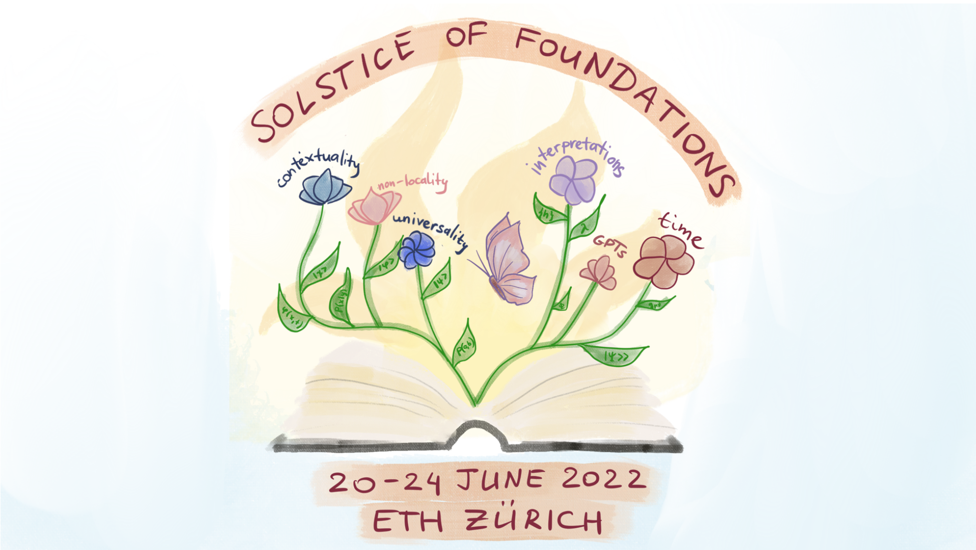
Solstice of Foundations 2022
ETH Zürich

Process Theories



Taking a diagrammatic approach to general physical theories

John H. Selby
20/06/2022

Lecture Notes

Abstract

In these lectures I'm going to introduce the formalism of process theories as a framework for studying general physical theories. This is a diagrammatic formalism, which, as we will see in the second lecture, subsumes the more widely studied framework of generalised probabilistic theories. This provides a complementary perspective which focuses on composition of processes, which can be contrasted with the standard perspective which focuses on the geometry of single systems. In the final lecture we will zoom out and consider the landscape of physical theories, and see how there is an interesting structure to this landscape which allows us, amongst other things, to talk about emergence of one theory from another.

*** WARNING ***

This is a rough first draft of these lecture notes. Proofs of many of claims are sketches or in places non existent, and there is an appalling lack of citations. I hope to fix these issues by the end of the summer school and get a new version of these notes out to you soon. Any feedback on these notes is therefore very much appreciated!

Contents

Lecture 1

The basics

It is well known now that quantum theory is a fantastically successful physical theory. From its origin as a weird explanation for a few peculiar phenomena, it has spread to become the foundation of all of modern physics[1]. It has passed every experimental test that we have thrown at it and now forms the basis of a vast range of technologies, and promises to lead to even more.

In these lectures, however, I'm not actually going to talk about quantum theory very much. Instead, I'm going to talk about alternative physical theories, that is, hypothetical ways that the world could be if it were not quantum. Given the fantastic successes of quantum theory you might well be asking yourself why you should care about these alternatives. Why should we care about theories which have absolutely no experimental support and which have not led (at least directly) to any technologies what so ever?

I think this is a pretty fair question so I'd like to start by trying to provide some motivation for looking at such theories.

1.1 Five reasons to study general physical theories

Reason 1: Quantum theory is probably not fundamental.

From a historic perspective, there is no good reason to believe that we should be at the unique point in time where we have happened to stumble upon a fundamental theory of nature, it seems much more likely that just as with all of our other theories that we have developed, that so too will quantum theory ultimately be replaced by a deeper theory of nature. Let us not be as naive as the Nobel prize winner Michelson, who in 1903 claimed that classical physics was complete, only to have to eat his words two years later when the photoelectric effect was discovered.

The difficulty that we have with unifying quantum theory with our other best physical theory, namely General Relativity, should also give us pause. Something is going to have to give, and it seems likely that both quantum theory and general relativity will need to be modified to some extend to achieve a satisfactory unification.

[1]With the notable exception of general relativity which we will talk more about later.

It is therefore interesting to study a general framework which can describe hypothetical theories, because it might be the case that one of them could one day replace quantum theory as our best theory. It might be unlikely that we can, absent experimental evidence, pinpoint which theory is more fundamental, but perhaps we can learn something about it anyway.

Alternatively, if quantum theory does happen to be fundamental, then should there not be some explanation for this? For example, could it be that all other theories are logically inconsistent? By trying to push beyond quantum theory we could also hope to find such an explanation.

Reason 2: To learn more about quantum theory itself.

Once we have situated quantum theory in some space of general physical theories, we can start to understand quantum theory itself on a deeper level. For example, we can start to ask, what distinguishes quantum theory from all of the alternatives? Can we find some set of physical principles which single it out? Ultimately, we can hope to find some explanation for why quantum theory seems to describe nature so well.

Moreover, we can also start to explore the possibilities and limitations of quantum theory for various information theoretic tasks, and, hence, what the scope is for developing quantum technologies. For example, we know that there are limits on the sorts of correlations that can be obtained within quantum theory relative to general physical theories, and this then translates into limits on communication tasks.

Finally, the different frameworks that have been developed for describing general physical theories each offer a different perspective on what a quantum theory is. For example, is it a theory of information, or of logic, or of processes and composition? While there may not be any way to choose one of these over the others, having a multiplicity of perspectives is useful for having different ways of thinking and different mathematical tools to attack interesting problems.

Reason 3: The other theories are relevant.

Well, maybe not all of them, but some of them are. The obvious example here is classical theory, which sits with quantum theory in these general frameworks. Whilst classical theory is not nowadays viewed as fundamental physics, it is still important to our understanding of the world. Similarly, we can also talk about different subtheories of quantum theory as living inside this framework, for example, the widely studied stabilizer subtheories can be nicely described, as can various subtheories that appear in the study of quantum resources such as entanglement, coherence, and so on.

Moreover, various other theories are relevant for various other lectures here at the summer school. Paul Skrzyczyk will talk about no-signalling theories, Matt Leifer will talk about epistemically restricted theories, David Schmid will talk about ontological theories, and Konstantinos Meichanetzidis will talk about theories for computational linguistics. All of these can be thought of as living inside the framework that I will introduce. So while they might not be designed to represent fundamental physics, it doesn't mean that they aren't interesting to study!

Reason 4: The relationships between quantum theory and other theories are relevant.

This isn't really an independent point but I thought it worth putting it separately to highlight it. The idea is that once we have a framework which can describe lots of different physical theories, then we can also start to ask about the relationships between them.

In particular, we can start to ask how one theory can be seen to emerge from another theory. For example, how does classical theory emerge from quantum theory as we know that it should to describe our every day experiences? Similarly, can we understand quantum theory as emerging from some deeper theory of nature?

Similarly, we can start to ask whether one theory can be used to provide some sort of explanation for another theory, for example, can we find a classical explanation for the quantum phenomena that we see? If so, is this a natural explanation or does it have some pathological features?

My reason: It's fun.

All of the previous reasons are the sorts of things that you would have to put on a grant application, and I do think they are important reasons, but they aren't really what drove me to study these things in the first place.

For me personally, I just find it intrinsically interesting to contemplate the different ways that the world could be. It's basically just Sci-Fi taken to the extreme!

1.2 Frameworks for general physical theories

Hopefully you now are persuaded that there is some significance to studying general physical theories, but the question then is how can you even get going with such a project? Well, there are broadly two approaches that can be taken.

The first is simply to try to cook up some new theory from scratch, or possibly, to create some new theory by modifying an existing theory such as quantum theory. There is a pretty big literature which takes this approach, but I find it quite unsatisfying, I don't see how going case-by-case will let us tackle some of the big questions that we talked about in the previous section.

The second approach is more principled, we try to find some basic principles that we think any physical theory should satisfy and from these we construct a mathematical framework which allows us to describe and compare many different general theories. One commonly considered principle is the idea that no physical theory should permit sending information faster than the speed of light – essentially, that it should be compatible with relativity theory (in some minimalistic sense). This lets us revisit some of the theories constructed via the first method by seeing if they define valid theories within a given framework. Often, by doing this you will find that they do not, in which case they violate some cherished principle, and, other times, you find that they are within the framework but that they might not be as radical a modification as you might have thought. For example, there are modifications of quantum theory which, when understood within a particular framework, turn out to just be classical theory in disguise!

There are, however, many different frameworks that have been constructed over the years, each one taking different perspectives on what should be taken as primitive to all physical theories. To give a non-exhaustive list, we have:

- non-classical logics
- generalised probabilistic theories
- effectus theories
- operational probabilistic theories
- process theories
- causal-inferential theories
- ...

There a pros and cons to each of these formalisms, it really depends on what aspect of quantum theory you are most interested in studying. They are also by no means independent of one another (e.g., any modification to logic leads to a modification of probability theory too) it is therefore also very interesting to study the relationship between these different approaches. This, however, is well beyond the scope of what I want to get into here.

There are a couple of important things to comment on here though. The first, is that the idea of studying general frameworks for physical theories is not at all new, indeed the origins can be traced back to very shortly after the initial formalisation of quantum theory. The second is that I don't believe that *any* of these frameworks are complete, in the sense of being able to answer all of the questions I raised in the introduction. In particular, if we want to find genuine beyond-quantum theories, such as may be necessary to unify with general relativity, then we may well need a new or further developed framework. Nonetheless, all of these frameworks have proved useful and a great deal of progress has been made on the questions that I posed earlier.

Finally, it is worth commenting on another approach to studying beyond-quantum physics, which is to focus attention on a particular phenomena such as nonlocality or steering. One can then consider the scope of possibilities for that particular phenomena, rather than worrying about developing an entire physical theory. The advantage of this approach is that we end up with a much simpler object to study, which helps to make concrete progress on understanding the particular phenomena. The disadvantage is that without a candidate physical theory underpinning the full scope of possibilities for the phenomena, then we can't be sure that it is worth contemplating this full scope in the first place! Luckily, for certain phenomena it can be shown that we can get the best of both worlds, that is, one can find a particular physical theory that can realise the full scope of possibilities for the phenomena. The classic example of this is the relationship between the study of Bell nonlocality and the generalised probabilistic theory known as Boxworld.

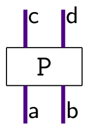
1.3 Process theories

In these lectures I'm going to largely focus on the framework of process theories[2]. The basic conceptual underpinning of the framework is that one can express (general) physical theories in terms of the physical processes that can happen in the (hypothetical) world, together with how processes can be composed together to make other processes.

This idea can be formalised in the mathematics of symmetric monoidal categories. This is great because mathematicians have been studying these things since the 60s, but is not so great because physicists don't know about these at all... Well, that isn't quite true, the truth is that physicists do know about these things, they just don't know that they know!

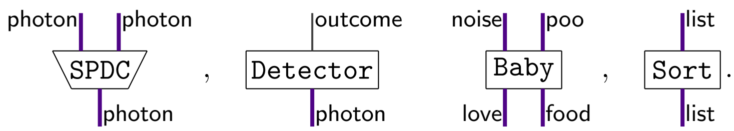
Physicists are used to drawing little sketches of protocols or experiments to keep track of what is going on, i.e., to keep track of the information flow within the protocol or the movement of physical systems between bits of lab apparatus in the experiment. What process theories allow us to do, is to take these sketches seriously. That is, rather than just being an informal picture to supplement the mathematics, these pictures become a formal diagram which is a mathematical object that we can directly use to prove things. From this perspective, the mathematics of symmetric monoidal categories is then just a way to make sure that the diagrams behave in the way that we already intuitively know they should.

