttir et al., 2006). Therefore, on top of the likelihood of being wrong in an "instinctual diagnosis" even if we were perfectly rational and well informed, our own human psychology is an additional problem. Our psychology might bias us to judge patients as falling into one or another diagnostic category just because we as clinicians happened to see patients showing a certain complaint as matching a stereotype of someone having a certain disorder, or because in the limited sample size of patients we have seen, patients with a certain complaint mostly turned out to have this disorder.

As an example, think of a patient who is harming himself without the intent to kill himself. Such behaviour may indicate the psychiatric symptom of nonsuicidal self-injury (NSSI) (Klonsky, Victor, and Saffer, 2014). NSSI is present, for example, in autism spectrum disorder (Johnson and Meyers, 2007), borderline personality disorder (Oumaya et al., 2008), bipolar disorder and dissociative disorders (Joyce et al., 2010), eating disorders (Rodríguez-López et al., 2021), depression, phobias, and schizophrenia (Singhal et al., 2014), non-suicidal self-injury disorder (Zetterqvist, 2015), and Munchhausen syndrome (Humphries, 1988). Looking at the available data, we learn that patients admitted to psychiatric hospitals with self-harm seem to suffer most frequently from depression or anxiety or and alcohol misuse, as well as attention deficit hyperactivity disorder (ADHD) and conduct disorder in younger individuals (Hawton et al., 2013). As pointed out by Hawton et al. "[t]hese findings are clearly at odds with the commonly held but misinformed view that the majority of self-harm patients do not have psychiatric disorders, or if they do then this is most likely to be a personality disorder." (Hawton et. al. 2013, p. 828).

However, there are also reasons for self-harming reported in the literature that do not seem to point towards psychopathology, such as religious reasons or the requirement to do so to be part of a certain subculture (Edmondson, Brennan and House, 2016). If a psychiatrist, knowing all this, briefly encounters a patient showing signs of self-harm or reporting having harmed himself, the first idea that springs to mind might be that this patient suffers from those disorders most frequently associated with this behaviour if it is a psychiatric symptom, and often enough the psychiatrist will be correct in their guess. However, in many cases this guess might also go wrong. Considering the example of self-harm, the patient may suffer from a different mental disorder associated with the suspected symptom(s) assumed based on the complaints (e.g., non-suicidal self-injury disorder rather than borderline per-

sonality disorder). Or the behaviour may not be a symptom of a mental disorder but rather a religious practice. Moreover, if we do not assume a perfectly informed and rational clinician but one whose decisions are potentially biased by availability and representativeness heuristics, the clinician might, after assuming the patient's behaviour to be a self-harm symptom, even more rapidly come to the conclusion that the patient suffers from borderline personality disorder. This might happen if the clinician worked for years in a hospital unit specialised in treating borderline personality patients who often showed this behaviour, so that there is now a tendency to equate self-harm as a symptom with the presence of a borderline personality disorder.

Looking at this example, it becomes clear why no responsible trained clinician should base their final diagnostic conclusions on their instinctual or educated diagnostic guesses. Thorough evaluation of diagnostic models against patients' presentations based on proper diagnostic information provides a better justification base for diagnostic conclusions than the above-described likelihood judgements. It does so because evaluating what indeed is the situation with a patient and matching this with our best psychopathological understanding of what is constitutive for a present psychopathological symptom tells us what is the case with the patient, rather than only telling us what the case with the patient might potentially be with a certain probability if a certain model fitted the patient. By following the proper process, the diagnosis also achieves diagnostic superiority, because if it is based on the process of model evaluation, it is supported by evidence that allows the inference of the presence of a certain symptom to be an inference to the best explanation. This inference occurs via the acceptance of a constitutive model that provides a constitutive explanation of how to understand the patient's complaint.

To avoid obviously problematic approaches by which diagnostic conclusions like the one discussed in this section may be reached, and also to make sure that there are no smaller mistakes in the process of diagnostics, there is an important tool at our disposal: critical diagnostic reasoning – that is, the critical diagnostic examination of one's own and others diagnostic work. This form of critical engagement with diagnostics is the topic of the next section.

4.6 Diagnostic Disagreement

Clinicians can be wrong about their diagnostic proposals for various reasons, some of which we explored above when we talked about misdiagnosis, malpractice, and diagnostic instinct. Knowing all too well that diagnostics is fallible, it is generally considered important to ensure that as many mistakes as possible are prevented or at least corrected.

Good clinicians try to do this with their own diagnostic conclusions once they have arrived at them by putting their own proposal and the way they arrived at it to the test again. If, after their self-assesment process, they still support their diagnosis, they will also evaluate it again later if interventions lead to changes that may require a diagnostic re-evaluation, or if any additional diagnostically relevant information is obtained that might require correcting their initial diagnostic judgements. But self-monitoring is not the only thing that happens. Besides monitoring their own work, clinicians also monitor each other if they disagree with a diagnostic conclusion and discuss this disagreement with each other, or at least they may ask colleagues to explain the reasoning behind a certain diagnostic conclusion – something that takes place in particular between new clinicians and their supervisors, to assess and train their diagnostic reasoning. Engaging in this kind of self-criticism and intrapersonal criticism of diagnostic decisions and resolving differences between two mutually exclusive evaluations is called critical diagnostic reasoning. This is thought to be an important feature of diagnostic reasoning as practised by clinicians, no matter their specialisation (Harjai and Tiwari, 2009; Mamede, Schmidt, and Rikers, 2007).

