

Causality and the Combination Problem

Gregg Rosenberg

July 2012

Summary

The combination problem is a question about nature which does not present traditional philosophical conceptual issues. The part of nature which must be confronted is causality itself, and a new view of causality provides a very satisfying solution to the combination problem for panpsychism. To embrace this view, panpsychists must embrace two key re-framings of traditional problems, and also re-evaluate which questions about causality and mind are philosophically most interesting.

1. Framing the Combination Problem

“Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them in a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence ... Where the elemental units are supposed to be feelings, the case is in no wise altered. Take a hundred of them, shuffle them and pack them as close together as you can (whatever that might mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feeling were set up, a consciousness *belonging to the group as such* should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its *creation*, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they *evolved* it (1890/1950, p. 160, original emphasis).”

-William James

This well-known quote from James describes a couple of difficulties which we have come to call “the combination problem” for panpsychists. The first is alluded to by the image of that poor twelve word sentence which will never exist, because its component words are trapped in the heads of twelve men. This is a *composition of feelings* problem. What rules of composition could blend separate feelings (or thoughts) into a single feeling or thought?

The second image of the hundred “windowless” feelings, imagines success overcoming the first problem, but suggests the cost of success is a new *composition of experiencers* problem. The only way to get a new feeling more complex than the originals is to create a new substantial identity to feel it, and this new substantial identity will be strongly emergent from the feelings which occasioned its emergence. This is a problem because panpsychism is supposed to be an alternative to emergentism.

James was substantially right that panpsychism leads to strong emergentism. However he and others have been wrong to frame panpsychism and strong emergentism as competing alternatives. Both panpsychism and strong emergentism are implied by a much deeper theory which is attractive because of the way it makes sense of the metaphysics of causality, and many other questions in metaphysics, as well as providing insight into important questions in the philosophy of mind. Panpsychism shouldn't be adopted simply because it might be a solution to the mind-body problem. It is also a solution to the carrier-causality problem, the problem of what intrinsic natures carry the schemas of causality in our world. This makes a large difference in how many things related to panpsychism should be viewed, particularly the motivations for adopting it, its explanatory power and its relation to emergence.

To reach this point of view, one has to switch to a new frame on certain problems, so that oppositions and concerning questions which recur in many contemporary discussions of panpsychism and/or the mind-body problem seem less relevant. It has been years and years since I have cared whether a view is dualist, monist or pluralist; or whether something was strongly emergent or weakly emergent; whether a view was panpsychist or merely neutral monist; whether causes have to precede effects or why there is something rather than nothing.

What follows is a discussion of a series of frame shifts which lead to an in-principle resolution of the combination problem, while changing one's point of view on whether certain ontological outcomes are “good” or “bad” and whether certain questions are interesting or not.

2. Changing the First Frame: From a Combination Problem to a Boundary Problem

James asks us to imagine twelve people thinking of twelve words. Crowd them together and we do not create a sentence. This seems like a hard problem to resolve. But is it any harder than the following problems: Imagine eight glasses of water. Crowd them together as close as you like, and they do not make a puddle. Imagine one hundred organs in a freezer. Make the freezer as small and cramped as you like, but they do not make an animal. Imagine a billion cells in a petry dish. Make the petry dish as crowded as you like, they do not form an organ.

There are related problems which are not hard: Imagine twelve waves flowing towards a beach. Have them meet, and the twelve waves combine into a single wave. Imagine two electric fields combined to create a more powerful electric field. Imagine a billion cells in a fetus, combining to make an organ.

This second set of related problems show there are good combination rules for many different kinds of things, in which these things can combine to become another unitary thing of either the same kind or a different kind. This is not a *conceptual* problem. It's a problem about understanding nature deeply enough to know when things interact and combine, and when they don't. In some cases, the combination rules are matrix or vector addition rules for things like waves and fields, in which different things can combine to make new things and the old things are lost entirely. In other cases, the combination rules are rules of tight coupling and interaction, in which the old things combine to make qualitatively new things while still maintaining an integrity of their own.

We can imagine a panpsychist world in which phenomenal characters combine via matrix or vector-like addition rules, while experiencers combine via tight interaction rules. The difficulty is really the odd empirical fact, that bounded phenomenal fields exist surprisingly at a mid-level of the physical world, at a scale corresponding to physical activity in animal brains.