In these lectures I'm therefore going to stick with the diagrammatic way of describing things, if you want to see how this connects to symmetric monoidal categories then there are plenty of good books that deal with the subject. The fundamental building blocks of these diagrams are *processes*, we denote these by labelled boxes



, (1.1)

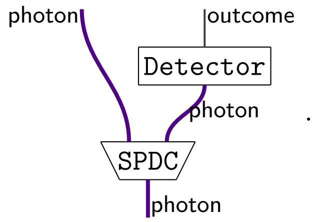
which have input *systems* A and B at the bottom and output systems C and D at the top. The kinds of processes that we consider come from a broad range of fields. They could be physical processes (what we really care about in this course) but they could also be chemical processes, or biological processes, or computational processes, or mathematical processes, and much more. To give a few illustrative examples all of the following are processes:



(1.2)

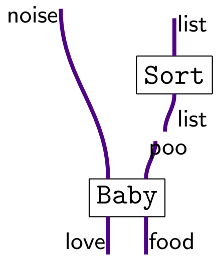
[2]Although, the special cases that I will consider next lecture will end up giving the framework of operational probabilistic theories which is itself closely tied to the framework of generalised probabilistic theories, but more on that later.

Now, the key thing that gives mathematical content to these processes is how the form *diagrams*. That is, given a bunch of processes we can wire them together to describe a new process, for example:



(1.3)

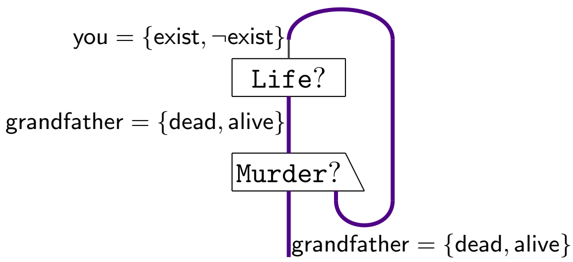
Not every wiring together of processes makes sense, in particular, it must be the case that we match system types when we do a wiring. That is, a diagram such as:



(1.4)

doesn't really make sense at all!

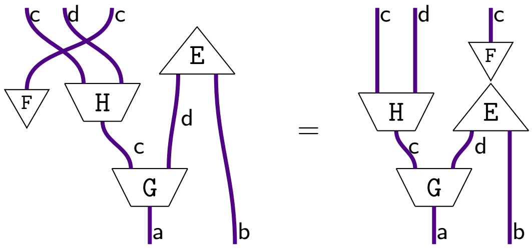
The kinds of processes that we want to consider here are physical processes, where we think of the inputs as physical systems that can potentially have a causal influence over the output physical systems. In this case then it also makes sense to demand that the wiring should be *acyclic*. This means that we can imagine that we have an arrow of time on our diagrams, and that we don't run into time-travel paradoxes! That is,



(1.5)

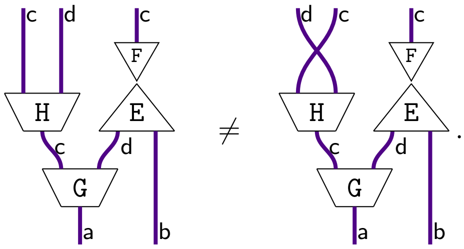
is not an allowed diagram in a process theory. It is interesting to study process theories which do not rule out cyclic wirings and so study things like this paradox, but it is beyond the scope of this course.

The sorts of diagrams that we draw contain some superfluous information, we aren't really interested in precisely how the processes are positioned on the page, only on how they are wired to one another. We therefore say that two diagrams are equal if they describe the same wiring of processes. For example



(1.6)

as both diagrams have all of the processes connected up in the same way, and, moreover, the way the systems connect "to the boundary" is the same. That is, in general

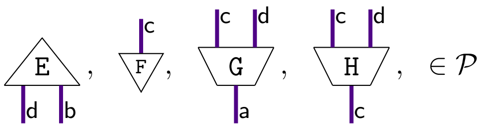


(1.7)

With these ideas in place I can now formally define what a process theory is.

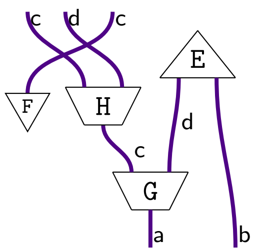
Definition 1.3.1 (Process theories)**.** *A process theory, $\mathcal{P}$, is a collection of processes which is closed under forming diagrams.*

This means that if we have some



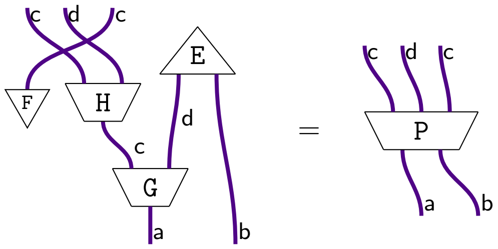
(1.8)

and construct a diagram out of them such as



(1.9)

then this diagram must be equal to some process $\mathtt{P} \in \mathcal{P}$

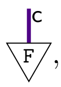


(1.10)

To give a potentially more familiar mathematical analogy, part of the definition of a group is that when we compose two group elements we get another group element.

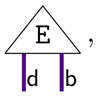
This gives us the basic definition of a process theory, it is useful to introduce a bunch more jargon for particular special cases, some of which we have seen already in some of the diagrams above.

In particular, we have *states*, which are processes, such as



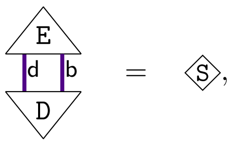
(1.11)

that have no inputs; we have *effects*, which are processes, such as



(1.12)

that have no outputs. We can wire states and effects together to give processes with neither inputs nor outputs, such as

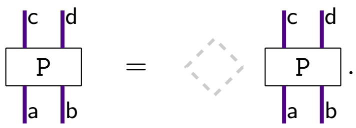


(1.13)

which are known as *scalars*. Every process theory has at least one scalar, namely the *empty diagram*

(1.14)

which acts as the 'number 1' in that it leaves every process invariant:

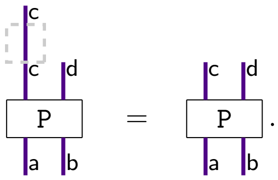


(1.15)

Every system, a has at least one process, namely the *identity process*, $\mathbb{1}_{\mathsf{a}}$, which diagrammatically is represented as an empty box

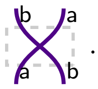
(1.16)

this acts as an identity in that it leaves every process invariant, for example



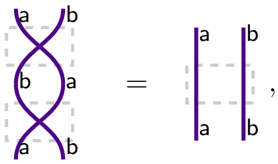
(1.17)

Every pair of systems, a and b, have a *swap process*, $\mathbb{S}_{a,b}$, which is diagrammatically just a swapping of the wires:



(1.18)

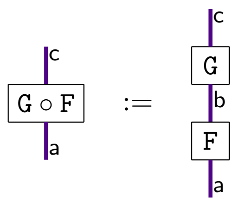
for which our rule 'that only connectivity matters' gives a bunch of interesting conditions, for example:



(1.19)

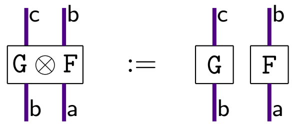
which tells us that the wires in process theories do not get tangled up[3].

Finally, we have two kinds of wirings that we highlight, sequential



(1.20)

and parallel

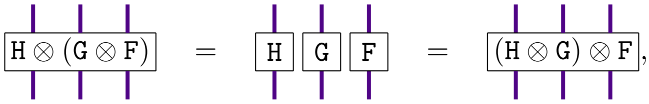


(1.21)

The diagrams automatically give us a bunch of nontrivial mathematical structure for these two special cases, for example, that sequential and parallel composition are

[3]There are generalisations which relax this condition but we won't go into them here.

associative, e.g.,



(1.22)

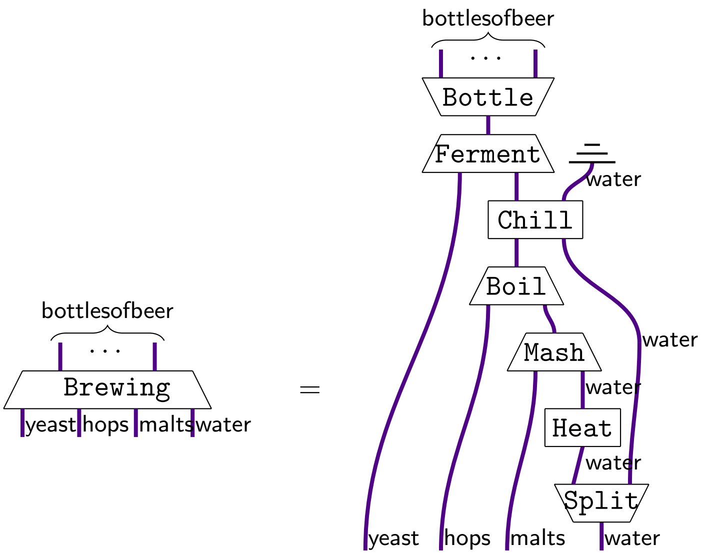
and that (as he mentioned earlier) sequential composition has the identity transformation as a unit.

One of the reasons why it is worth considering all of these special cases is that it turns out that we can use them as primitive building blocks for our diagrams.

Theorem 1.3.2. *Any (acyclic) diagram can be nonuniquely constructed by composing using the above speical cases.*

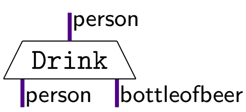
In fact, this theorem is essentially what gets us to an equivalent definition of process theories in terms of the symmetric monoidal categories that I mentioned earlier.

Now to try to give a bit of a nontrivial example, lets think about the process theory of brewing. This has a bunch of different processes and so we can, for example, write that:



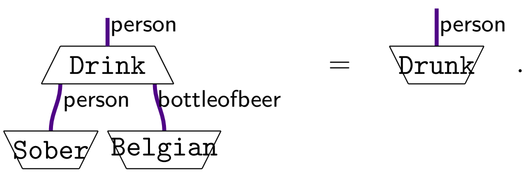
(1.23)

and we could go further and compose this with other processes such as



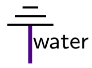
(1.24)

which satisfies the equation



(1.25)

There's one process in the diagram on the right hand side of Eq. (1.23) that we are yet to discuss, namely, the process



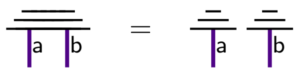
(1.26)

which, in this case, represents pouring the hot water coming from the chiller down the sink. More generally, we use this symbol to represent any sort of discarding process.

1.4 Discarding

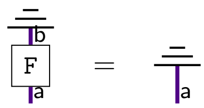
In many process theories there is a way to discard any process. For example, in the context of brewing we can think of this as 'throwing something in the bin' or 'pouring something down the sink', but in other contexts it may not be a physical process but just simply ignoring some system due to lack of interest or not having access to it.

Regardless of the context they show up in, these processes satisfy some obvious but very important equations. Firstly



(1.27)

which tells us that if we discard two systems together, that this is the same as discarding the two systems individually. Secondly



(1.28)

which tells us that if we discard all of the outputs of some process then we may as well have just discarded the input.