To engage in critical diagnostic reasoning, clinicians ask themselves or others questions that make them check their diagnostic decisions. For example, "Why exactly did I/you draw this diagnostic conclusion?", "What could be an alternative explanation?", "Did I/you consider all available and potentially relevant information?". Answering these questions by presenting a valid inferential path leading to the diagnosis, in support of which relevant information was gathered and adequately considered, can support one's confidence in one's diagnostic judgement, or, if the answers hint at flaws, undermine it. Alternatively, if there is a disagreement between clinicians, answering this question on both sides of the conflict and demonstrating how the diagnostic reasoning process on each side meets or fails to respond to these questions may lead to a rational agreement as to whether one or the other or maybe neither option seems to be right, or whether there is a residual uncertainty about whose the right diagnosis is. Now, how does the model-based account make sense of these intra- and interpersonal procedures?

Intra or interpersonal critical diagnostic reasoning is structurally equivalent to the procedures that can be employed in the case of diagnostic uncertainty discussed earlier. Therefore, the relevant points are quickly made. At the top level of syndromal diagnostics, the model-based account has nothing particularly interesting to say beyond what is to be found in the diagnostic manuals considering disorders to be sets of symptoms and using additional criteria to tell us straightforwardly whether a diagnosis is correct or not. Critical reasoning on this level simply requires double-checking whether all diagnostic criteria have indeed been met. This may be done for oneself (intrapersonal) or between clinicians (interpersonal). And again, it is the symptomatic level that seems to be more interesting. In other words, while there is

little to no room for disagreement about what must be present for a major depression, because we can look it up in the manual we are using, whether the required symptoms are present (i.e., whether a patient's report that he no longer has fun when pursuing his hobbies is indeed a case of anhedonia) offers a livelier ground for diagnostic disagreement.

Critically evaluating whether attributing or not attributing any specific psychiatric symptom is adequate provides more room for the application of the model-based diagnostic reasoning framework. Its application is in principle like the method discussed earlier in cases of diagnostic uncertainty, since doubting the adequacy of one's diagnostic decision basically amounts to intentionally introducing artificial uncertainty. If a clinician is coming back to a diagnostic evaluation of a complaint, they may ask themselves whether they did carry out the initial evaluation (screening) of the patient in a way that covered all relevant areas, whether they considered the models for all encountered complaints, whether they considered all models relevant to the encountered complaints, whether they did what was required to generate data that allowed for the evaluation of the relevant models in the in-depth evaluation, and whether as a result of the comparison they chose the right model to apply.

The same may take place on an interpersonal level. Here, the debate between clinicians may start from various points. A supervisor or chief may want to discuss a diagnostic conclusion of a trainee to test and exercise their diagnostic reasoning skills based on a patient case that the supervisor themselves has never seen. Or a debate might result from a chief physician reading the case formulation supporting the chosen syndromal diagnosis of a patient but being unsatisfied with the justification provided by it. Or maybe colleagues in a team end up disagreeing about a diagnosis of a patient they are treating together and have to sort out this disagreement. In any of these cases, the clinician whose diagnostic conclusion on the level of symptom attribution is in question will have to make transparent the actions undertaken to gather initial and additional information about the patient, the models considered to apply, and why each model based on detailed diagnostic information was accepted or rejected. Making transparent this process then opens the field for interpersonal criticism. The colleagues or supervisors may point out that some models were not considered or sufficiently evaluated, suggest that the diagnostic data were insufficient to assume that one of the tested propositional models indeed applies to the patient, or raise many other points regarding any stage of the diagnostic process. If the interpersonal disagreement comes to a point where both debaters agree that each other's diagnostic evaluation is in principle valid, they might nonetheless think that their diagnostic choice is to be preferred because the model they picked better suits the patient's case. This situation may then be debated further, considering the theoretical solution strategy for diagnostic ambivalence earlier in this chapter, with the same potential outcomes: a solution in favour of one diagnostic conclusion or a

residual uncertainty. To sum up: the way in which we can understand the occurrence of diagnostic disagreement and critical diagnostic reasoning in the context of psychiatric diagnostic reasoning is well covered with the resources of the model-based account.[7]

4.7 Change in Diagnostics

The final topic I wish to cover in this chapter concerns how an answer to the Methodological Question is capable of making sense of the possibility and limits of integrating changes into our understanding of psychopathology and the means we use to assess it. That an answer to the Methodological Question should have something to say about this is desirable for at least two reasons. First, because a good answer should be able to show that it will be able to assimilate modest changes in our understanding of psychopathology and methods of assessment. Small to modest changes occur all the time, and for an answer to provide a somewhat stable proposal that applies to psychiatric diagnostic reasoning at least in the recent past and will probably apply in the near future, it should be flexible enough to incorporate such changes. Second, it is important because only if the proposal can display its limits on implementing changes will it appear to be usefully precise. If significant changes that we could imagine taking place in a potential or fictional future of psychiatric research could be accommodated by the proposal without problem, it would seem too arbitrary to be considered a specific understanding of the diagnostic practices at hand.