In a world with a “combination problem”, a panpsychist might predict primordial experiencers at the base of the world, simple fireflies of flickering experiencers at low-levels of space time, without any experiencers at higher levels: feeling, feeling everywhere, but not a drop can think.

A panpsychist can avoid this surreal picture via appeal to known types of combination rules. In a world where experiences combined according to matrix or vector rules, for example, one might predict an ocean of experience constituting universal consciousness, but no experiencers below that level: a comically cosmic god-mind, experiencing itself alone forever.

The Scylla and Charybdis of these two alternatives are what I call the Boundary Problem for

Phenomenal Individuals. How can one allow experiences to combine from the low-level to the mid-level, but in a structured way that does not run away from us? It is a re-framing of the combination problem in terms of a different but related problem.

The hard nut of the Boundary Problem is that animal experiencers possess a kind of inherent individuality at a physical mid-level of reality, which is hard to explain. If panpsychism is true, why do the boundaries exist *just so*? Panpsychists can easily find types of combination rules which could work to overcome the combination problem. However, once a panpsychist recognizes a known kind of combination rule might be in play, he or she can coherently hypothesize almost as many ways of determining boundaries for experiencing subjects as there are of seeing organization in the world's pattern of microphysical interaction. Constraining application of the combination rules to produce a meaningful outcome on the Boundary Problem is hard.

Moral: boundaries are harder to explain than combination. We are faced with the need to understand more deeply what it is to be an inherent individual in the natural world.

There is one relation at the boundary of physics and metaphysics which divides the world by its very nature: the relation of causal interaction. Causal interactions imply partitions: they divide the world into different, mutually influencing spaces, and do so at many levels of nature. We might discover, by looking at causality and causal interaction, that causal interactions have certain important aspects which distinguish natural individuals. For panpsychists, these individuals would be attractive candidates as the supporting ontology of experiencing subjects.

3. Changing the Second Frame: From Cause-and-effect to Causal Significance

Mainstream discussion of causality almost always begins and ends with discussions about judgments of cause and effect or the language of cause and effect. The guiding image is that of a probability tree. Time is a tree of different outcomes, the actual world is a path through the tree, and causality is a way of influencing or setting probabilities for each branch of the tree.

A “probability tree” in which causality is a way of setting/changing probabilities is the fundamental frame from which discussions of causality depart. Significant questions within this frame include: is causality epistemic or metaphysical? Can causality work only from the past to the future? How do we understand the relation between “cause” and “effect”? What are dispositions? This fundamental causality frame does not help the panpsychist, whose interest in causality is driven by a desire to understand causal interactions and the way causal interactions partition the world.

The probability tree is a flawed metaphor. The major problem with the probability tree is that the cause-and-effect relationship is a poor starting point for understanding the metaphysical basis of interactions and its relation to the world's structure, as our cause-and-effect concepts contain many assumptions which could be false in a world with in which things interact, and are heavily intertwined with human psychological components.

For example, dynamics are necessary for cause-and-effect but not necessary for interesting causality in general. If different parts of the world constrained one another so that counterfactuals about states of the world as a whole were true, one could say interesting things about laws and about causal constraints between different parts of the world, even without dynamics. Concretely, imagine a timeless world which was nothing but an unchanging Rubik's cube. Some combination of colors on its six faces is showing. Other combinations *could* show, and there is an atemporal physics describing which ones are

possible and which ones are not (there is no physics of transitions). There are meaningful causal constraints in terms of constraints between components in the Rubik's cube world, even though there is no cause and effect.

Of course examples of the subjective nature of cause-and-effect are well-known. The idea of “cause” is closely related to agency and blame, and “effect” to goals and means. Negative facts such as absences can be said to cause things. Figure-ground relations, levels of granularity and so forth all impact cause-and-effect judgments. And just try to get agreement about cause-and-effect in a political situation, such as the Israeli-Palestinian conflict.