Hidden in this last condition is a subtle but important point about process theories. We might try to think of situations in which discarding the output is not the same as discarding the input, but what is happening in such cases is always that there is some extra output which we are only treating implicitly. If we instead treat them as explicit outputs then we again will find that this equations holds. Generally a good rule of thumb for working with process theories is that you need to explicitly denote *all* of the relevant systems, if you don't then you're likely to make mistakes and to lose important diagrammatic rules such as the above equation.

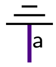
If a process theory has discarding maps satisfying Eq. (1.27) then any process $\mathtt{F} \in \mathcal{P}$ satisfying Eq. (1.28) is said to be 'causal' or 'discard-preserving'.

Definition 1.4.1 (Causal process theories)**.** *A causal process theory is a process theory in which every process is discard-preserving.*

There are a couple of important results that hold for arbitrary causal process theories that are worth commenting on. First, the empty diagram,

, (1.29)

is the unique scalar for the theory. Secondly for each system a, the discarding effect,

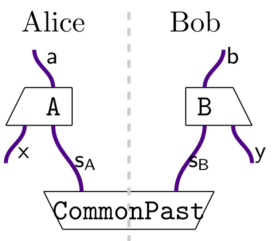


, (1.30)

is the unique effect for the system.

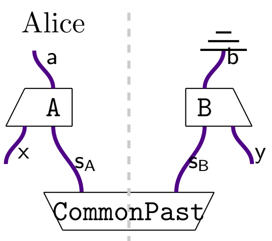
With these results in place then we can start to see why these process theories are called causal. Namely, that they have a sort of 'compatibility' with space-time. That is, if we imagine that these diagrams are actually living inside some space-time such that the wires are describing world-lines of some physical systems, then any causal process theory will obey the no-signalling conditions implied by relativity theory!

Let's illustrate this with a simple example. Consider two parties, Alice and Bob, who are space-like separated. They have some shared past in the intersection of their backwards lightcones, which we represent by some bipartite state, and then they can each do some local process of their part of this bipartite system. The global picture therefore looks like



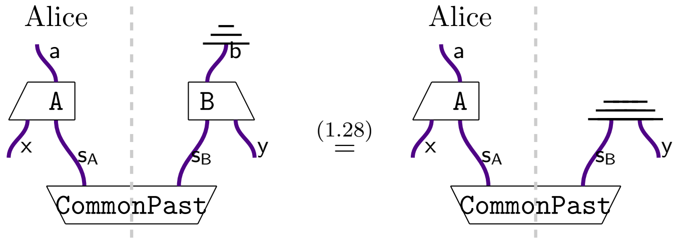
. (1.31)

However, let's now consider what things look like from the perspective of Alice. Alice doesn't have access to Bob's output, and so treats it as if it's been discarded

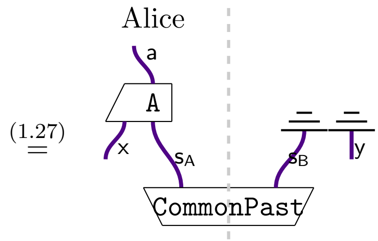


. (1.32)

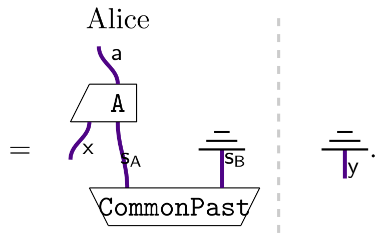
At which point if we are working with a causal process theory then we can use show the following:



(1.33)



(1.34)



(1.35)

This means that from Alice's perspective whatever Bob does is entirely disconnected from what she does, and, hence, he has no way to send a signal to her. That is, there can be no superluminal signalling from Bob to Alice.

Of course, the situation is symmetric, so if we look at things from Bob's perspective then we will see that there is no possibility of Alice signalling to him either.

By a very similar argument we can also see that if Alice is in the causal future of Bob, then it is possible for him to signal to her but not vice versa. A more general argument, but based on the same ideas, is put forward in Ref. [1].

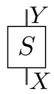
The basic message that I want you to take away from this is the following: **complex physical ideas follow from extremely simple process theoretic concepts**.

1.5 Example: stochastic dynamics

Any good physical theory should be able to describe the classical world that we see around us every day. In the next lecture we will see how this idea can be used to underpin a formalism of generalised probabilistic theories by demanding constraints on how classical systems interact with non-classical systems, and in the third lecture we will see how we can take these classical systems to be emergent rather than fundamental. Before we get to any of this, however, we need to first describe the process theory which describes classical physics.

For various reasons that I won't get into here but other lecturers are sure to do justice to, it is natural to consider stochastic rather than deterministic classical physics. Whether this stochasticity is fundamental or just capturing our ignorance as to what is going on is an important question but not one that we're going to get into right now. The reason for this is that it's not really pertinent to the process theory that you get at the end of the day.

We call this process theory **Stoch**. The systems are given by finite[4] sets which represent the set of ways that the classical system can be. The processes are then stochastic maps between these sets. If a process has a single input and output, such as



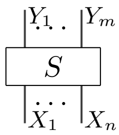
, (1.36)

then it is described by a stochastic map:

Definition 1.5.1 (Single input and output stochastic map)**.** *A stochastic map S from the set X to the set Y can be described by a function $S : X \times Y \to [0,1]$ :: $(x, y) \mapsto S(y|x)$ such that $\sum_{y\in Y} S(y|x) = 1$ for all $x \in X$.*

That is $S(y|x)$ tells us the probability that the classical system labelled by Y will take value y given that the classical system labelled by X took value x. In other words, it is a conditional probability distribution.

If we have multiple inputs and outputs, such as

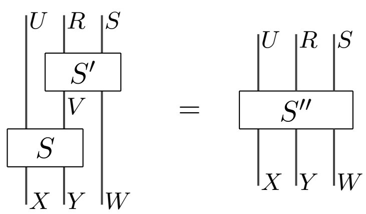


, (1.37)

then things are somewhat more complicated but the basic idea is the same:

Definition 1.5.2 (Stochastic map)**.** *A stocahstic map S from the sets $\{X_i\}_{i=1}^n$ to the sets $\{Y_j\}_{j=1}^m$ can be described by a function $S : \prod_{i=1}^n X_i \times \prod_{j=1}^m Y_j \to [0,1]$ :: $(x_1, ..., x_n, y_1, ..., y_m) \mapsto S(y_1, ..., y_m|x_1, ..., x_n)$ such that $\sum_{y_i\in Y_i} S(y_1, ..., y_m|x_1, ..., x_n) = 1$ for all $x_i \in X_i$.*

Next lets look at how these compose, rather than looking at sequential and parallel composition independently we'll define the generic case:



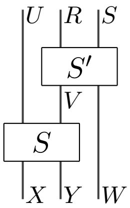
(1.38)

[4]This is purely for technical convenience rather than any fundamental reason.

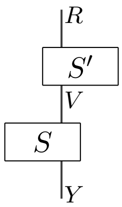
where S'' is another stochastic map defined by

$$S'' : (X \times Y \times W) \times (U \times R \times S) \to [0,1] :: (x,y,w,u,r,s) \mapsto \sum_{v \in V} s(u,v|x,y)s'(r,s|v,w). \tag{1.39}$$

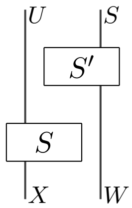
This can be seen to include both sequential and parallel by suitably removing systems, that is,



$\mapsto$

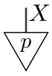


and



respectively. (1.40)

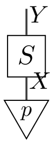
This is all much clearer (to me anyway) once we focus on some special cases. For example, suppose we have a state with a single output,



$$: \; X \to [0,1] \;::\; x \mapsto p(x) \quad \text{s.t.} \sum_{x \in X} p(x) = 1. \tag{1.41}$$

This is simply a probability distribution over the set X.

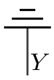
If we then compose this with a stochastic map with a single input and output then we obtain



$$: \; Y \to [0,1] \;::\; y \mapsto \sum_{x \in X} S(y|x)P(x) \tag{1.42}$$

which is using the standard belief propagation formula to use a stochastic map from X to Y to convert a probability distribution over X into a probability distribution over Y.

Processes without outputs are a bit weirder, they correspond to the constant unit function, for example,



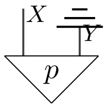
$$: \; Y \to [0,1] \;::\; y \to 1 \tag{1.43}$$

hence, we see that these are unique for each system and so we have that:

Proposition 1.5.3. Stoch *is a causal process theory.*

Proof sketch. The stochasticity condition for a stochastic process S from X to Y, namely $\sum y \in YS(y|x) = 1$ for all $x \in X$, ensures that all S are discard-preserving, that is, they satisfy Eq. (1.28). □

It is clear what this is useful for when we consider discarding one part of a bipartite state:



$$: X \to [0,1] \;::\; x \to \sum_{y\in Y} p(x,y) \tag{1.44}$$

that is, discarding in **Stoch** gives us a way to compute marginal probability distributions!

Next we will consider the convex structure of processes in **Stoch**.

Proposition 1.5.4. *The set of stochastic maps from X to Y forms a convex set.*

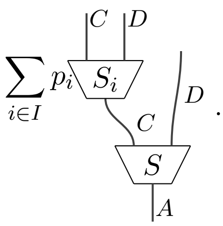
Proof. In particular, given $S_i : X \times Y \to [0,1]$ we can define $\sum_i p_i S_i :: (x,y) \mapsto \sum_i p_i S_i(y|x)$ it is easy to check that this is indeed a valid stochastic map.

Firstly, p_i and $S_i(y|x)$ are all greater than zero so $\sum_i p_i S_i(y|x) \geq 0$. Moreover, $S_i(y|x)$ are all less than one, hence $\sum_i p_i S_i(y|x) \leq \sum_i p_i = 1$. This means that $\sum_i p_i S_i : X \times Y \to [0,1]$ as we require.

Secondly, $\sum_y \sum_i p_i S_i(y|x) = \sum_i p_i \sum_y S_i(y|x) = \sum_i p_i 1 = 1$ for all $x \in X$ as we require. □

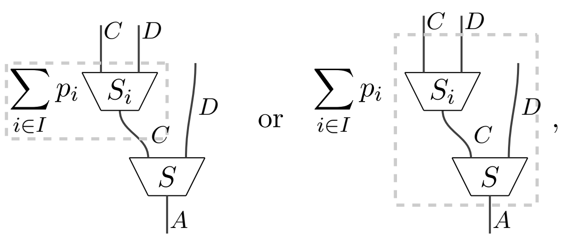
This geometry is particularly simple in the case of states. Let us define the delta-function distributions as $\delta_x : X \to [0,1] :: x' \mapsto \delta_{x,x'}$, then any other probability distribution p can be uniquely written as a convex combination of these δ_x, that is, $p = \sum_x p(x)\delta_x$. In other words, the set of probability distributions over the set X form a *simplex* with vertices labelled by $x \in X$.

It is tempting to incorporate these convex sums into our diagrammatic formalism, for example, to write things like:



(1.45)

However, apriori we do not know if this is consistent as there are different ways to understand this diagram, that is as:



(1.46)

where in the first case we first take the convex combinations and then compose the processes, whereas in the second case we first compose the processes and then take the convex combinations.

Conveniently for us these turn out to be equivalent, and so we can use these convex sums as we would like!

Proposition 1.5.5. *Composition in* **Stoch** *is convex linear.*

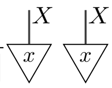
Proof sketch. The crux of this is that

$$\sum_{v\in V}\left(\sum_i p_i s_i(u,v|x,y)\right) s'(r,s|v,w) = \sum_i p_i \left(\sum_{v\in V} s_i(u,v|x,y)s'(r,s|v,w)\right).$$

□

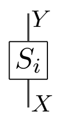
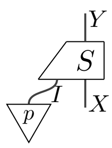
This means that our convex sums are free to float around diagrams, which is great because it means that our idea that 'only the wirings in a diagram matter' still holds.