In the following, I want to show how the model-based approach holds up to both requirements. To show the robustness of my account against small to modest changes but its sensitivity to relevant changes in psychopathology, I will discuss aspects of the two levels of diagnostics. The higher level of diagnostic decision-making will be discussed in terms of providing a symptom-based syndromal diagnosis,

7 What has been discussed in the previous sections on instinctual diagnosis and diagnostic disagreement, especially intrapersonal diagnostic disagreement, can also be found under discussion – sometimes in normative terms, sometimes in descriptive terms – in the medical education science literature on diagnostic reasoning. The error-proneness of quick and intuitive judgements and the relevance of analytic reasoning as their corrective have been discussed in the context of dual-process theories. These theories consider human cognition to consist of two interrelated systems, one of them intuitive, the other one analytic, with the intuitive being more prone to several kinds of bias (Monteiro and Norman, 2013). Applications of this idea in medical education assume that the same is true for diagnostic reasoning: quick intuitive judgements pay the price of being open to all sorts of biases, such that any judgement made in this way (if one is using this approach at all) requires the monitoring influence of analytic reasoning (Croskerry, 2009; Elstein, 2009; Marcum, 2012).

and the lower level of psychiatric diagnostics will be dealt with in relation to psychiatrists' evaluations of the presence of symptoms. Considering the case where the lower level remains the same and only the top level is changed, I will discuss what the changes may look like such that the model-based account may still be useful to understanding psychiatric diagnostic reasoning, and also under which circumstances it may no longer be useful. Then, for the lower level of symptoms, I will look at potential changes in the understanding of symptoms by homing in on the symptom of anhedonia. I will first discuss varying historical understandings and the current understanding of this symptom. I will argue that the variations in these understandings, though real, is small enough in its relevance to how the symptom would be evaluated that adopting each version of it would square with the model-based account's understanding of symptom evaluation. This argument will demonstrate the flexibility of this level of the model-based proposal for clinical diagnostics.

Next, I will discuss the current science of anhedonia falling within the field of computational psychiatry and how it is changing our understanding of anhedonia. Although the changes in our understanding of mental symptoms like anhedonia that computational psychiatry is currently encouraging have not yet led to widely adopted change in the clinical evaluation, this may happen in the future. I will therefore discuss, mainly using the example of anhedonia, some of the options for how computational psychiatry may soon change diagnostic evaluations and point out which changes would not, but also those that would, undermine the model-based approach. This will demonstrate the fallibility of my approach in light of more significant changes in diagnostics on this level. Finally, I will provide a brief discussion of some possible though perhaps unlikely changes to psychiatric diagnostics that would significantly transform our understanding of both levels of diagnostics. I will argue that these significant changes would render the model-based account a chapter in the history of psychiatric diagnostic reasoning rather than part of its present. I will conclude that the model-based account is flexible and thus robust enough, but at the same time sensitive and thus fallible enough, to fulfil the desideratum in question. Let me begin by discussing the current format of syndromal diagnostics and how its changes might or might not affect the plausibility of my answer to the Methodological Question.

If we look at the contemporary format of psychiatric diagnostics, which is based on syndromal diagnosis consisting of clusters of symptoms and signs, changes may appear on two levels: either on the higher syndromal level or on the lower symptom level. On both levels there may be changes. Let us talk about the higher level first. Changes on this level may entail, and have entailed, new diagnostic categories such as the gaming disorder introduced in ICD-11 (Aarseth et al., 2017). The criteria for existing diagnoses may be changed, as occurred with the criteria for PTSD from ICD-10 to ICD-11 (Barbano et al., 2019). Or diagnostic categories might be abandoned, like the subtypes of schizophrenia in DSM-5 (Tandon et al., 2013), or intro-

duced, like the subtypes of neurocognitive disorders in DSM-5 (Regier, Kuhl, and Kupfer, 2013).[8]

Although the central diagnostic manuals DSM and ICD may change in this manner, these and future changes of diagnostic taxonomy will not impact the ways in which these manuals are used as long as they keep operating in this framework of symptom-based syndrome diagnostics – that is, using identified symptoms and signs plus the additional diagnostic criteria to diagnose disorders. Accordingly, the symptom-based pattern recognition approach would perhaps not be influenced by these changes if the straightforward formal process of inferring syndromal diagnosis from patterns of symptoms remained the same. However, if the way in which we diagnose psychiatric disorders on the top level changed (i.e., if we still identified symptoms and signs but used them differently in a second step to make a higher-order diagnostic judgement), symptom-based pattern recognition approach might of course change too. To look at just one scenario that somewhat realistically might take place (or at least one that is argued for in the literature), namely that inferring disorder diagnoses as syndromes from specified clusters of necessary and sufficient sets of symptoms is no longer used, imagine that instead we only diagnose present symptoms. The rationale behind this could be, for example, that we can better target specific symptoms with specific interventions than syndromes that allow for very heterogenous clinical presentations under one label (Park et al., 2017). In this case, the overall model-based proposal would be no longer be correct but would contain superfluous components. Of course, superfluous components (i.e., everything that goes beyond symptom diagnostics) could be cut out to make the proposal adequate again, but for the time being it would be inadequate. This shows that my model-based proposal is in principle robust to some changes on the higher level of diagnostics (disorder diagnostics) but would also be open to falsification if deeper changes were to take place. Now we can move on to consideration how the model-based theory of diagnostic reasoning can handle changes in the context of the evaluation of symptoms and signs.