Our common concept of cause-and-effect also builds in several other parameters: locality, directionality, categorical constraints on relata, the arity of the relation, all of which are baggage on a more basic notion. To think of causality in a fresh way, we need to get rid of the baggage and pare it down to its core essential truth: the world is in a constrained state, in that having one part of the world in one state, places constraints on what states other parts of the world can be in. This truth captures the fundamental mystery of non-Humean causation.

It doesn't matter if the parts of the world are in different time slices, or if they are local, or if there are two of them versus three, four, or an infinite number. It doesn't matter if we think of these parts of the world as events or states or facts. It doesn't matter if one is a “cause” and the other is an “effect”. If one part of the world being in one state, places a meaningful physical or metaphysical constraint on the state of another part of the world, we are presented with a mystery of causality.

To express this mystery, I like to talk about the Two Canvases of Causation.

Imagine two great, blank canvases that you paint with color one drop at a time. Imagine also that the two canvases are two very different kinds of surfaces to work with. You call the first canvas the Humean Canvas. It acts like a normal canvas, as it will accept any drop of paint anywhere on its surface in any color that falls on it. If you let a drop of red paint fall onto the Humean canvas, it will stick where it lands. The same will happen if you then drop a speck of yellow paint somewhere else on the canvas. You can fill the whole canvas this way, with whatever colors you want, anywhere you want.

You call the second canvas the Canvas of Causation. It is a marvel. If your first drop of paint is a bit of green, and then you try to place a drop of red next to it, the red paint will bounce off. The canvas will not accept it. But it will accept yellow. And the more paint you put on the canvas, the pickier it seems to become. Each bit of color you put on the surface seems to place a constraint on what colors may appear elsewhere on its surface. Although the canvas will allow you to paint it in many different ways, it will accept only combinations of color that make for a beautifully covered canvas. If you try to pore many colors on it all at the same time, some will stick and some will run-off, and each time what remains after the run-off will be a beautiful pattern of color. Every color and every drop matters, jointly enforcing or excluding the colors that will finally appear on the canvas.

Although the Humean Canvas is ordinary, the Canvas of Causation seems like magic. Yet the Canvas of Causation seems to be more like the world we actually live in. It is a world in which nature includes and excludes membership based on what else has made it into the club. Making our world be like a Canvas of Causation requires some extra ingredient over and above simply having a world in which things can happen. This extra ingredient, whatever it is, should be what a theory of causality should be about.

The central concept for causality, in this view, is *causal significance*, not cause-and-effect. The *causal significance* of a thing is the constraint its existence adds to the space of possible ways the world can be.

An emphasis on *causal significance* is a second kind of re-framing which panpsychists need, to make progress on their world view. They need to re-frame the problem of causality from being a problem about cause-and-effect to being a problem about causal significance, because a theory of causal significance yields many unexpected and very important explanatory benefits, not only in the philosophy of mind but elsewhere.

The shift away from cause-and-effect as the core concept of causality also requires shifting away from probability trees as the core framing device for discussions of causality. If we emphasize causal significance, it seems the central job of causation is to act on potentialities in ways which constrain possible states of the world. In other words, causality doesn't just make *future probabilities*. It makes the *world actual*. The proper framing device for causal significance is to think of causality as a *potentiality filter*, rather than a probability tree. On one side of the filter is a set of potentialities, and on the other side is a smaller set. [[note: in the past I've called this a *possibility filter* but I think potentiality filter is more appropriate, given that it acts on the potentialities of nomic individuals.]]

As soon as one starts thinking of causality using potentiality filters as a tool, it just becomes obvious that there is nothing inherently temporal about a potentiality filter. The potentialities fed into a potentiality filter don't need to flow from past states to future states of individuals. Efforts to restrict the formalism that way look obviously artificial and hard to justify.

The potentialities fed to the potentiality filter could come from individuals at a lower level of nature and be fed into an individual at a higher level. This scenario is mathematically, logically and metaphysically coherent. It also happens to cohere well with known physics.

Indeed, it becomes equally obvious that a core, unargued assumption of all classical metaphysics is that things at lower-levels of nature are already, in and of themselves, in determinate states, just as we find things at higher levels in ordinary experience. It is also assumed that things at higher levels of nature *must* inherit their determinate states from the determinate states of the lower level things which compose them. In philosophical parlance, all classical metaphysics assumes that determinate macro-level states of the world *must* strongly supervene on independently determinate micro-level states of the world. This statement is a kind of *micro-determination thesis*, or *mic-d* for short.