Nonetheless, there is something a bit unsatisfying about using sums in this way, as they add a connection without having it as an explicit part of the diagram, for example, we can write the perfectly correlated bipartite distribution as:

$$\sum_{x\in X}\frac{1}{|X|}$$  , (1.47)

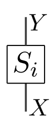
which leaves us without any explicit diagrammatic connection between the two systems even though they should not be viewed as being independent of one another. Luckily in **Stoch** there is a nice way to get around this, we can always turn the convex coefficients p_i into a probability distribution and work with that instead.

Theorem 1.5.6. *For any convex combination of stochastic maps $\sum_{i\in I} p_i S_i$ where S_i have input X and output Y, we can always find a stochastic map S with inputs I and X and output Y such that:*

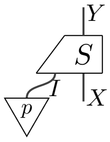
$$\sum_{i\in I} p_i$$  =  . (1.48)

Proof. We can simply define p by $p : i \to [0,1] :: i \to p_i$ and S by $S : (I\times X)\times Y \to [0,1] :: (i,x,y) \mapsto S_i(y|x)$. It can then easily be verified that this is indeed a valid stochastic map as $\sum_{y\in Y} S_i(y|x) = 1$ for all $i \in I$ and $x \in X$.

The definition of convex combinations means that the LHS is described by a stochastic map

$$\sum_{i\in I} p_i$$  $: X\times Y \to [0,1] :: (x,y) \mapsto \sum_{i\in I} p_i S_i(y|x)$ (1.49)

whilst the definition of composition mean that the RHS is described by a stocahstic map

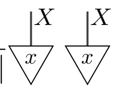


$$: X \times Y \to [0,1] :: (x,y) \mapsto \sum_{i\in I} p(i)S(y|i,x) \tag{1.50}$$

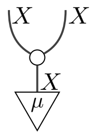
which we can then show, by the definition we gave of p and S that $\sum_{i\in I} p(i)S(y|i,x) = \sum_{i\in I} p_i S_i(y|x)$. This means it is the same stochastic map as the LHS which completes the proof. □

If we apply this to the case of the perfectly correlated distribution then we find that:

$$\sum_{x\in X} \frac{1}{|X|}$$



$$=$$



$$, \tag{1.51}$$

where the white dot is the 'copy dot' and μ is the uniform distribution over X. On the RHS we therefore can see that there is an explicit diagrammatic connection between the two versions of X, and so correlations are now manifest in the diagrammatic language.

1.6 Summary

In this lecture I hope to have persuaded you of the interest in studying physical theories other than quantum theory, and given some idea of what this is useful for. I've introduced the basic diagrammatic formalism of process theories, and in particular, shown how we can define a particular class of process theories – causal process theories – which have the nice property that signalling is necessarily mediated by systems, and so they are (in a minimal sense) compatible with relativity theory. Finally, I introduced a key example of a process theory **Stoch** which represents classical stochastic dynamics. This will be a crucial process theory in going forwards to general probabilistic theories in the next lecture. Essentially, we will see that general probabilistic theories are those that are compatible with **Stoch** in the 'obvious' way.

Lecture 2

General probabilistic theories

As I have mentioned already, the scope of process theories is extremely broad covering examples from all natural sciences as well as mathematics, computer science, and so on. There are therefore many important questions which we can't hope to answer at this level of generality. For example, suppose we are interested in the information processing capabilities of some candidate physical theory, maybe we want to see whether it permits better secure key distribution than quantum theory, then in many process theories this is not even a meaningful question.

In this lecture we will therefore narrow our scope to a smaller class of process theories for which such questions are well posed. This (for all practical purposes) leads us to the framework of Operational Probabilistic Theories (OPTs) which can be thought of as a variant of the Generalised Probabilistic Theory (GPT) framework. These frameworks have been used to study various features of quantum theory in a generalised setting, such as: cryptography, computation, thermodynamics, and much more. One of the most exciting things to come out of this framework are various *reconstructions* of quantum theory, that is, finding small sets of physical principles which can be articulated within the framework and which together single out quantum theory as the only GPT satisfying them all. This provides insight into 'why' quantum theory – if nature was anything else then one of these physical principles would necessarily be violated!

I will now introduce the basics of this formalism in three steps. In the first step we will consider a class of process theories that have a *classical interface*, in the second step we will throw away irrelevant information about processes in order to obtain a GPT, finally, in the third step we will obtain a representation of the GPT in terms of real vector spaces.

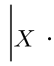
2.1 Step 1: the classical interface

Suppose we have some process theory $\mathcal{P}$ which we want to represent a candidate theory to explain our world. Then we know that this must contain the classical world, described by **Stoch** as a subtheory. That is, we have certain systems in the world that are classical and evolve under stochastic dynamics. This is not to say that we necessarily have systems that are fundamentally classical, but that at least on an effective level these must be around[1].

Formally, we are demanding that **Stoch** $\subseteq \mathcal{P}$. It will be useful to diagrammatically distinguish the systems in this subtheory from the more general systems, we will do so by drawing general systems as:

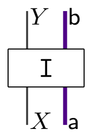
(2.1)

and those in the classical subtheory as:



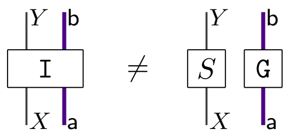
(2.2)

Now, it is not enough to simply have classical systems around in our theory, we want to be able to use these classical systems as an interface with the full theory. That is, we think about the classical systems as our probes that we can manipulate which let us learn something about the wider world. For example, these classical systems represent the settings that we have on bits of experimental apparatus, or the pointer positions that indicate different outcomes of a measurement. This means that there must be some minimal kinds of interactions that we demand between our classical and general systems. That is, there must be certain processes



(2.3)

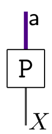
such that



(2.4)

for any S and `G`, as otherwise, the interaction between the classical and general world would be trivial!

Certain types of interactions correspond to some familiar sorts of processes. For example:

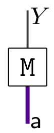


(2.5)

is a controlled preparation procedure, that is, we have some setting X which controls

[1] Indeed, in the last lecture we will see how to understand these classical systems as emergent from more fundamental systems.

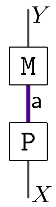
which state of a is prepared. Moreover, we have:



(2.6)

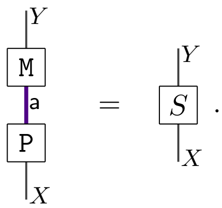
which is a destructive measurement procedure, that is, we have some system a which is measured generating some probability distribution over the outcomes Y.

We will return to more adding more stringent requirements to these interactions shortly, but first let us consider what happens if we construct a diagram with only classical inputs and outputs. The fact that it only has classical inputs and outputs still leaves open the possibility that there are some non-classical systems in the interior of the diagram, for example, if we compose the controlled preparation with the destructive measurement we obtain the diagram:



(2.7)

we demand that this is simply a stochastic map from X to Y, that is, that there exists a stochastic map S such that:



(2.8)

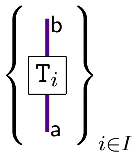
This is because if we perform such an experiment then at the end of the day we will collect data about which outcomes we observe for each settings, and given sufficient runs of the experiment this will converge to a conditional distribution $S(y|x)$ which defines the stochastic map S. The same holds for more complicated diagrams so long as the inputs and outputs are all classical.

Formally we can succinctly state this as:

Principle 1. **Stoch** *is a* full *subtheory of* $\mathcal{P}$*.*

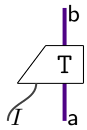
Where the term 'full' simply means that every process with inputs and outputs from **Stoch** is necessarily a process from **Stoch**.

Next, it is natural to assume that if we have some set of processes



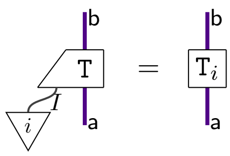
(2.9)

that we can implement, then we should be able to choose which one to implement with a control variable, I. Formally this means that there must exist a classically controlled transformation



(2.10)

where the classical system I is now a setting variable controlling which of the $\mathtt{T}_i$ are implemented, i.e., satisfying



(2.11)

for all $i \in I$.

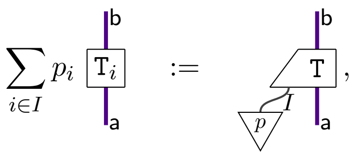
This brings us to our second principle:

Principle 2. *All classically controlled transformations exist in $\mathcal{P}$.*

Note that this holds also just within **Stoch**, indeed, this is what we used at the end of the last lecture to turn sums into explicit systems. That is, the stochastic map S in Eq. (1.50) is the classically controlled process for the stochastic maps $\{S_i\}_{i\in I}$ in Eq. (1.49).

A natural question to ask is, we used controlled stochastic maps to represent convex combinations of stochastic maps, can we extend this to general transformations? The answer turns out to be yes! That is, we can define:

Definition 2.1.1 (Convex combinations)**.** *We can define a convex combination of general processes as:*



(2.12)

where T is the classically controlled process for the $\{T_i\}_{i\in I}$ and $p : I \to [0,1] :: i \mapsto p_i$.

Showing that this is a consistent definition is a bit beyond the scope of these lectures, but hopefully the intuition is clear: a convex mixture of the processes $\mathtt{T}_i$ can be thought of as having some setting which toggles between the different processes which is chosen according to some probability distribution.

Similarly, we can then show that composition with other processes is convex linear with respect to this notion of convex combination that we have defined. This is essentially because composing $\mathtt{T}$ with some other process $\mathtt{T}'$ gives a classically controlled process for the composite of the $\mathtt{T}_i$'s with the $\mathtt{T}'$.

We can use this notion of convex combination to argue that the set of processes from a to b forms a convex set, but, unlike in the case of **Stoch** it isn't really clear at the stage what this convex set actually looks like. This is resolved in the next step which will bring us to the standard geometric picture of GPTs.

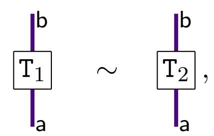
2.2 Step 2: removing operationally irrelevant information

Within the diagrams that we are drawing in $\mathcal{P}$ there is (typically) a great deal of information which is irrelevant to making predictions about outcomes of experiments. Such superfluous information is known as the *context* of the process[2]. The idea of a generalised probabilistic theory is that it should contain no superfluous information, that is, it is a bare bones description which contains only what is strictly necessary for making experimental predictions.

This idea should be fairly familiar from quantum information theory, we can work with ensembles of pure states $\{(p_i, |\psi_i\rangle)\}_{i\in I}$, but we know that there are multiple ensembles of pure states which lead to the same outcome statistics for all measurements, and hence we work with density matrices $\rho = \sum_{i\in I} p_i |\psi_i\rangle\langle\psi_i|$ instead. Density matrices can then be seen as a minimal description capturing only the information that is strictly necessary to make predictions about experiments.

We can capture this idea much more generally by introducing the notion of *operationally equivalent* processes:

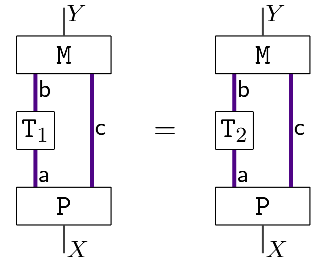
Definition 2.2.1 (Operational equivalence)**.** *Two processes* $\mathtt{T}_1$ *and* $\mathtt{T}_2$ *are said to be operationally equivalent, denoted*



(2.13)

[2] This is (formally) the same notion of context that David will talk about in depth in his lectures.

if and only if:



(2.14)

for all finite sets X, Y, general systems c*, controlled preparations* P *and measurements* M*.*

This means that two processes are operationally equivalent if there is no possible experiment that could distinguish them.