Whatever changes take place on a level of diagnostics higher than the level of symptoms – whether changes in the taxonomy of syndromes or a whole new way of making of attributed symptoms – they do not affect the way in which symptoms themselves are evaluated. However, there might also be changes in diagnostics that

8 Such past decisions regarding single changes in the diagnostic taxonomy, as well as the whole diagnostic approach of syndromal diagnosis based on symptom clusters (now supplemented with dimensional diagnostics of certain symptomatic features), have been heavily criticised by researchers, clinicians, and philosophers (e.g., Kendler and Parnas, 2012; Casey and Kelly, 2013; Demazeux & Singy, 2015; Hengartner and Lehmann, 2017; Ghaemi, 2018). But regardless of the validity of concrete categorisations of disorder entities, the delineations between them, or even the whole approach of syndromal diagnostics, diagnostic practice must apply it.

would influence the way we would identify symptoms. The way this may occur is through changes in how we understand these symptoms. Such changes in understanding may, on the one hand, lead to change regarding what we look for to evaluate the presence of a symptom by our usual means of diagnostic information-gathering and use, or it may be that our changed psychopathological understanding is accompanied by new means of evaluating the presence of a symptom. I will discuss both cases considering the model-based approach I have proposed, beginning by showing how the model-based account would accommodate for the first case: changing understanding with no general change of diagnostic approaches.

The idea in this case would be that our ways of grasping psychiatric symptoms *via* propositional models used to evaluate the presence of such symptoms, would change in so far as those propositions in the model change. However, despite modifying the model structure that we then use we would still follow similar process of screening, in-depth diagnostic information-gathering, and conclusion-drawing. To make this possibility more vivid, let us consider a historical example and ask how these different understandings would have been used in the context of temporary diagnostic reasoning as explained by the model-based account. Let us look at anhedonia.

As Berrios and Olivares (1995) point out in their historical investigation of anhedonia, we have seen many understandings of this symptom in the past hundred years or so. Although the phenomenon itself was described and discussed earlier, it was Ribot (1897, p. 53) who coined the term anhedonia and characterised it as a general inability to experience pleasure, found in individuals suffering from melancholia. Since then, anhedonia has been described clinically as present in patients suffering from depressive disorders as well as psychosis (especially schizophrenia) (Pelizza and Ferrari, 2009; Lambert et al., 2018).

Earlier discussion of ostensibly the same clinical phenomenon can be found in Griesinger (1861), calling it "mental anaesthesia": a state in which "the patient can no longer rejoice in anything, not even the most pleasing" (ibid., p. 223). Going into more detail, he described this phenomenon as a "continual dissatisfaction with the external world" and as involving "abnormal states of emotional dullness [Gemüthsstumpfheit], and even of total loss of emotions [volligen Gefühllosigkeit]" (ibid., pp. 66–67).

Later authors, not picking up on the term anhedonia, described the same phenomenon differently again. Kraepelin (1919, p. 33) wrote:

> The singular indifference of the patients towards their former emotional relations, the extinction of affection for relatives and friends, of satisfaction in their work and vocation, in recreation and pleasures, is not seldom the first and most striking symptom of the onset of disease (dementia praecox). The patients have no real joy

> in life, "no human feelings"; to them "nothing matters, everything is the same"; they feel "no grief and no joy", "their heart is not in what they say".

Jasper (1963, p. 93) talked about a clinically relevant "feeling of having lost feeling" (das Gefühl der Gefühllosigkeit) in which "patients complain that they no longer love their relatives, they feel indifferent to everything. Food does not gratify. […] All sense of happiness has left them. They complain they cannot participate in things, they have no interest".

Myerson (1920) and others picked up on the term *anhedonia*. Myerson proposed an understanding of the phenomenon in light of a developmental model, summed up by Berrios and Olivares (1995, p. 463):

> [F]irst, by the disappearance or the impairment of the appetite for food and drink and failure in the corresponding satisfactions […] Second, there is a failure in the drive or desire for activity and the corresponding satisfaction.… Third, the appetite or desire for rest and the satisfaction of recuperation are also involved in the anhedonic syndrome. The tired feeling […] may be supplanted by a final absence of the feeling of fatigue.… Fourth, the sexual drives and satisfactions are conspicuously altered in the acquired anhedonic states. […] Finally, the social desires and satisfactions, which belong indissolubly to the nature of the herd animal known as man, become disorganised, deficient and even destroyed.

Klein's (1974) understanding arguably went on to have the largest impact on the understanding of anhedonia that made its way into the DSM-III and later editions (De Fruyt, Sabbe, and Demyttenaere, 2020). He described anhedonia as "a sharp, unreactive, pervasive impairment of the capacity to experience pleasure or to respond effectively to the anticipation of pleasure" and as "a phasic, temporary, severe lack of present or anticipated satisfaction associated with the conviction that one cannot perform adequately" (Klein, 1974, p. 175). Later, Klein (1987) also added two dimensions to pleasure and its loss, distinguishing between consummatory pleasure, which is the pleasure of consuming or doing something that should be expected to bring pleasure, and appetitive pleasure, which is the pleasure gained from the expectation of a future usually pleasurable stimulus.

