

MANUEL VARGAS



Professor of Philosophy and Law, University of San Francisco

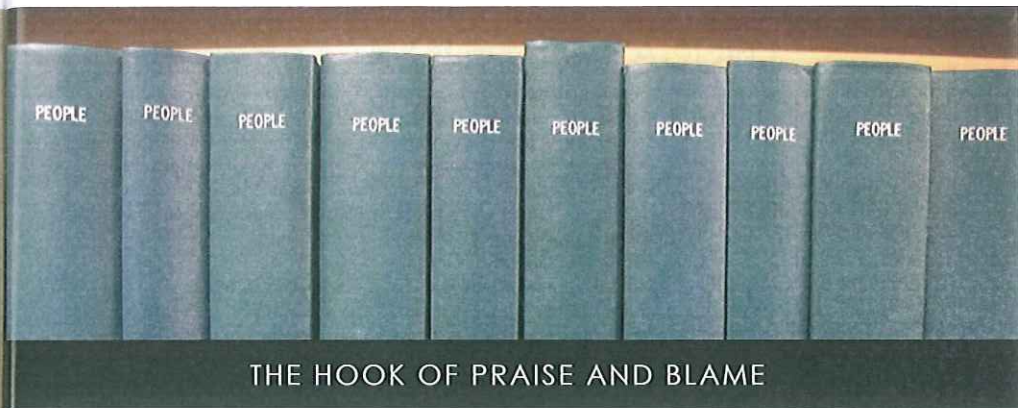
Distinguished Research Award, University of San Francisco, 2012

A.B. University of California, Davis, 1995

PhD Philosophy and Humanities, Stanford University, 2001

"The key, I think, is to acknowledge that we are part of the physical causal fabric of the world, but to also recognize that we are the part capable of acting with malicious intent.

That's the hook of praise and blame."



Nadia Rosenberg: So what is philosophical research, anyway?

Manuel Vargas: Philosophy is what we do when we have no idea how to solve a problem in some domain. You can do philosophy of nearly any subject matter – if you are interested in a question and you're unsure of what it would mean to give an adequate answer. Unfortunately, to do it well often requires such technical and painstaking detail that it's of little or no interest to those who are not specialists.

Let's take out the painstaking details. Why free will? Why is that the focus of your research?

I don't know. When I was an undergraduate, I thought the free will problem was incredibly uninteresting. I could see what the problem was supposed to be, but it just didn't grip me in any way.

Then I went off to graduate school at Notre Dame. I took a class from Peter van Inwagen who wrote this landmark essay on free will, arguably the most influential work written on free will in the past 100 years. I took a seminar from him where we were looking at this issue. I remember being really impressed at how difficult the issue was and how interesting it was.

Then I transferred to Stanford, and it

just so happened that a lot of people were teaching classes on free will and moral responsibility or related issues. I was struck by how the presentation of the puzzle was completely different. It was a presentation of the view that gets called compatibilism.

The basic upshot of that view is that there is no deep problem about free will. On this view, we can have free will even if everything happens deterministically.

People who defend compatibilism typically insist that there is only a puzzle about free will because we are confused about a few things. Once we give up the confusions, the puzzle

drains away and it turns out to be a philosopher's construct rather than a real puzzle. It seemed that everybody at Stanford had this view and they provided very compelling, powerful defenses of it.

Ok, I thought, there's no free will problem. Then I thought, wait a minute! What struck me as interesting was that there were these two powerful ways of presenting a subject matter: on the one hand, there is a deep puzzle, and on the other hand, there isn't.

I WAS WONDERING HOW IS IT POSSIBLE THAT ALL OF THESE REALLY SMART, WELL-EDUCATED AND NON-DOGMATIC FOLKS COULD HAVE SUCH WILDLY DIVERGENT VIEWS ABOUT THIS.

So I was gripped by that puzzle, in some sense, a meta-philosophical puzzle, a sociological puzzle. Ten years later, here I am.

Can you try and explain 10 years of research?