Why do we believe *mic-d*? There is, in fact, no evidence for it and quite a large amount of evidence against it.

The only real reason for believing it, is it is “intuitive”. But that is no reason at all. We have had no conceptual tools for thinking of anything different and so it is “intuitive” by default. Additionally, we know enough now to say with confidence that questions about fundamental nature tend to not have intuitive answers, and that things at the lowest levels of nature are not in-and-of-themselves in determinate states.

The problem is, panpsychists (and others interested) haven't had the vocabulary to raise this as an issue, or a formalism to model how it could work or be resolved. The assumption has been buried so deep, there's been no good way to talk about it. There also has been no good way to model worlds in which it is true and worlds in which it is false, to examine the differences between those worlds, and the

implications of those differences. To get anywhere, we need to shift from modeling probability trees extended through time, to modeling ladders of potentiality extended through levels of nature. A significant part of my project in *A Place For Consciousness* (2004) was to provide a vocabulary and formal model for asking these questions, representing differing answers, and exploring implications of those representations. What I found was that model worlds in which mic-d is false

- i. tend to have physics that make the physics of our world with its randomness, backward causality, indeterminate states, non-local causality and measurement problem, look expected rather than weird;
- ii. deductively require, just to make causality work, the presence of subjectivity with many of the funny properties consciousness seems to have (e.g., unified fields of intrinsic properties structured according to information-theoretic constraints);
- iii. possess higher-level individuals which can easily extend to the mid-levels of the natural world.
- iv. recast traditional oppositions between things like emergence and panpsychism, or dualism and monism, or mental causation and epiphenomenalism in ways that make them uninteresting.

4. The Theory of Causal Significance

In *A Place for Consciousness*, I introduced a vocabulary and formalism for representing both the truth and falsity of mic-d within a single metaphysics for causality, and for exploring consequences of a metaphysics in which mic-d could actually be an open question. The result of this effort was the Theory of Natural Individuals (TNI), in which the Theory of Causal Significance (TCS) was its heart.

TCS describes the causal connection as a potentiality filter, of a very general nature, in which the same metaphysical device produces temporal connections of cause-and-effect, causal interactions between individuals within a time slice, and also higher-level individuals which can make indeterminate lower-levels more determinate. TCS is a unifying framework, in which these three things are all aspects of the same thing, looked at from different angles.

By adopting TCS, we can ask metaphysical questions which previously were hard to ask and we can provide answers which previously were hard to even represent. Importantly for the panpsychist, TCS provides a solution to the combination problem which does not fall foul of the boundary problem.

TCS has a maximally simple ontology, consisting of two kinds of properties, one relation, and causal laws. Here is the furniture of TCS,

Property Types

Effective properties – Properties that contribute to constraints on a causal nexus.

Receptive properties – Connective properties that bind to effective properties, creating a causal nexus

Fundamental Relation

Binding – A primitive metaphysical relation in which one property becomes part of the completion of another property.

Causal laws – Laws describing relationships of compatibility, incompatibility and requirement between effective properties.

The ideology of TCS is slightly more complex than the ontology, as it introduces two new concepts: receptive properties, and the concept of *completion* which occurs in the definition of the binding

relationship.

Receptive properties in TCS play the role of the potentiality filter – they are the causal connection existing between effective individuals. If binding occurred directly between effective properties, the model could not simultaneously represent their interactions and establish a plurality of individuals, some at different levels of nature than others. The world would necessarily be a single-level, homogenous mesh.

In TCS, two or more effective properties can bind to the same receptive connection, which then establishes a context where the intrinsic constraints between them can be activated and structured. Think of receptivity as a kind of neutral background field, whose essence contains the possibility of interaction between effective properties, and which establishes the structure of interaction and causality in the model.

The concept of *completion* is an important ideological innovation in the theory. It is part of a cluster of concepts which also include *determination* and *concreteness*. The idea is that properties considered in isolation from their binding relation to other properties, do not have *determinate* states, which is interpreted to indicate a kind of essential incompleteness in their isolated nature. Properties *need* context to be anything in particular. Otherwise they are none of many things.

