Research statement: future work

My dissertation asks what one of the most common methods in philosophy, *reasoning from examples*, can show. When employing this method, a philosopher describes a case, asks what readers judge about it, and treats their judgements as evidence for or against a general claim. I argue that every form of the method faces challenging problems to do with how difficult it is to transform amorphous real-world situations into well-defined cases or problems. My future research extends this work in three directions. These connect it with the study of literature, machine learning and decision theory.

The first strand concerns a way of reasoning from examples that the dissertation puts under pressure. Many philosophers treat cases as experiments, comparing cases that differ in a single respect so as to isolate the contribution of that respect. Others, especially in ethics, work with cases that are richly described and often literary, and reach a verdict by interpreting the case as a whole (Murdoch 1970; Nussbaum 1990; Diamond 1991). Their cases need not be rich, however. An impressionist painting can represent an experience or a person better than a photorealistic image, and a sparsely drawn case can in the same way be more accurate than a detailed one, partly because of what it leaves out. This is the 'interpretive' tradition.

I will ask which of the problems raised by the dissertation this interpretive way of reasoning avoids. It makes no attempt to isolate the contribution of a single feature, and so is not committed to guarding against the errors that such attempts invite. I will ask what precedent it has. For Dewey (1938), turning an unsettled situation into a well-defined problem is itself a stage of inquiry, and my dissertation is largely concerned with that step. Casuistry and the law's reasoning from precedent supply further models of how particular cases can yield general understanding.

Additionally, I will ask how far the careful study of literary texts can answer questions of philosophical interest. Some philosophers of literature deny that literature's value lies in the truths it conveys (Lamarque and Olsen 1994). Literary study might nonetheless identify which features of a situation matter, and so which features any further representation of it should include. If it can, it would have a use even in formal modelling, where a decision about which features to represent must be made *before* any model is built. The question connects philosophical methodology with the philosophy of literature and with literary studies.

The second strand concerns whether the failure modes of machine learning have analogues in philosophy. The dissertation already draws its analogues from classical statistics and causal inference. Machine learning has since documented failure modes of models that learn from examples, and reasoning from examples is a way of learning from them. On some accounts, intuitions are themselves the outputs of statistical learning (Railton 2014). I will ask which of these failure modes have analogues in philosophy, and what the remedies developed for them would require of philosophers. Williamson (2024) and my paper on the tradeoff about detail have addressed overfitting; I intend to examine at least four others.

The final strand asks how a reasoner should respond when she discovers distinctions or possibilities that she had overlooked. A doctor may learn that a condition she had treated as one comes in two forms; an investor may learn of a risk she had never considered. Such discoveries can change what a reasoner believes and chooses, in ways that standard Bayesian updating, which assumes a fixed set of possibilities, does not govern (Bradley 2017). Philosophical verdicts are exposed to the same kind of change, since a verdict about a case can reverse once a distinction that the case ignored is drawn. I will ask when a verdict survives the drawing of new distinctions and, where it does not, what determines the level of detail at which it should be assessed.

One answer, which the dissertation motivates, is that the right level of detail is that of the model that best represents the reasoner's situation, given what she is trying to find out. The task is to show that this answer gives definite guidance in the cases that matter. Reasoners seldom have a complete list of the possibilities. In science and public policy as much as in philosophy, an answer would have wide application.

The three strands form a single programme. The dissertation offers suggestions for how to turn a real situation into a formal problem, but leaves open two questions: which features of a situation a model should represent, and how finely it should divide the possibilities. The first strand approaches the first question through the interpretation of richly described cases, and the third approaches the second through decision theory. The second extends the dissertation's diagnostic use of statistics to the failures documented in machine learning.