Well, I think it is important to distinguish between two kinds of questions. One is a question about what people think about free will and moral responsibility. That's distinct from thinking about what we ought to think about free will and moral responsibility. The first step is to just make the distinction.

With respect to the first question – what do we ordinarily think about free will? – there is good reason to think the answer is mixed. In our everyday pedestrian beliefs, people have conflicting beliefs about the nature of

their own agency.

For example, suppose we could roll the universe back to a few minutes ago, and suppose it then ran forward again in time. Could you or I be having a different conversation, given the actual laws of nature and the starting conditions at the point at which we rolled back the universe?

Some people have thought that we do have the ability to do otherwise given the same starting conditions and the same laws of nature. For such people, the idea that we have an alternative possibilities feature is important to their self-conception.

For other people, that's a less robust feature about how they think about themselves. So you can see how you get a philosophical problem up and running if some people have this view and other people don't.

Speaking about determinism, does that come from physics? Like a domino effect where this eventually pushes this to that and everything is pre-determined? When I was reading your work, there was no discussion of fate. Is determinism a philosophical thing or a physical concept?

There are a lot of related concepts banging around here: fate, determinism, causation, and so on, that have overlapping significance, but that are discrete ideas. First of all, let's just start with the idea of causation. You might think that everything in the universe is caused, but there are some cases in which it's probabilistic: with the exact same physical conditions, you will end up with an array of ways that things could go.

The other kind of causation is deterministic causation: given a set of fixed initial conditions, you will get the

same effect every time. In the context of philosophical disputes about determinism, we are being asked to imagine that all causes in the universe are like the second kind. So if determinism is true, there is only one way the universe could go given the starting conditions.

What is the relationship of that view – determinism, as understood by philosophers – to physics? The short view is that for philosophical purposes, physics is a big mess. A lot of people are inclined to think that the best way to understand quantum mechanics is as indeterministic, but larger-than-atom-sized things tend to be deterministic, with some important exceptions.

It's a bifurcated picture and it's an ongoing project to figure out how to merge these pictures. There are people who argue about whether that picture is right, but this is roughly the standard picture.

I'm inclined to think that the familiar focus on determinism is frequently a distraction. Most of the free will and moral responsibility debates can get entirely rewritten without appealing to determinism. Determinism was just the old traditional way of talking about it, and it turns out that it's helpful for getting people to see the problem.

I think the more global problem, the version of the problem that I'm interested in, is that irrespective of whether determinism is true, the mere fact that we are embedded in a system of physical causes is enough to get the free will problem up and running.

Suppose it's a purely random event whether you get up or stay in your chair. Then, in some sense, it will be

“A FIVE YEAR OLD IN A PADDED ROOM CAN ONLY DO SO MUCH DAMAGE. A FIVE YEAR OLD SITTING BEHIND A TANK CAN DO WAY MORE DAMAGE. A FIVE YEAR OLD WITH HIS FINGER ON A NUCLEAR WEAPON CAN DO EVEN MORE DAMAGE.

IN THE WORLD WE HAVE BUILT FOR OURSELVES, WE ARE NO LONGER FIVE YEAR OLDS IN A PADDED ROOM.”



disconnected from your character. If whether you tell the truth or are honest is entirely random, that seems to be of no help when we get to the moral evaluation of one another.

After all, it would appear like what we do is just part of a long chain of causes going back to the Big Bang or whatever the current story is of how it got started.

How is it that some people say there is no free will?

Some people are driven by concerns of determinism. Increasingly, it's a minority view. I think the bulk of philosophers who insist that we don't have free will are driven by the following kind of thing: there are a set of powers we need in order to be held responsible, and what we know about humans being causally embedded shows that we cannot have those powers. People get to that conclusion by different paths.

A lot of philosophers have been impressed by arguments that start by thinking about imaginary cases, like what if it turned out that we knew that your thoughts were caused by somebody actively manipulating your brain, and you were unaware of it. You can gradually modify these examples until you say, well, ordinary causal forces are brain manipulation.