Considering this sample of historical views on what constitutes anhedonia as a symptom of mental disorder, linking those making similar proposals, and translating them into a propositional model would result in five different model: the Ribot model, the Griesinger model, the Kraepelin-Jasper model, the Meyerson model, and the Klein model. According to the Ribot model, the only proposition that would have to be shown to apply to a patient to justify the attribution of anhedonia is that the proposition "fully lacks the capacity for consummatory pleasure" applies to an individual. According to the Griesinger model, the propositions to apply to a pa-

tient would be that the patient has "dullness or loss of emotional reactions" and a "permanent state of dissatisfaction". The Kraepelin-Jaspers model would require the proposition "no expression or repot of emotional experience", "general indifference to occurrences in the surrounding world". The Meyerson model would require that the content of the following propositions apply to the patient and have arisen in the stated order: "loss of appetite and pleasure in food", "loss of drive for activity and the corresponding satisfaction", "loss of desire for and enjoyment of relaxation", "loss of sexual drive and satisfaction from sex", "loss of interest in and satisfaction from social interactions". And finally, the Klein model requires three propositions to apply, namely "loss of consummatory pleasure", "loss of anticipatory pleasure", and "believing that one would perform poorly in usually pleasant activities".[9] In contrast with these historically informed models we may also consider the diagnostic features of anhedonia in the DSM-5 text revision. Here, we have a list of features, where each of the features, separated by a comma, would make one proposition of the model:

> Feeling less interested in hobbies, not caring anymore, not feeling any enjoyment in activities that were previously considered pleasurable, reduction from previous levels of sexual interest of desire. Family members may notice social withdrawal or neglect of pleasurable avocations. (APA, 2022, p. 187)

Considering all these propositional models, including the current DSM-5 presentation, we can imagine how information sufficient to plausibly accept or reject the relevant propositions can be gathered by means of behavioural observation and interviewing of patients and conversations with relatives and friends (i.e., the typical current means of information-gathering), and therefore that while each of the models could in principle be adopted to determine the presence of anhedonia, all that would have to change for this would be the propositions to be evaluated in the otherwise similar diagnostic process. We would still use the same type of model and the same means of evaluation. This little look into the history of psychiatry therefore seems

9 Note that while all these models address anhedonia, they do not consider its occurrence in the context of the same disorder. Kraepelin's comments consider the occurence of anhedonia in dementia praecox (schizophrenia) while Klein describes anhedonia in the context of depression. Whether the psychiatric symptom of anhedonia in both patients is indeed the same across contexts which is usually assumed in the literature (e.g., Harvey et al., 2007; Pelizza and Ferrari, 2009) and also in the DSM-III, is challenged by more recent neuroscientific research. A better understanding of the neurobiology of anhedonia (Kuhlmann, Walter, and Schläpfer, 2013; De Fruyt, Sabbe, and Demyttenaere, 2020) begins to suggest that the cross-diagnostic symptom anhedonia may indeed represent two different conditions in the contexts of schizophrenia and depression. In depression, anhedonia may be characterised by impairments in anticipatory pleasure and integration of reward-related information, while anhedonia in schizophrenia is associated with neurocognitive deficits in representing the value of rewards (Lambert et al., 2018; Liang et al., 2022).

to support the idea that the model-based account shows some robustness, allowing us to integrate some changes on the level of symptom diagnostics and helping us to understand how they are integrated.

Instead of going down this very speculative path, I would like to bring up an example that seems more likely to be relevant to psychiatric diagnostics in the near future and see whether the methods accompanying it would necessarily or likely make the framework of model-based diagnostic reasoning obsolete. For this I will look at computational psychiatry.

Computational psychiatry as a field of research "consists of applying computational modelling and theoretical approaches to psychiatric questions" (Seriès, 2020, p. 12).[10] In this way, "Computational Psychiatry seeks to understand how and why the nervous system may process information in dysregulated ways, thereby giving rise to the full spectrum of psychopathological states and behaviors. It seeks to elucidate how psychiatric dysfunctions may mechanistically emerge and be classified, predicted, and clinically addressed" (ibid., p. 13).

In this endeavour, computational psychiatry came to merge insights and methods from the field of computational neuroscience – itself concerned with "formaliz[ing] the biological structures and mechanism of the nervous system in terms of information processing" (Seriès, 2020, p. 10) in terms of mathematical models – with recent changes in approaches to research in psychopathology, especially the research domain criteria (RDoC) (Cuthbert and Insel, 2013). RDoC is a research framework attempting to move beyond the supposedly stagnating current approach to psychopathology and treatment, by substituting the focus on psychiatric syndromes with a focus on mechanisms of specific dysregulations of cognition and behaviour relevant in the context of psychopathology. This approach was supposed to be better suited to integrating into psychiatry the increasing amount of knowledge gained from research on neural systems and behaviour in clinical and non-clinical populations. With this focus, RDoC and the attempt to use computational neuroscience for the purpose of psychiatric research have immense synergies, making them natural partners. As Seriès (2020) puts it:

> Rather than considering psychiatric diagnosis a cluster of symptoms, RDoC functional domains and constructs can be conceptualized as resulting from sets of underlying computations taking place across interacting neural circuits. In theory, these neural processes can, in turn, be described by algorithmic representations that describe information processing in the system. (p. 9)

10 Other earlier bird's-eye-view discussions of computational psychiatry can be found in, e.g., Montague et al. (2012), Walter (2013), and Friston et al. (2014).

Hence these neural processes could be described in terms of computational models, as used in computational neuroscience. Questions that research may at least in principle be able to address by pursuing these pathways would be questions such as "*What* are the main biological components involved in psychopathology and what are the mathematical relationships between these?", "*How* do dysfunctions in the individual biological units or in their interactions lead to the behavioral changes seen in mental illness?", and "*Why* have these changes occurred?" (ibid., p. 13).

