To build a successful feature extractor, we have to utilize domain knowledge. One approach to this is to parametrize the extractor so that it contains a useful inductive bias, i.e. convolutions for computer vision and attention for natural language processing. However, it is often not clear how to parametrize our domain knowledge. For example, knowing that various augmentations of our data should not change the corresponding representations is intuitive to us, but sometimes difficult to parametrize in feature extractors. In such cases we resort to pretext tasks that modify the training procedure instead, so that the feature extractor learns to be invariant to specified augmentations. Constructing pretext tasks that capture our intuition and domain knowledge is crucial for surpassing supervised learning. Since pretext tasks do not require labels, we call this unsupervised training *self-supervised*, since the model sets its own pretext tasks. I believe that approaching self-supervised learning broadly holds much promise for the future of deep learning, because extracting fundamental self-supervised learning principles would not only mitigate the issue of learning from limited data, but it would also enable us to better utilize distilled knowledge of basic research in natural science. Through my work I hope to enable scientists to incorporate their knowledge into deep learning models without the need for labeled data. Working on this through self-supervised learning would scale their work rapidly.

The *contrastive* objective as a pretext task is particularly important, not only due to its resurgence for learning representations for computer vision, but also because it has the potential to become the “go-to” approach for learning representations for scientific discoveries. For example, in physics the research community has identified symmetries that provide considerable information regarding systems’ properties. Exploiting such symmetries via data augmentation and contrasting the representations is a useful way to incorporate our physics knowledge into deep learning models. We are beginning to observe the fruits of contrastive learning for photonic crystals, but we believe that the potential for this approach is much larger. Hence, it is important to answer certain questions about contrastive learning in general. I discuss these questions below and explore them in parallel through my work.

- Why do current contrastive learning methods, such as Bootstrap Your Own Latent (BYOL) work, given that they only rely on similarity of positive examples, i.e. why does BYOL not collapse to a single mode? Answering that question is important as not having to rely on negative examples is great, because often it is unclear what a negative example should be (and it even turns out that some negative examples should be *positive* for downstream tasks). I have a conjecture that the success of this method comes from good initial representations, and I am actively exploring this by tuning properties of datasets and architectures. Currently, I am replacing CNN feature extractors with fully connected ones, and from preliminary results I see that we can bootstrap (without collapsing) even from fully connected feature extractors. Another strange thing about BYOL is that it sets up an optimization problem, but it actively avoids arriving at the global optimum. This type of optimization is particularly suitable for deep learning since it exploits some of the properties of neural networks that were recently suspected to be drawbacks. For example, it is hard to learn the identity with a neural network, which turns out to be useful for BYOL, since if the predictor part of the model can easily become an identity, then a mode collapse is incentivised. In my opinion, rethinking the optimization problems that we solve with BYOL, i.e. setting an optimization objective where we actually aim for the global optimum is necessary, since it will help us to think about contrastive learning in simple terms. To achieve a better and more general BYOL I am exploring more principled repulsive and attractive regularization terms in the feature space with the hope that I arrive at a principled optimization objective that also does not rely on negative examples.

- How to do self-supervised learning with a single epoch? Arguably, a single epoch is a more realistic scenario (as humans we cannot afford to replay some experiences over and over again), where we strive to get the most from seeing an example only once in a dataset. Recent studies have demonstrated that a single image with heavy augmentation can tell us a lot about the image manifold. I believe we should exploit that phenomenon for realistic self-supervised learning via a *single* epoch. However, if we see a datapoint only once, then we need to augment it heavily, but taking expectations over the infinite pool of augmentations is costly. I propose that we learn those expectations through a neural network module. A predictor module that takes a feature vector and returns a feature vector in expectation would be very useful in contrastive learning, and recent empirical work already suggests that such a component is necessary for the success of BYOL and other related models. I can pre-train such a predictor on existing pools of data and augmentation, and then enforce it during self-supervised learning. The conjecture is that pre-training the predictor would allow us to get useful representations from a single epoch *without* heavy augmentation. When pre-training the module I would aim to make it transferable, so that it does not memorize the input data points, but rather focuses on learning how to predict expectations over augmentations. This assumes that the features, which the predictor receives, already contain useful information about the data point, which is in line with recent work on the “deep image prior,” i.e. that even randomly parametrized convolutions already capture useful low-level information about images. If my attempt is successful, then I would focus on contrastive learning beyond image data as, in principle, my proposal is data agnostic. Then, I would focus on improving feature extractors via such predictors in a lifelong learning setting, where the data distributions evolve with time.

- Can we find unknown symmetries through self-supervision? Recent studies have made initial steps in parametrizing and meta learning equivariances. This is an important direction, because it would allow us to discover novel properties of physical systems. Designing efficient parameterizations of data augmentations is an important next step for scalability, and here I will draw upon my past work on efficiently integrating unitary parametrizations for improving RNNs. Furthermore, designing the parametrization so that it can be learned by, say the contrastive objective, would be another crucial step since labels are not needed. Here, I would start with existing parametrizations of equivariance. Then I would perform self-supervised learning using augmentations coming from these parameterizations, and using early stopping on suitable measures of feature quality, I would derive a criterion that guides the training of the parametrizations. It is possible that along the way I find that parametrizations using complex or quaternion numbers are more efficient both for more sophisticated symmetries and datasets coming from a single distribution, which I could study in depth through the feedback loop that I propose. Coming up with general parameterizations and testing them on scientific datasets in natural science will be the next step, and then inspecting the learned parametrizations could help us revisit our knowledge about these sciences, or even perhaps learn new scientific facts.

Finally, due to the fundamental nature of the above questions, answering them through my proposed action plans can yield further advances across deep learning. Results from developing my ideas on setting up good optimization problems for self-supervised learning, using data efficiently in a single epoch, and distilling (discovering) knowledge in basic scientific research will further be fruitful to semi-supervised learning, active learning and beyond.