The state of such a context-free property needs to be represented as a disjunction of its potential states. TCS says such properties considered in isolation are incomplete and need to bind to other properties to (take on context and) become complete.

Yet, the model also holds that no properties exist in the world unbound from other properties, so such property essences are *abstract*. This is why in TCS the terms *complete*, *determinate* and *concrete* are cognates and need to be understood together, as do their opposites, *incomplete*, *indeterminate*, and *abstract*.

The model example for thinking about this is to consider a property like Spin. Inherently, spin can be either up or down. Considered in isolation – as an essence - spin as a property has an indeterminate state which is represented as a disjunction of those two values. According to TCS, spin, considered in isolation, is therefore an *abstract*.

Spin becomes concrete by binding to a receptive property, which may also bind to other effective properties to form an individual such as an electron (so a particle like an electron is modeled as multiple effective properties bound to a common receptivity, which gives it its causal unity as a particle). If that electron is not further constrained, it will inherit the indeterminateness of its spin, so the electron itself may need to bind to become determinate. This process may repeat again and again, until the constraint structure is strong enough to make spin take on a determinate value.

Some people who have read *A Place for Consciousness* have reacted to this ideology as if it were strange or exotic. In a sense, it is, compared to classical metaphysics. In another sense, it is not, as it is clearly a metaphysical recasting of well-evidenced aspects of standard physics. One might say it scores low on *intuitiveness* (it is not the first ideology one might think of) but high on *plausibility* (it is by far the ideology most coherent with our actual evidence about nature and causality). Overall, TCS is a simple theory ontologically, a moderately complex theory ideologically, and a very plausible theory from the standpoint of what we actually have reason to believe about nature.

TCS is the heart of a theory of natural individuals because it contains at its basis a recursive combination rule, which describes how causality creates natural individuals and can do so above the base levels of reality. The combination rule is simple,

Base rule: Any primitive receptive or effective property is a natural individual.

Recursive combination rule: Any receptive property which completes itself by binding to two or more other natural individuals, is a natural individual.

The combination of effective properties which occurs within a binding relation is treated like a matrix or vector combination of properties, occurring within the boundary provided by the receptive connection. The receptive property itself is irreducible causality, an emergent.

The purpose of a receptive connection – its occasion for existing and what it produces – is a determinate state, or a *more* determinate state, for otherwise indeterminate individuals. Receptive connections as emergents are not otiose or indulgent. They are metaphysical ground zero. They have an essential function in causality as the engine of creation, which must somehow make a determinate world from a set of indeterminate and abstract potentials. Completion happens because a receptive context enables additional constraints to be active on indeterminate individuals, which make them more determinate (and as a consequence, less abstract and more complete. Binding does not occur if all individuals are already determinate).

TCS thus describes causal significance: it gives a formalized account of how a potentiality filter – called a receptive connection – can metaphysically enable causality in the world by producing a determinate state from relations between indeterminate properties. Once formalized, the theory allows us to model causality in a world which is intrinsically determinate at its lowest level, as in classical physics; but it also allows us to model worlds which are *indeterminate at the lowest level but made determinate by the presence of higher-level individuals*. Additionally, it shows how the same mechanics at work between levels can produce dynamic causality across time slices.

Of the many kinds of worlds which can be modeled using TCS, the non-classical, multi-level worlds with indeterminate lower-level individuals are by far the most natural. Also, the intra-level structures which model interactions between individuals, and the inter-level composition relations between levels of nature, are rich and provide more insight into important metaphysical questions than does investigation of cause-and-effect relations across time.

For example, one can clearly show a causal role for higher-level individuals that is different than anything ever pictured in philosophical discussions of interactionism or emergent causation. The role of higher-level individuals is to act more like final causes, because the determination of the state of the higher-level individual is a cause of determination at the lower levels, without imposing any interaction force across levels or impacting energy: it is simply a selection force. In the reverse direction, the relation between the lower-level individuals and the higher-level individuals is more like material causation rather than pure composition (because of the presence of irreducible receptive connections).