Within computational psychiatry, we can differentiate between two broad classes of computational modelling: data-driven and theory-driven (Huys, Maia, and Frank, 2016). In data-driven modelling, machine learning is applied to large, multidimensional datasets from psychiatric patients, including genetic, neuroimaging, behavioural, and self-report data, and without considering any pre-established psychological or biological theories. Instead, the algorithm is supposed to find novel associations within the data structure that might give rise to new theories. Theory-driven approaches, on the other hand, attempt to provide a mathematical description of relations between types of behavioural performance or self-reports of psychiatric subjects and the performance of relevant biological mechanisms (such as brain anatomy or physiology) or higher-level functions (such as perception and learning) assumed to be relevant based on what we already know from previous work in computational neuroscience. By comparing the performance in self-report and behaviour with the underlying biological mechanisms and cognitive functions in healthy and clinical populations we may then generate a computation model of the dysregulations occurring in the clinical population.

Among the many examples of how computational psychiatry may in the future impact clinical diagnostics, I will select one from the branch of theory-driven computational psychiatry, and via this route return to my previous example, anhedonia. Anhedonia has more recently become an object of investigation in computational psychiatry (Kuhlmann, Walter, and Schläpfer, 2013; Huyes et al., 2013; Lambert et al., 2018; De Fruyt, Sabbe, and Demyttenaere, 2020; Walter, Wellan, and Daniels, 2020; Walter, Daniels, and Wellan, 2021; Liang et al., 2022).

Insights from research on reinforcement learning, including its neurobiological basis[11] and its relation to the phenomenon of pleasure, are especially important

11 Reinforcement learning is a strand emerging from the combination of two longstanding areas of theory: control theory and learning theory (Dayan, 2002). Control theory is an area of mathematics in which one attempts to provide value functions and dynamic programs that achieve optimal control of a dynamical system's behaviour. For this purpose, the theory attempts to identify a suitable control law for a system such that a given optimality criterion is matched by the system if the system is manipulated accordingly (Sutton and Barto, 2018). Learning theory, on the other hand, focuses on learning from trial and error and originated in psychology and the early investigations of animal learning in terms of Pavlovian (classical) and instrumental (operant) conditioning (Resorla, 1988; Staddon and Cerutti, 2003).

for this research. In the context of research on pleasure and its disruptions, phenomena are often considered in terms of the classical so-called pleasure cycle (Sherrington, 1906; Craig, 1918) assuming an appetitive phase (wanting), signified by the motivation for or the incentive salience of a reward; a consummatory phase (liking), signified by the pleasure of an actually achieved reward; and a satiety phase (learning) signified by representations and predictions about future rewards based on past experience (De Fruyt, Sabbe, and Demyttenaere, 2020). This basic model has been further developed by Rizvi et al. (2016), who describe the reward process as initially building a stimulus–reward association, which then leads to interest (wanting a reward), anticipation (a state of readiness for a reward), motivation (initial energy expenditure to attain a reward), effort (sustained energy expenditure to attain reward), hedonic response (enjoyment of reward), and feedback integration (updating reward presence and values). These aspects map quite well onto the aspects of the RDoC construct of positive valence systems: reward valuation (reward, delay, effort), reward responsiveness (reward anticipation, initial response to reward, reward satiation), and learning (probabilistic and reinforcement learning, reward prediction error, habit) (NIMH, 2018). On the neurobiological level, several regions are relevant, especially in the mesolimbic reward system consisting of a network of parts of the ventral tegmentum, the nucleus accumbens (part of the ventral striatum), and the amygdala (Schultz, 2002).[12] These regions are connected by dopaminergic signalling that seems to play a major role in reward-directed and consummatory behaviours in rodents as well as humans in general (Berridge and Robinson, 1998; Schultz, 2002; Egerton et al., 2009).

In however fine-grained a way we decide to think about anhedonia – whether we go with Rizvi and colleagues (2016) or with those researchers preferring a three-part model of wanting, liking, and learning (Bossini et al., 2020) – we end up with an understanding of anhedonia that, compared with that assumed in the DSM-5

Later evidence from lesion studies, pharmacological interventions, and imaging studies in animals and humans linked reinforcement learning with brain structures and functions of neurotransmitters, especially dopamine (Schultz, Dayan, and Montague, 1997; Heinz, 2017; Bogacz, 2020).

12 Besides these classically mentioned regions, other brain areas also appear to code and perhaps contribute to pleasure processing: for example, one site of the mid-anterior and mid-lateral part of the orbitofrontal cortex seems to track changes in subjectively reported pleasure (Kringelbach, 2005). For an overview of further regions and their (potential) implication in reward and pleasure processing, see Ellingsen, Leknes, and Kringelbach (2015). Due to the involvement of regions such as parts of the frontal lobe, researchers have proposed an alternative to the mesolimbic reward system in the form of the frontostriatal reward-processing network in frontal areas such as the ventromedial prefrontal cortex (vmPFC), orbitofrontal cortex (OFC), and midbrain limbic areas, including the ventral striatum (VS), insula, and thalamus (Sescousse et al., 2013).

discussed earlier (which assumes an impairment in wanting and liking), has more components, and therefore has more propositions whose presence might be evaluated as part of a propositional model to determine the presence of anhedonia. But since we are interested in how the improved understanding of anhedonia *qua* computational psychiatry might also impact the ways in which we diagnose, let us focus on this, instead of on the changes that we would see in a potentially new propositional model.