As a sanity check, it is quite straightforward to see that if we focus on processes with classical inputs and outputs, then two stochastic maps are operationally equivalent if and only if they are equal.

An important property of this equivalence relation, is that it is compatible with the various structures that we care about. In other words:

Theorem 2.2.2. *Operational equivalence, $\sim$, is a* congruence *relation.*

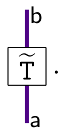
Proof sketch. There are two key things to prove here. Firstly, that $\sim$ is preserved by composition, and secondly, that $\sim$ is preserved by convex combinations.

TO DO

□

The key benefit of this, is that it allows us to *quotient* $\mathcal{P}$ with respect to $\sim$, and that the mathematical object that we get out of this quotienting will inherit a bunch of nice properties from $\mathcal{P}$. In particular, it will be a process theory with **Stoch** as a full subtheory that allows us to take convex combinations of processes.

Let's try to make all of that a bit more concrete. To start with, what do I mean by quotienting? Well, we're going to try to define a new process theory $\mathcal{P}/\sim$ in which the processes correspond to equivalence classes of processes in $\mathcal{P}$. Let us denote the equivalence class of processes that are operationally equivalent to T as $\widetilde{\texttt{T}}$, this means that the processes in $\mathcal{P}/\sim$ can be diagrammatically denoted by:

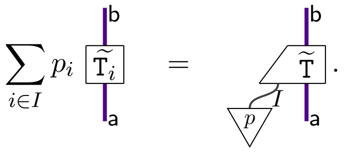


(2.15)

Note that as for stochastic processes we have that equality and operational equivalence are the same, then quotienting leaves the subtheory **Stoch** invariant, and so we will write S rather than $\widetilde{S}$ for the processes in this subtheory.

Now, having defined the processes in our quotiented theory, we need to define how they compose. Well one way to do this is simply to pick some representative element from each equivalence class, compose those within $\mathcal{P}$ and then to find the equivalence class of the composite process. However, in order for this to be well defined it had better be the case that this is independent of the choice of representative elements. Luckily for us this is indeed the case, this is part of what we proved in order to demonstrate that $\sim$ was a congruence relation!

We also want to understand the convex structure of the quotiented theory. In particular, we want to show that we again have the two equivalent perspectives, we can either talk directly about convex combinations of equivalence classes, or, we can add a classical control system and a probability distribution over the control variable, that is



(2.16)

An important feature that be expressed, is that in GPTs it is assumed that:

Principle 3. *All effects for a given system are operationally equivalent, that is, ways to discard a system only vary in their contexts.*

This immediately implies that:

Proposition 2.2.3. *$\mathcal{P}/\sim$ is a* causal *process theory.*

This description of generalised probabilistic theories in terms of equivalence classes, is a mathematically elegant way to go about things, but it isn't really clear how to compute anything.

2.3 Step 3: linear representation

In order to get to a formalism that is more amenable to actually doing calculations, we want to find a concrete representation of our process theory in terms of real vectors. This is generically possible, however, due to lack of time we're going to focus only on the states and measurements for a single system, that is, on *preapre-measure scenarios*. Moreover, we will make the simplifying assumption, that all of the convex sets that we abstractly obtained by quotienting are *finite dimensional*. This can be motivated by the following principle:

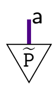
Principle 4. *A finite set of experiments suffices for characterising operational equivalence, i.e., for process tomography.*

This means that if two processes are not operationally equivalent to one another, then this will be revealed by the statistics produced by a particular finite set of experiments. That is, we don't need to test all possible experiments, as we have in the definition of operational equivalence).

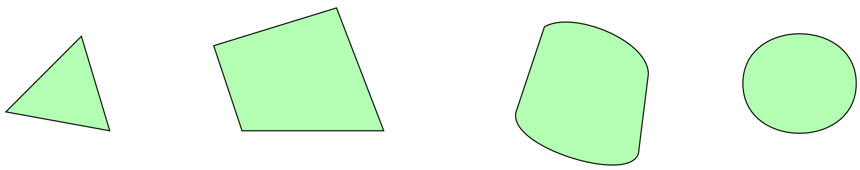
This kind of finite tomography works well enough for finite dimensional quantum systems but fails when we get into the realms of infinite dimensional systems. This principle therefore tells us that we are focusing on generalised theories that are finite in a similar sense to the study of finite dimensional quantum theory. Whether or not we believe that finiteness is appropriate for fundamental physics, it is undoubtedly true that we have managed to understand a great deal about quantum theory by focusing on the finite dimensional case. It therefore seems sensible to start off considering finite dimensional generalised probabilistic theories before we start worrying about the technicalities of infinite dimensional systems.

Let us start by considering the geometry of states. We have already seen that for **Stoch** that the states for X form a simplex where the vertices correspond to delta-function distributions, and so can be labeled by $x \in X$. There is a key feature of the simplex which is generic to all GPTs which is that this convex set is *bounded*. Let us try to now to give some intuition for this property.

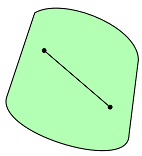
For some GPT system a let us denote the convex set of states as Ω_{a}, that is

$$\Omega_{\mathsf{a}} := \left\{ \right. $$  $$\left. \right\}, \tag{2.17}$$

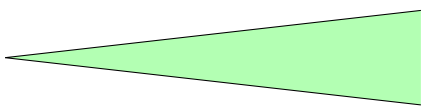
and we will assume (so that I can actually draw it on the page) that this convex set is 2D, for example:

 , (2.18)

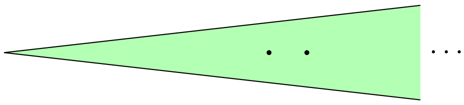
where the first of these is a three-vertex simplex so corresponds to a state space in **Stoch**. Convexity of these sets means that if I take any two points in these sets, and draw a line-segment between them, then this is contained within the set, for example,

 . (2.19)

Now, convexity alone does not rule out the possibility of sets such as

 $\cdots$. (2.20)

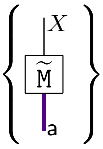
This is an unbounded set as there is a direction that 'goes to infinity'. However, such convex sets are ruled out as they contain points which cannot be distinguished by any measurement, and so should have been identified when we quotiented, for example, in the above picture the two points:



(2.21)

cannot be distinguished by any measurement.

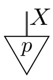
We know that the set of measurements with a particular outcome set X, form a convex set,

$\mathcal{M}_X :=$  , (2.22)

and we could characterise the geometry of each of these sets. However, we will instead focus on a closely related geometry which is the geometry of *effects*. The reason for this is that effects are often taken as part of the fundamental structure of a GPT so it is important to gain some familiarity with them.

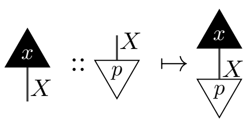
Essentially, if we want to talk about particular outcomes of measurements rather than the measurement as a whole, then we start talking about effects.

Lets think just about **Stoch** for the moment. We have considered states as probability distributions



$$: \ X \to [0,1] \ :: \ x \mapsto p(x) \quad \text{s.t.} \sum_{x \in X} p(x) = 1. \qquad (2.23)$$

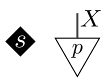
However, suppose I am interested in a particular $x \in X$, and I want to have some diagrammatic representation of $p(x)$, then, as things stand this is not possible. An easy way to see this is that typically $p(x) \neq 1$ and so if we have $p(x)$ as a scalar in our theory then we are necessarily no longer in a causal theory. Nonetheless, let us introduce a diagrammatic element that does just this:



$$= p(x), \qquad (2.24)$$

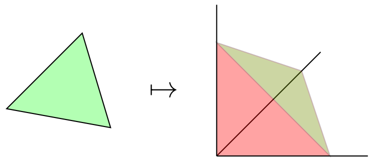
which we denote as an effect with white text on a black background to highlight the fact that this is not a discard-preserving process.

The issue with adding in this new effect, is that we still want to have a process theory after adding this effect, and so we need to worry about all of the processes that we can now construct using this new effect. Well, the first thing to observe is that this gives new states in our theory. That is, noting that $p(x)$ can be an arbitrary value s in $[0,1]$, then we can obtain any state:



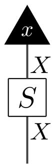
. (2.25)

Geometrically what this gives us is an embedding of the simplex of probability distributions into a real vector space of one dimension higher, that is, into the vector space $\mathbb{R}^X$,


(2.26)

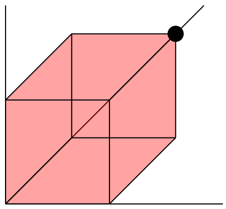
and the set of states now correspond to arbitrary vectors between the zero vector and the simplex of probability distributions. These can be thought of as *subnormalised* distributions, i.e., those for which $\sum_x p(x) \leq 1$.

Similarly, we can create a bunch more effects by pre-composing these new effects with stochastic maps:



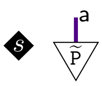
. (2.27)

Geometrically these can be thought of as dual-vectors in $\mathbb{R}^{X^*}$, that is, as linear functionals on $\mathbb{R}^X$. Using the Riesz representation theorem, however, we can always represent linear functionals on $\mathbb{R}^X$ as vectors in $\mathbb{R}^X$ and compute the probabilities via the dot-product. Pictorially, we can therefore draw these as:


, (2.28)

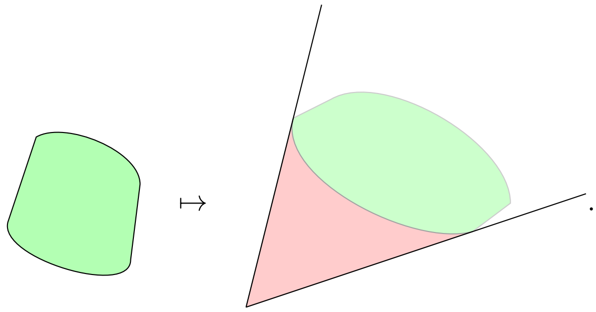
where the black dot in the upper right corresponds to the discarding effect. These are often known as *response functions* for X.

Adding in these extra processes also then feeds through to the GPTs. For example we now obtain subnormalised states such as:



, (2.29)

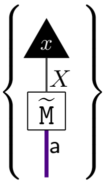
which geometrically again gives us an embedding of Ω_{a} into a vector space, which we will denote as V_{a}, of one dimension higher and then taking all convex combinations with the zero-vector, e.g.:



(2.30)

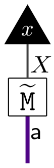
More interestingly, we also get a bunch more effects for the GPT, defined by

$\mathcal{E}_{\mathsf{a}} :=$



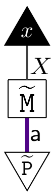
. (2.31)

These are very useful mathematical objects as they allow us to compute the probability of particular outcomes of the measurements. For example, if we compose the effect



(2.32)

with some state $\widetilde{\mathtt{P}}$, then this will tell us the probability that we obtain outcome $x \in X$ when measurement $\widetilde{\mathtt{M}}$ is performed on the state. That is,

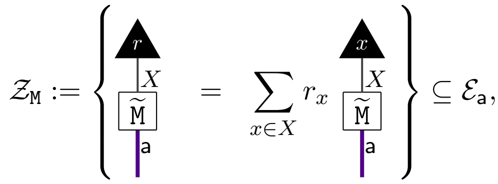


$= \mathsf{Prob}(x|\mathtt{P},\mathtt{M}).$ (2.33)

Like with the response functions that we created in **Stoch**, these can be thought of as linear functionals on V_{a} and so living in the dual vector space V_{a}^*. Often, however, we will introduce some (typically arbitrary) inner product on V_{a} and use the Riesz representation theorem to allow us to represent these effects in V_{a} as well.