This model then provides a potential third way to think of mental causation, aside from interactionism and epiphenomenalism. Mental states are just states of certain higher-level individuals, and they are part of a universal process of causality in which there are two-way causal relations between levels of nature. Higher-levels provide contexts for constraint, which make them final causes for the determination of otherwise indeterminate base levels; and the base levels, made determinate, provide material causation for the higher-levels. Effective causation remains limited to within-level relations.

There is a kind of strong emergentism in TCS, but it is a kind which renders the usual discussions and concerns about it uninteresting. In TCS, receptive connections bind individuals at different levels of nature, and they are not reducible. Each receptive connection is a constitutive property of its own individual, not composed from lower-level properties. But recall, this is just *causality*. It is happening just the same everywhere, and is not a special thing brought in suddenly at one level of nature, or just to explain mental properties. It is in no way ad hoc or surprising.

Other traditional discussions also become less interesting. One could call this a dualism of receptive and effective properties. Or a monism of the causal nexus. Or even a pluralism of intrinsic and extrinsic aspects of causality. One could claim this is a kind of Russellian neutral monism, a kind of physicalism, or a kind of non-physicalism. However, from the perspective of TCS, this sort of counting question does not advance any issue and does not grip the imagination.

5. The Carrier Theory of Causation

The relationship between TNI and panpsychism comes by noticing that TCS, for all its virtues, does not avoid the kinds of critiques of physical theory put forward by Russell and Whitehead. The natural individuals in TCS need intrinsic properties to perform the functional roles laid out by the theory. These intrinsic properties are called carriers, as they carry the extrinsic descriptions for effective and receptive properties within the theory. The theory which adds this postulate to TCS is called the Carrier Theory of Causation and together the two theories make up the Theory of Natural Individuals.

The carrier theory is an additional postulate to TCS, but because TNI contains both theories as components, we are able to use TNI to deduce specific requirements about what the intrinsic carriers of effective properties must be like. One can deduce, for example, that the carriers of effective properties would have to possess a kind of unity similar to the unity we find in consciousness, and would have to have intrinsic relations of compatibility and incompatibility similar to what we find among phenomenal properties like red and green. Also, using TNI we can deduce that the intrinsic carriers of effective properties must support scalar relations also similar to what we see among phenomenal properties in their intensity dimensions (e.g., the loudness of a tone).

When we turn our attentions to receptive carriers, we can deduce the carrier of receptivity must have a kind of contentless openness, similar to what meditative practices report as characteristic of the pure experiencing self, and that the receptive carrier's relationship to effective properties would have to be much like what we see as the relationship between experiencing and phenomenal properties.

For these and other reasons, someone trying to make sense of TCS in the world will feel compelled to hypothesize that phenomenal properties are the intrinsic basis of effective properties; and that an experiential property is the intrinsic basis of receptive connection; and that the causal nexus in our world is carried by the experiencing of phenomenal properties. In more ordinary language, we would say that the experiencing of phenomenal properties *is* the causal nexus in our world. In other words, a variety of panpsychism is likely true in anywhere there is direct interaction between natural individuals, there we will find experiencing occurring.

Panpsychism and TNI make for an agreeable partnership. For not only can TNI benefit from panpsychism by being able to answer questions about the intrinsic basis of the world, but panpsychists can help themselves to the combination rules in TNI *for free*. This is a strong position, because these rules resolve both the traditional combination problem and the more difficult boundary problem. Also,

because TNI is a theory of causality, it allows for deep answers to issues about mental causation, about determinism and perhaps about free will. It can provide a framework for actually deducing explanations for a variety of traditionally mysterious aspects of consciousness, such as its unity and relationship to time as well.

Very importantly, TNI vastly strengthens the motivation for being a panpsychist. The TNI-based panpsychist is no longer making *ad hoc* conjectures to resolve a specialty problem in the philosophy of mind, back-filling with more conjecture to resolve further problems which arise. Rather, he or she is making a principled conjecture to solve fundamental problems in the philosophy of nature, from first principles, while at the same time proposing to resolve problems in philosophy of mind and in metaphysics more broadly. The explanatory power of panpsychism when made part of TNI is vastly expanded, while nothing about it retains the whiff of *ad hoc* metaphysics (*speculative*, but not *ad hoc*).