To return to diagnosis, let us look at studies that have used tasks to investigate the presence or absence of certain behavioural patterns and neural features in individuals suffering from anhedonia. Let us focus on research regarding the *wanting* component of anhedonia. Studies interested in this aspect have employed a variety of behavioural tasks, such as the "effort expenditure for rewards" task (Treadway et al., 2009), effort-based cost/benefit valuation tasks (Croxson et al., 2009), incentive motivation tasks (Anselme and Robinson, 2019), the "monetary incentive delay" task (Lutz and Widmer, 2014), reward-guessing tasks (Ubl et al., 2015), the wheel-of-fortune task (Dichter et al., 2009), and a slot-machine task for reward anticipation (Fryer et al., 2021).[13] While scientific evidence collected in these investigations is still not extensive, several interesting findings have been generated. I will focus on one of these. As a meta-analysis has shown, there are patterns of middle frontal gyrus and anterior cingulate cortex hyperactivation, as well as caudate hypoactivation, during different reward-anticipation tasks carried out with MDD patients, including monetary incentive delay tasks, card-guessing tasks, and wheel-of-fortune tasks (Zhang et al., 2013).

If we assumed for a moment that these findings are valid, in the sense that brain activation in individuals carrying out these tasks would show patterns of middle frontal gyrus and anterior cingulate cortex hyperactivation as well as caudate hypoactivation across these tasks if they suffered from the liking component of anhedonia, then these tasks combined with neuroimaging could be included in clinical diagnostic procedures to evaluate whether patients suffer from the symptom of anhedonia. The evaluation of this symptom would no longer be based on behavioural observations and self-reports of patients; instead, an objective bio-neuro-cognitive test could be used as part of the evaluation. Staying with this example, we may ask, would this step in the evaluation of anhedonia (or a similar step in this direction for any other psychiatric symptom) change the diagnostic procedure as described in my elaboration of my model-based account? The answer is: not necessary, but possibly.

Not necessarily, because the new psychopathological understanding of anhedonia can also be taken to offer the material for a different set of propositions telling us what it means for a patient to suffer from anhedonia and therefore for an alternative

13 For systematic overviews of behavioural tasks in combination with neuroimaging for the evaluation of reward processing, see Borsini et al. (2020) and Geugies et al. (2022).

constitutive propositional model of anhedonia. What may change given our neurobiological insights would then be an aspect of the assessment. After the screening phase of diagnostics that suggests a complaint that might be the psychiatric symptom of anhedonia, instead of evaluating this possibility by asking the patient questions or talking to their relatives, we might implement neuro-behavioural testing. If, for example, we took the proposition "Shows significant lack of motivation for initial energy expenditure to attain a reward (wanting component)" to be part of a propositional model of anhedonia, and we would accept that this lack is realised by a certain pattern of neural activity shown across monetary incentive delay tasks, card-guessing tasks, and wheel-of-fortune tasks. We might use these tasks and the recordings of brain activation patterns to evaluate the applicability of the proposition and thus the fit of this aspect of the model, via objective testing instead of self-report in the context of interviewing. Thus nothing changes in the overall order of diagnostic evaluation steps that I discussed in earlier chapters, and nothing about the use of model's changes. Only the means by which propositions are evaluated changes from interviewing to the new means of objective biological and cognitive testing – which, though so far for only a few psychiatric conditions, is sometimes already assumed to be part of the evaluation in the model-based approach.[14] In conclusion, it seems that changes that might occur as a result of developments in computational psychiatry could be readily integrated into the framework I have presented with my model-based account. However, when I said that our changing understanding of anhedonia would not necessarily change the procedure of diagnostics such that it would endanger the model-based account, I left open the option that it could do so. Let me come to this possibility now.

There are changes deriving from research in computational psychiatry – for example, in the research on anhedonia discussed here – that might in principle lead to changes in the overall diagnostic procedures in psychiatry that would make the account of psychiatric diagnostics discussed here obsolete. This would be the case if these changes impacted overall diagnostic practices and what is considered proper

14 For more examples of how computational psychiatry might inform diagnostics in a similar manner (i.e., by new means of evaluating diagnostic propositions), see Słowiński et al. (2017). They propose social biomarkers for identifying motor abnormalities that contribute to the deficits in nonverbal behaviours and in nonverbal synchrony that impair the structured and unstructured social interactions of schizophrenia patients, and that supposedly underlie patients' feelings of incompetence, confusion, and overwhelm in social contact, leading to the social withdraw of typical schizophrenia patients. The behavioural biomarker they use is motor behaviour in a "mirror game", a coordination task in which two partners are asked to mimic each other's hand movements, where the partner is a computer avatar or humanoid robot. With the help of statistical learning techniques applied to participants' movement data, they were able to provide a classification with 93% accuracy and 100% specificity.