While the geometry of states is quite straightforward (either some arbitrary convex bounded set in the case of normalised states, or the embedding of this into one dimension higher and taking convex combinations with the zero-vector in the case of subnormalised states) the geometry of effects is much more interesting.

For a given measurement, M, we obtain a *zonotope* of effects:



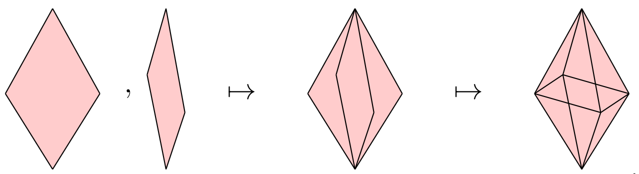
(2.34)

where $r_x \in [0,1]$. Note that r is an arbitrary response function here, and so there is no demand that $\sum_x r_x = 1$. Note now, that each of these $\mathcal{Z}_{\mathtt{M}}$ share (at least) two common elements, that is, if we set $r_x = 0$ for all x, then regardless of which measurement we consider we end up with the zero-effect, whilst if we set $r_x = 1$ for all x, then regardless of which measurement we consider we end up with the discard-effect.

Now, it can be shown that an arbitrary effect can be written as a convex combination of effects lying in these zonotopes, that is,

$$\mathcal{E}_{\mathsf{a}} = \mathsf{ConvHull}\left[\bigcup_{\mathtt{M}} \mathcal{Z}_{\mathtt{M}}\right]. \tag{2.35}$$

That is, we can think of the effect space as a bunch of zonotopes that are glued together at two points and then we take the convex hull of these. To give a simple pictorial example with just a pair of binary measurements, we have something like:



(2.36)

where in the first step we take the union of the two zonotopes gluing them together by the zero and discarding effects, and in the second step we take the convex hull of these.

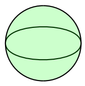
2.4 Example: quantum theory

We will now illustrate the generalised probabilistic theory formalism by casting quantum theory in this language. We will call the quantum GPT **Quant**. Our focus here again will be on preparations and measurements.

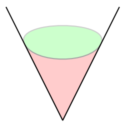
Typically we represent quantum systems by (finite dimensional) Hilbert spaces, and represent pure states as vectors in the Hilbert space or pure states as Hermitian operators on the Hilbert space. However, we've now seen that generalised probabilistic theories represent everything in terms of real vectors, so in order to see how to represent quantum theory as a GPT we need to see how to represent quantum theory with only real numbers.

It sometimes surprises people to hear that this is possible, but it is actually very straightforward – in fact, it is just a generalisation of the Bloch-ball representation of a qubit! In particular, one can note that the space of Hermitian operators on a Hilbert space defines a *real* vector space. This is because given some Hermitian operator h and a complex number α, then αh is Hermitian if and only if $\alpha = \alpha^*$, that is, if α is real. Hence, we can only take real linear combinations of Hermitian operators, $\sum_{i\in I} r_i h_i$ where $r_i \in \mathbb{R}$, and so this forms a real vector space. Let's denote this real vector space as H. It is clear that quantum states form a convex set within this real vector space, as $\sum_i p_i \rho_i$ is always a valid density matrix.

In the case of the qubit, the geometry of normalised states, i.e., those satisfying $\mathsf{tr}(\rho) = 1$, is given by:

 (2.37)

which is a 3D convex state space, this is then embedded in a 4D real vector space such that we can represent subnormalised states as well. A particular 3D slice of this is given by:

 . (2.38)

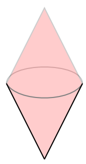
This tells us that we can represent quantum states in terms of real vectors, but what about quantum effects?

Well, typically the probabilities of outcomes of measurements are computed using the Born rule:

$$\mathsf{Prob}(k|M,\rho) = \mathsf{tr}(M_k \rho) \tag{2.39}$$

where $M = \{M_k\}$ is a POVM where k labels the outcome of the measurement and the M_k are positive Hermitian operators which sum to identity, $\sum_k M_k = \mathbb{1}$.

To see that this is the same as we have in the GPT formalism, first note that $\mathsf{tr}(M_k\_)$ is a linear functional on the space of vector space of Hermitian operators, and so lives in the dual vector space – these are our effects for the quantum GPT. We can then see that $\mathsf{tr}(\_\_)$ defines an inner product on the vector space of Hermitian operators, and so the POVM elements M_k are simply the Riesz representation of the quantum effects! Geometrically, a 3D slice through the vector space with these effects for a qubit can be drawn as:

 . (2.40)

We can see this as being of the form:

$$\mathcal{E}_{\mathsf{a}} = \mathsf{ConvHull}\left[\bigcup_{\mathtt{M}} \mathcal{Z}_{\mathtt{M}}\right]. \tag{2.41}$$

by noting that the number of different binary measurements for a qubit is infinite.

In particular, we get one zonotope per point on the surface of the Bloch sphere. The effect space is then the union of all of these!

This kind of representation can also be extended to quantum transformations, we just need to note that CPTP (or CPTNI) maps between quantum systems can be viewed as linear maps between the real vector spaces of Hermitian operators. The details of this however go beyond what we have time for here.

Understanding quantum processes in this language gives some interesting insight into these mathematical objects that we often take for granted. For example, what should we now say if we are asked "what is a density matrix"? Well, from the GPT perspective a density matrix is a mathematical way to represent an operational equivalence class of preparations for the system. Similarly, CPTP maps are a mathematical way to represent operational equivalence classes of transformations. Finally, POVM elements are then a mathematical tool for computing probabilities of obtaining particular outcomes when performing a measurement on a quantum system.

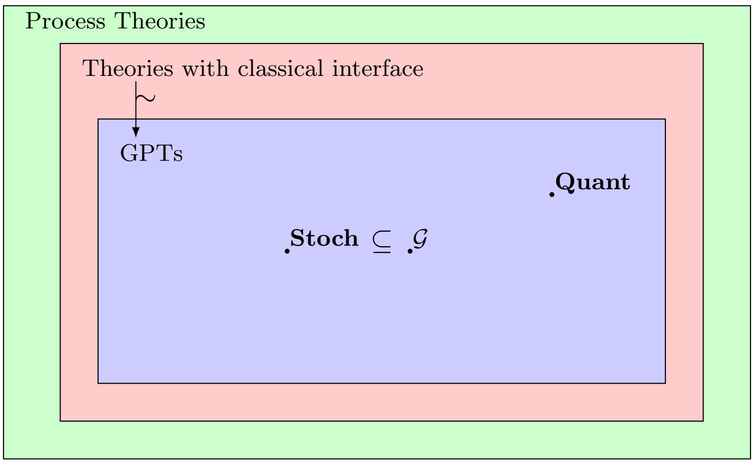
2.5 Summary

In this lecture you've seen how a particular class of process theories correspond to the well studied formalism of GPTs. We saw that there were three steps involved in this, the first was to restrict our attention to process theories that have a well defined classical interface, the second was to identify processes that differed only by context, and the third was to give a concrete linear representation of the involved processes. We then briefly saw how the geometry of quantum states, effects and transformations looks like from this perspective. In particular, this highlights the fact that our standard quantum processes such as density matrices, and CPTP maps, are in fact representatives of operational equivalence classes of processes. We have not yet, however, touched on how **Quant** and **Stoch** interact with one another (even though this interact was central to defining GPTs) the reason for this is because we will discuss this in depth in the next lecture, albeit from a complementary perspective.

Lecture 3

The landscape of physical theories

In the first two lectures we saw how to describe a particular theory within some framework for describing physical theories. In particular in the first lecture we talked about a very general framework known as process theories, and in the second we focused on a special case of these known as generalised probabilistic theories. We can consider either framework as describing a 'landscape' of hypothetical physical theories, with quantum and classical theories as particular points within this landscape. Moreover, we know that there is some structure to this landscape, for example, we know that there is a subset of the theories within the landscape of process theories which are generalised probabilistic theories. More interestingly, we have also started to see relationships between different process theories. For example, we know that **Stoch** is a subtheory of any other GPT, and we know that quotienting my operational equivalences maps one process theory to another.



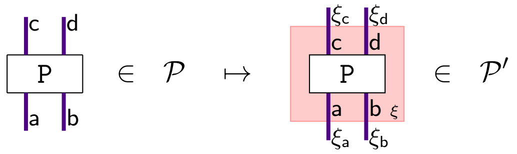
(3.1)

In this final lecture I want to flesh out this picture to some extent. I want to formalise the structure of the landscape of physical theories. Ultimately I will use this to understand the emergence of one physical theory from another via a generalisation of the notion of decoherence.

3.1 Diagram preserving maps

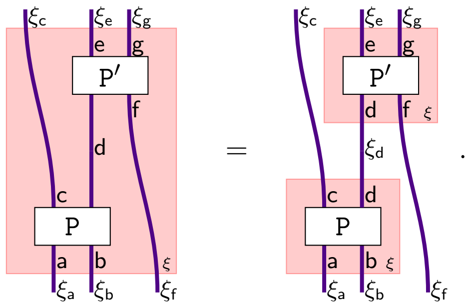
The key tool that we will use to explore the landscape of process theories is the notion of a diagram preserving map. These are a flexible tool which formalise the relationships discussed above and much more. They map the systems and processes of one process theory, $\mathcal{P}$, to the systems and processes of another, $\mathcal{P}'$, in a way that commutes with forming diagrams. We denote these as $\xi : \mathcal{P} \to \mathcal{P}'$ for reasons that will soon become apparent.

These map systems $\mathsf{a} \in \mathcal{P}$ to systems $\xi_\mathsf{a} \in \mathcal{P}'$ and processes as



(3.2)

which commutes with forming diagrams, for example:



(3.3)

This means that we can always merge these shaded boxes together or split them up around individual processes in a diagram.

PRESERVATION OF IDENTITY AND SWAP

It turns out that we have already encountered a couple of diagram preserving maps in this course already, we just didn't formalise them as such.

The first instance we encountered comes from the fact that whenever we have a subtheory $\mathcal{P} \subseteq \mathcal{P}'$ we can define a diagram preserving inclusion map $\iota : \mathcal{P} \to \mathcal{P}'$. So, for example, when we said that **Stoch** is a subtheory in any GPT $\mathcal{G}$ then we can define the inclusion map $\iota : \mathbf{Stoch} \to \mathcal{G}$.

The second instance we encountered comes from the fact that whenever we have a congruence relation $\sim$ for a process theory $\mathcal{P}$, we can define a diagram preserving quotienting map $\sim : \mathcal{P} \to \mathcal{P}/\sim$[1]. So, for example, when we said that the equivalence relation defined by operational equivalence was a congruence relation, then this means that going to the theory of equivalence classes $\mathcal{P}/\sim$ can be seen as an application of the diagram preserving map $\sim$.

[1]There is a bit of overloaded notation here, in that we are using $\sim$ both for the congruence relation itself, but also the diagram preserving quotienting map, hopefully it is always clear from context which way this symbol is being used.

In the next section we will see a more interesting application of this formalism.