diagnostic evaluation. Let us consider a few examples. If, for example, we developed neuro-cognitive objective tests for every single psychiatric symptom, it would in principle be possible to do no screening with patients as a method for deciding which potential psychiatric symptoms we should do an in-depth evaluation of. Instead, we might immediately have every patient do all the objective tests. We could move directly to the in-depth evaluation. While this might still be understood as an evaluation of the applicability of diagnostic models, this shift would change the procedure I discussed in the last chapter because there would no longer be a screening phase. As a result, the model-based account as it stands would be inadequate. Or take the current physiological and biochemical candidates for diagnostic biomarkers of major depressive disorder (e.g., Targum et al., 2022) or some of its symptoms (e.g., Stout et al., 2022) as measurable in clinical contexts. If they turned out to meet the specificity and sensitivity requirements for use in clinical contexts, they might supplement our current clinical practices. After identifying initial complaints that might indicate symptoms of depression, or that might point towards psychiatric symptoms that can occur in the context of major depressive disorder, we might then simply order the physiological or blood tests relevant to evaluate this possibility, providing us with a clear negative evaluation of whether the symptom or disorder in question is present. No mental modelling process, no comparing models to clinical observation, no evaluations of alternative sets of propositions that are part of qualitative models of symptoms would take place. Although there are still a number of problems in the pursuit of diagnostic biomarkers – such as underpowered and biased studies (Carvalho et al., 2020) for transdiagnostic biomarkers and low test-retest reliability and strong response to placebo intervention in psychophysiological biomarkers (Rapp et al., 2022), as well as ethical concerns (Glannon, 2022) – overcoming these obstacles and establishing biomarkers for clinical use would mean major progress in psychiatric diagnostics. If genuine, such progress would make my account a matter of philosophy of the *history* of psychiatry. These examples suffice to show the sensitivity of the model-based account to larger changes on the level of symptom diagnostics. Next, in order to underline the account's sensitivity to large-scale changes, let me come to changes in psychiatric diagnostics that are perhaps more unlikely to occur but are at least conceivable, and that might render the model-based account obsolete.

So far, I have focused on somewhat more realistic changes in psychopathology and clinical assessment that one might argue are already detectable in the current psychiatric literature. Now let me come to more extreme potential changes that would rapidly transform psychiatric diagnostics. These examples will make the point that in principle, such changes may falsify the model-based account. Let us consider two such scenarios. I will call the first one the Place-Feigl-Smart psychiatry scenario, the second one the Churchlandian psychiatry scenario.

What I call the Place-Feigl-Smart psychiatry (see Place, 1956; Feigl, 1958; Smart, 1959) would take place if two things were true. First, if the identity theory of mind and brain (i.e., types of mental states are identical to types of brain states) were correct, at least for those mental states interesting for psychiatry. Second, if we attain complete knowledge about how brain states and psychopathological mental states relate, such that these mental states and the behaviors they exhibit are fully intelligible in terms of structural or functional brain features. If this were the case, we would no longer need self-report, behavioral observation, or anything else from the patient. We would simply have to investigate their brain (let's say with some kind of neuroimaging) and let a program identify the present brain features that would then tell us what symptoms are present in the patient.

Alternatively, we may in principle end up with a Churchlandian psychiatry (Churchland, 1981) in which, since all talk of the mental in our language would be abandoned for brain talk anyway (to adopt Churchland's sketch of the future), pure brain and behavioural talk would also be all that we have when we talk about symptoms. Then mental symptoms would be out of the game and in their place we would have talk about brain states whose presence could be evaluated again by investigating the brain.

Although such radical scenarios seem unlikely to occur any time soon – even if the metaphysical framework that would have to be true to allow those scenarios to become reality were shown to be correct – what we can take from these two examples is that straightforwardly reading off symptoms from brain data would certainly make obsolete all the steps of the model-based account as spelled out here. When direct inference from brain data to psychopathological mental states which are mental symptoms or causes of pathological behaviour is possible, no modelling efforts as described by me seem necessary. We can also conclude that if we were, in a Churchlandian manner, to abandon mental talk entirely, the model-based account would collapse because we would drop talk about mental symptoms that need diagnosing from our diagnostic approaches altogether. Thus there would no longer be any efforts to engage in modelling to evaluate whether mental symptoms are present. The model-based account as presented would clearly be obsolete in both cases. Hence psychiatry could change in ways that would make the model based account an inadequate proposal to understand psychiatric diagnostic reasoning.

In conclusion, it appears that the model-based account is sensitive to changes in the reality of psychiatric diagnostics but at the same time general enough to encompass certain potential changes in psychiatric diagnostics. It is in touch with the reality of diagnostic practice and is thus a falsifiable theory of psychiatric diagnostic reasoning that is also not so overfitted that it loses all robustness against change. There is a spectrum of changes that it could integrate and accommodate.

4.8 Conclusion

In this chapter I discussed how the model-based account addresses the desiderata for a theory of psychiatric diagnostic reasoning providing an answer to the Methodological Question. I discussed whether the model-based account can be comprehensive and cognitively realistic, whether it helps us make sense of the difference between misdiagnosis and diagnostic malpractice, and whether it can account for the occurrence and resolution of diagnostic uncertainty, and concluded that it performs well in all these domains. Moreover, I argued that it helps us to understand and evaluate the phenomenon of good instinctual diagnostics and the occurrence and resolution of diagnostic disagreements. For each of these points, I set out how the model-based account fulfilled the criteria and thus meets all desiderata. In the next and final chapter, I will discuss alternative accounts to the whole of psychiatric diagnostic reasoning or aspects of it, and compare them to the model-based account to show the advantages it has over them.