3.2 Decoherence

In quantum theory, the leading explanation for the emergence of the classical world around us comes from the notion of decoherence. At the most simplified level, we can think of decoherence for some d-dimensional quantum system as a CPTP map $\mathcal{D}$ defined by $\mathcal{D}(\rho) = \sum_{i=1}^{d} |i\rangle \langle i| \rho |i\rangle \langle i|$, this maps general density matrices ρ to those that are diagonal in some orthonormal basis described by the $|i\rangle$'s. Such density matrices are in one-to-one correspondence with probability distributions over the finite set $\{1, ..., d\}$, and so, intuitively, it seems that decoherence is leaving us with something that is essentially classical.

Now, there are many things to explain about decoherence as an explanation for emergent classicality, for example: what picks out the particular basis $\{|i\rangle\}$ and in what physical situations does $\mathcal{D}$ actually capture the dynamics of what is going on? This is a big field of study and I am not even going to attempt to do justice to it here. Instead, what I'm going to do is show how we can capture the essential features of decoherence in the language of process theories.

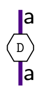
There are two key things that I hope to achieve by doing this. On the one hand, I hope to see not just how classical states (i.e., probability distributions) emerge from quantum states under decoherence, but how the entirety of classical theory emerge from quantum theory. On the other hand, I also want to go beyond just looking at the quantum to classical transition, but look at transitions between arbitrary theories! Of particular interest here, will be in the next section when we ask the question, is it possible that there is some beyond-quantum theory that decoheres to quantum theory?

Let's go back to the quantum decoherence map $\mathcal{D}$. In particular, we want to ask what properties of this map can be captured in an elegant diagrammatic form. We know that this is a CPTP map, so is automatically discard-preserving. More interestingly, we know that this is an idempotent map, that is, $\mathcal{D}(\mathcal{D}(\rho)) = \mathcal{D}(\rho)$ for all states ρ.

If we want to think about the image of this map as being the set of classical states, then idempotence tells us that i) all states are mapped to classical states by $\mathcal{D}$, and ii) classical states are left invariant by $\mathcal{D}$.

This motivates the following definition:

Definition 3.2.1 (Decoherence processes)**.** *In any process theory $\mathcal{P}$ with discarding, a process*



(3.4)

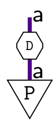
is said to be a decoherence process if and only if it is discard-preserving and idem-

potent, i.e.,

(3.5)

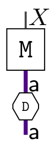
If we turn back to quantum theory for a second, then it will probably be clear to you that there are idempotent and discard-preserving processes that are not simply $\mathcal{D}$ with respect to some basis, to give a trivial example, we have the identity process! We also have processes that map general density matrices into a block-diagonal (rather than strictly diagonal) form, and a bit more too. We'll return to see what this added generality buys us later in the lecture.

We can then talk about the decohered states as those of the form:



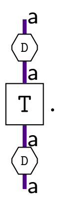
(3.6)

and the decohered measurements as those of the form:



(3.7)

and more generally decohered transformations for the system as those of the form:



(3.8)

In the case of the quantum decoherence map $\mathcal{D}$, we find that this gives us states, transformations, and measurements that are isomorphic to classical probability distributions, stochastic dynamics, and classical 'measurements'[2], respectively. Hence, we see that decoherence leads to much more than just the emergence of classical states!

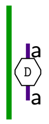
The idea now, is that we want to construct a new theory in which decohered systems live. However, as mentioned before, there are often many different decoherence processes for any given system, and so how do we pick which ones to use? Well,

[2]I.e., stochastic maps from the classical system to the outcome set.

the simple answer is don't! To start with at least we're going to construct a new theory in which arbitrarily decohered systems live. This is actually a well studied construction in category theory known as the Karoubi envelope[3], so we don't need to go about reinventing the wheel in this case!

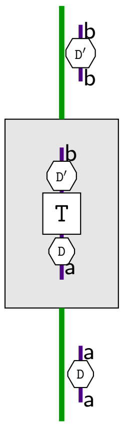
Definition 3.2.2 (Karoubi envelope)**.** *Given a process theory with discarding $\mathcal{P}$ we can define a new process theory $\mathcal{K}[\mathcal{P}]$ in which systems correspond to decoherence processes in $\mathcal{P}$ and in which processes correspond to decohered processes in $\mathcal{P}$. Composition of processes in $\mathcal{K}[\mathcal{P}]$ is then given by simply composing the decohered processes in $\mathcal{P}$.*

Let us start with some very explicit notation for this theory, and later we can introduce a shorthand for this. To begin we will denote systems in $\mathcal{K}[\mathcal{P}]$ with green wires labelled with an idempotent process in $\mathcal{P}$, that is, as:



. (3.9)

We then denote proesses as gray shaded boxes labelled by appropriately decohered processes in $\mathcal{P}$, that is, as:

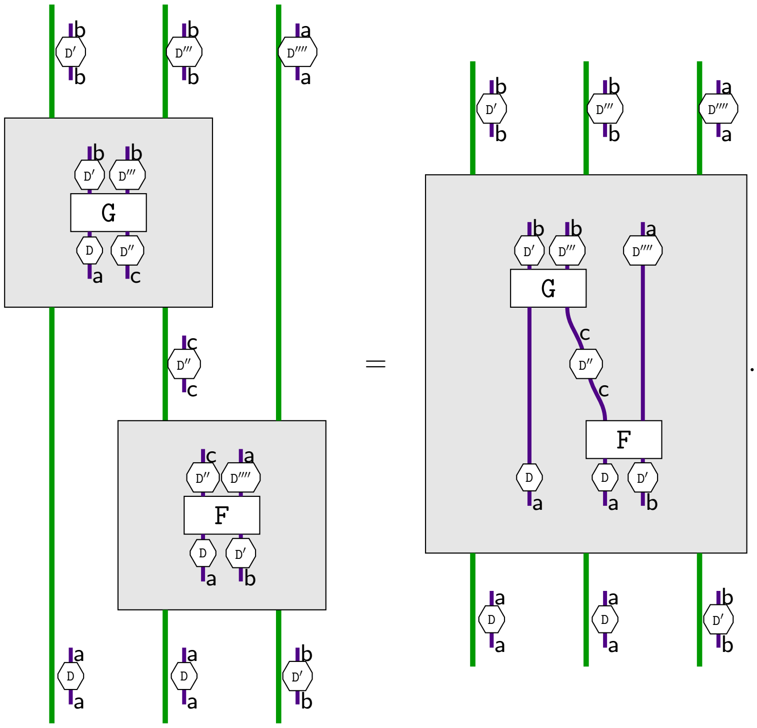


. (3.10)

Composition of these processes in $\mathcal{K}[\mathcal{P}]$ is then given by composing the associated

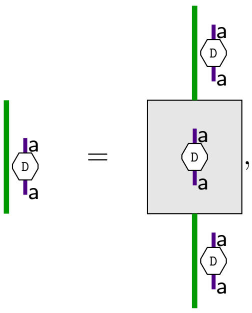
[3]Ok, so there is actually a subtle difference which I'm glossing over here, is that the Karoubi envelope strictly considers all idempotents, whilst here I'm only talking about discard-preserving ones. This doesn't really change anythign though.

decohered processes in $\mathcal{P}$, for example:



(3.11)

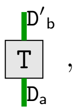
It is interesting to consider the identity processes in $\mathcal{K}[\mathcal{P}]$ as we have that:



(3.12)

as the identity transformation for $\mathtt{a}$ is not a decohered process (relative to $\mathtt{D}$) and so the idempotent $\mathtt{D}$ (which is a decohered process) acts as the identity instead.

As mentioned above, this notation is quite heavy and this use of "nested diagrams" takes up far too much space on the page, so we tend to use a simplified notation. It is worth always having this more explicit notation in mind however, as it helps to avoid mistakes. In particular, we can label wires in $\mathcal{K}[\mathcal{P}]$ by $\mathtt{D_a}$ without losing any important information. We can also leave the fact that the label on the box is a decohered process in $\mathcal{P}$ implicit. Then, for example, we can write a process in $\mathcal{K}[\mathcal{P}]$ as:



(3.13)

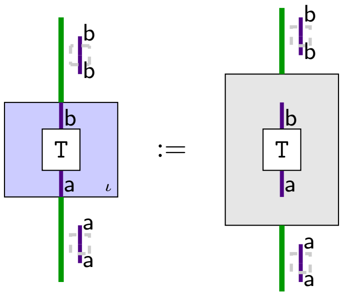
where we now need to make sure that T is indeed a decohered process in $\mathcal{P}$ with input a and output b, such that it satisfies $\mathtt{T} = \mathtt{D_b} \circ \mathtt{T} \circ \mathtt{D_a}$. The reason being that this is no longer explicitly enforced by the notation.

Now, if we consider the subtheory of $\mathcal{K}[\mathcal{P}]$ defined by restricting to only the systems labelled by the identities in $\mathcal{P}$, that is, those of the form:

(3.14)

then it is hopefully clear that the theory we get is basically the same s $\mathcal{P}$. For example, the set of processes between this kind of systems is always the full set of processes in $\mathcal{P}$, as decohering processes by the identity does nothing. However, how can we formally capture this intuition? The answer is diagram presrving maps!

What we can do is define a diagram preserving map from $\iota : \mathcal{P} \to \mathcal{K}[\mathcal{P}]$, defined by:



(3.15)

for an arbitrary process T in $\mathcal{P}$. It is straightforward to verify that this is indeed diagram preserving, it follows immediately from the definition of composition in $\mathcal{K}[\mathcal{P}]$ and the above definition of ι. There are a couple of important properties of ι which capture this intuition that $\mathcal{P}$ is the kind of subtheory I described above.

Firstly, it is *full* (i.e., surjective on processes). This means that any process in $\mathcal{K}[\mathcal{P}]$ between systems $\mathbb{1}_{\mathtt{a}}$, is in the image of ι.

Secondly, it is *faithful* (i.e., injective on processes). This means that if two processes in $\mathcal{P}$ are distinct, then they are distinct when ι is applied to them.

These conditions together capture the intuitive idea that if we restrict $\mathcal{K}[\mathcal{P}]$ to particular systems (namely $\mathbb{1}_{\mathtt{a}}$) that we obtain the theory $\mathcal{P}$. If ι were not full then we would not simply be restricting to particular systems, but also to particular processes for those systems, whilst if ι were not faithful then this subtheory would not be capturing all of the information about $\mathcal{P}$.

While this is hopefully a good example to get you some intuition for this definition, it is not really a particularly exciting example, we've done some elaborate construction only to get back to the thing we started with! What is much more interesting, is whether **Stoch** is a subtheory of $\mathcal{K}[\mathcal{P}]$ in the same sense. That is:

Definition 3.2.3 (Decoherence)**.** *We say that a process theory $\mathcal{P}$ can decohere (to classical theory) if and only if there exists a full and faithful diagram preserving map ι :* **Stoch** $\to \mathcal{K}[\mathcal{P}]$*.*

Let us now consider quantum theory, and in particular, let us start by constructing $\mathcal{K}[\mathbf{Quant}]$. One can show that

Proposition 3.2.4. *There is a full and faithful diagram presrving map* $\iota : \mathbf{Stoch} \to \mathcal{K}[\mathbf{Quant}]$*, which, in particular, maps a stochastic system* X *to a system in* $\mathcal{K}[\mathbf{quant}]$ *which is a quantum decoherence map* $\mathcal{D}$ *for a* $|X|$*-dimensional quantum system.*

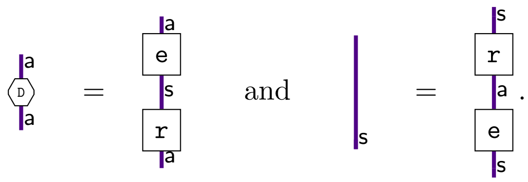
This means, in the formal sense I defined above, that quantum theory decoheres to classical theory. Now, the formal proof of this is all a bit beyond the scope of this course, but there are some nice things that we can understand given this result.

The above result tells us that $\mathcal{K}[\mathbf{Quant}]$ is a theory in which we have both **Stoch** and **Quant** as full subtheories. It is therefore interesting to ask about the sorts of interactions that we get between them.

The key result that we will use is the following:

Proposition 3.2.5. *Any decoherence process in* $\mathcal{P}$ *splits in* $\mathcal{K}[\mathcal{P}]$*.*

Firstly, what does it mean for a decoherence process to split? Suppose we have a decoherence process `D` then it splits if we can find a system `s` and processes `e` and `r` such that:

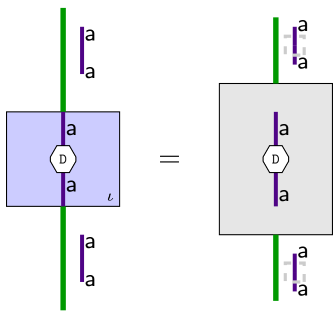


(3.16)

Intuitively, we think of `s` as the subspace of `a` which `D` is mapping everything into. `r` is then mapping everything into this subspace, and then `e` is mapping this subspace back into the full space. The second equation then tells us that if we map the subspace into the full space, and then back into the subspace, that this leaves everything in the subspace invariant.

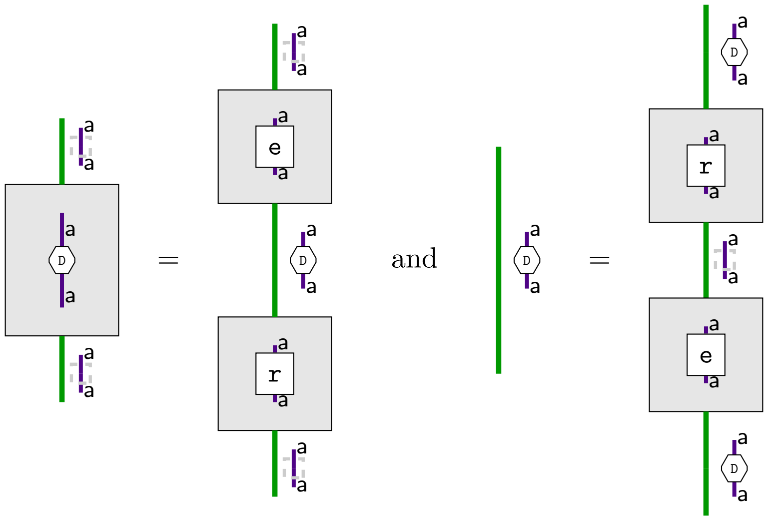
Note that the systems that `D` can split through are all isomorphic to one another, and that we can in fact derive idempotence of `D` from these two equations!

Now, when we say that any decoherence process in $\mathcal{P}$ splits in $\mathcal{K}[\mathcal{P}]$ what we really mean is that:



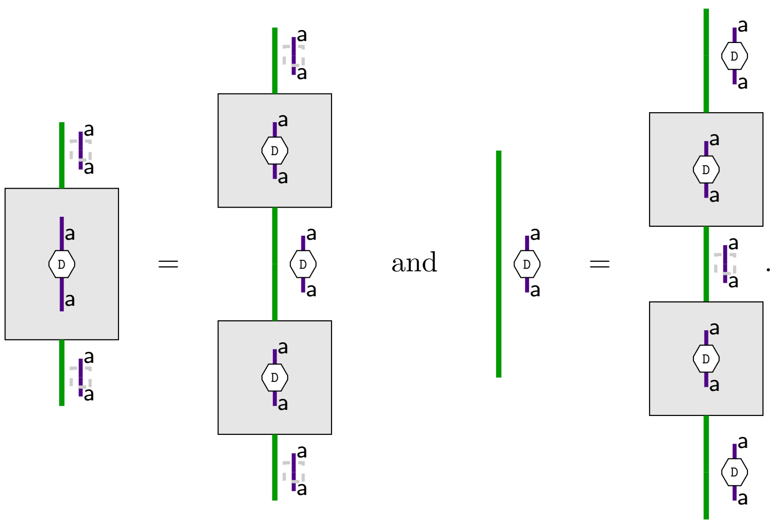
(3.17)

splits in $\mathcal{K}[\mathcal{P}]$. In particular, it splits through the system labeled by itself! That is, we can write:



(3.18)

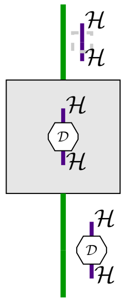
In order to prove this we just need to find suitable processes `e` and `r`, it turns out that each of these can be taken to be `D` itself, that is, it is straightforward to verify that the following equalities are valid:



(3.19)

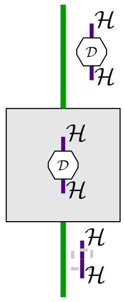
The tricky bit in showing this is remembering how the identities in $\mathcal{K}[\mathcal{P}]$ were defined in Eq. (3.12).

Let us now consider what this means for $\mathcal{K}[$**Quant**$]$. Well, we have already seen that if we consider decoherence channels $\mathcal{D}$ that these give us **Stoch** as a full subtheory. And that if we consider identity channels $\mathbb{1}$ then we get **Quant** as a full subtheory. If we now consider the splitting of $\mathcal{D}$, then what this is giving us is a perfect encoding and decoding of a classical system into a quantum system. In other words:



(3.20)

is a classically controlled state preparation and



(3.21)

is a measurement of the quantum system which perfectly distinguishes the set of prepared states. These encodings are also *minimal* in the sense that the dimension of $\mathcal{H}$ matches the number of classical states being encoded. We then know from quantum theory that this must be preparing an orthonormal basis of quantum states $|i\rangle\langle i|$ and that the measurement is measuring this orthonormal basis.

We therefore have seen that not only does $\mathcal{K}[\mathsf{Quant}]$ contain both Quant and Stoch as subtheories, but that they interact in a non-trivial way as well. One can go further and show that in fact all logically consistent ways that these could interact is described in $\mathcal{K}[\mathsf{Quant}]$, but we will focus on two results that are relevant for GPTs.

Proposition 3.2.6. $\mathcal{K}[\mathsf{Quant}]$ *contains all classically controlled quantum processes.*

Proof sketch. ***TO DO*** □

Proposition 3.2.7. $\mathcal{K}[\mathsf{Quant}]$ *has sufficient measurements for quantum process tomography.*

Proof sketch. ***TO DO*** □

What this tells us is that if we restrict to the processes in $\mathcal{K}[\mathbf{Quant}]$ belonging to either the classical or quantum subtheory (or composites thereof), then we have a GPT description of quantum theory with a classical interface!

What this means is that we do not need to treat the classical interface as part of the fundamental description of nature, that is, we do not need to imagine that there are systems floating around that are intrinsically classical. Instead, we can view the systems in the classical interface as being emergent from some deeper description of nature.

What is interesting is that this is not true for all GPTs! There are some GPTs for which there is no way to see the classical interface as emergent from the other systems. If such a theory were to be a real theory of physics it would therefore require that one posits that there are systems in the world that are fundamentally classical. The idea that classical systems should be emergent rather than part of fundamental physics seems quite a natural one, it is therefore interesting to ask what features must a GPT have in order to allow for emergent classicality.

3.3 Hyperdecoherence

We have seen the formal process-theoretic sense in which classical theory can emerge from quantum theory. This offers at least a partial explanation for why we don't see quantum phenomena in our daily lives, and why for so many years classical theory adequately explained everything we saw in experiments.

It is natural to speculate that there might be some deeper theory of nature that will one day replace quantum theory as our best theory of nature, just as classical theory was replaced by quantum theory. However, if there is such a *beyond-quantum* theory, then it should offer some sort of explanation for why we haven't yet seen it in our experiments. One candidate explanation could be by some mechanism analogous to decoherence. We call this mechanism *hyperdecoherence*.

One of the great things about the process-theoretic definition of decoherence, is that it is immediately obvious how to generalise it to talk about this sort of situation. That is, we can define:

Definition 3.3.1 (Hyperdecoherence)**.** *We say that a process theory $\mathcal{P}$ can hyperdecohere (to quantum theory) if and only if there exists a full and faithful diagram preserving map ι :* $\mathbf{Quant} \to \mathcal{K}[\mathcal{P}]$.

There is not a huge amount of research on hyperdecoherence as of yet, however, there are some interesting preliminary works.

3.3.1 No-go theorem

For example, one can show that

Theorem 3.3.2. *Any GPT which hyperdecoheres must either violate the principle of purification, and/or causality.*

We have talked about causality quite extensively already, however this is the first time that I have mentioned purification. Essentially, it is the GPT analogue of quantum purifications, this has two parts. Firstly, the idea that any mixed state can be understood as being a pure state (the purification) on a larger system with part of the larger system discarded. Secondly, that such purifications are essentially unique, that is, can be mapped to one another by reversible transformations on the purifying system.

This result is interesting because it shows how we can still manage to prove results about beyond-quantum theories, even without having any experimental evidence of what they will be like. That is, simply consistency conditions with what we observe in our current experiments can still teach us a lot!

3.3.2 Density hypercubes

One can also show that if we slightly modify the notion of hyperdecoherence, then it is possible to explicitly construct a beyond-quantum theory. In particular, if we remove the requirement that decoherence maps are discard-preserving, then we can find a theory that hyperdecoheres to quantum theory.

This is the theory known as density hypercubes, which, uses rank-4 tensors to represent states (compared to quantum theory that uses rank-2 tensors, i.e., density matrixes, and an earlier theory known as density-cubes which uses rank-3 tensors).

3.4 Summary

In this lecture I have started to flesh out the structure of the landscape of physical theories. Whilst the mathematics behind this has existed since the 60s it is only very recently that this structure has started to be studied in the context of generalised physical theories. Nonetheless, I hope to have given you a flavour of the utility of these tools. We've seen how they can be used to see the emergence of classical theory (as a whole) from quantum theory, and how this can easily be generalised to start considering , and proving results about, theories beyond quantum too!

3.5 Resource theories

This topic really needs at least another lecture to do it justice, but I wanted to flag up a couple of interesting points here and to point you to references for further reading.

The first is that there is an elegant formulation of the idea of a resource theory in the language of process theories. Resource theories provide a way to quantify and reason about things (e.g., quantum phenomena) that we deem to be useful for various tasks. For example, resource theories of entanglement, nonlocality, and so on, have all been developed.

The basic idea behind the process-theoretic formulation of resource theories, is that if we have a process theory that describes what is possible in the world, then in order to understand resources we partition these processes into two. We have the processes that we can implement (essentially) for free and those that we cannot. We can then order the resourcefulness of resources, by saying that one resource, r_1, is more resourceful than another, r_2, if we freely transform r_1 into r_2. The reason being that anything that we achieve using r_2 could equally well be achieved by r_1 simply by first transforming r_1 into r_2.

Anyway, the details of how all of this formally works are not important for now, what I want to highlight is the fact that resource theories can be expressed very generally in the language of process theories. We have also seen how decoherence can be very generally discussed in the language of process theories. The conjunction of these two leads us to being able to understand resource theories of coherence in this very general setting as well!

Bibliography

[1] Aleks Kissinger, Matty Hoban, and Bob Coecke. Equivalence of relativistic causal structure and process terminality. *arXiv preprint arXiv:1708.04118*, 2017.