So if you think that the first case, the manipulation case, is not a case of responsibility, then it may seem appealing to conclude that we never were responsible because all brain activities are products of a kind of manipulation — that is, ordinary causal forces.

The neuroscientists that I like to complain about, that have argued against free will, have typically done so on the

grounds that free will requires disembodied souls. Neuroscience has no place for disembodied souls. Therefore, free will doesn't exist.

At least 99% of the people who have written about free will in the philosophical tradition in the last 100 years have not accepted that initial premise, that you have to have an immaterial soul to have free will.

A LOT OF THESE DECLARATIONS ABOUT THE NON-EXISTENCE OF FREE WILL ARE NEUROSCIENTISTS GETTING CHEAP HEADLINES.

You just won an award at USF for the research you're doing. What is the research?

It's connected to a book I just published, titled *Building Better Beings*. It's a theory of moral responsibility. It's a theory about the conditions under which we praise or blame one another. The bulk of the book is concerned with the prescriptive project.

It goes something like this: suppose it turns out that we can't initiate new causal changes that are disconnected from what came before, that there is no scientifically plausible picture on which we have immaterial souls, and that we are just part of the physical fabric of the world.

If we don't have the kind of agency we'd like to have, one that affords us a degree of radical independence from the physical world, what is the basis on which we can hold each other responsible? That's the initial puzzle that motivates the book.

The answer I give in the book is roughly the following: the basis of responsibility rests in both our ability

to recognize and respond to moral considerations, and also in the way in which our social practices of judging and holding responsible foster the ability to appropriately respond to moral considerations.

You might ask, why would any of that be a thing that licenses praise or blame for the choices you make? Partly, it is a matter of knowing that we run the risk of being praised and blamed for our choices. The fact of this social practice helps us modify our own behavior as we go along. Knowing that helps us maintain and expand the domains where we are sensitive to moral considerations.

It reminds me of the question: if a tree falls in the forest and nobody bears it, did it fall? If you do something and no one is there to praise or blame you, isn't that dangerous? Does it rely on some outside witnessing?

This is a really great question. It's not that it's wrong to do this only if someone will catch you and give you negative feedback. Rather, it's wrong to do this, and in virtue of being wrong to do this, it would be appropriate to give negative feedback. We internalize those norms and they take on a life of their own — with any luck. We are worried about trees falling in forests even if nobody is there to observe them.

Where does your interest in culpability and blame come from?

I'm just hoping to avoid getting into trouble, that's the main thing! I just want to know what it is so that I can try and avoid it. [Laughs] That's partly true.

Also, the notion of a *should* or an *ought*, and the business of praise and blame, is a distinctively human dimen-

sion of our lives. It's probably human narcissism, this interest in something that seems to be distinctive about us.

It seems that the fragmentation of responsibility has increased. You hear about this in the military. Misconduct is brought

to light, and when people are being shamed about their actions, they say, 'I was just following orders.' That's where it becomes tricky. When do you take responsibility for your actions? Maybe you don't build a bomb, but you are putting the screws on it.

The military is a nice example, because there is a lot in the way in which military life is constructed that is going to undermine the individual's ability to recognize and respond to suitable considerations. The same is going to be true of any large-scale organization in which you have routinized behavior, where folks are inclined to think and say things like, 'Look. This is what we do here. This is the routine.'

In such contexts, we are not supposed to question the norms or practices. Organizations cannot function if everybody is questioning every detail about what they are being asked to do.

The great gift of bureaucratic capitalism brings this to us. However, one of the consequences is that folks who are in positions of authority, who set up those institutions and the policies that govern them, bear a heck of a lot more responsibility than we might ordinarily assume. One of the things I've tried to show in my own work is the importance of circumstances, of social context, for individual moral responsibility.

We are prone to thinking, no, no,

everyone is responsible for their own sphere of conduct, no matter the context. That's an admirable aspiration, but there is a special culpability for people at the top. This is the sense in which, metaphorically, it is appropriate for heads to roll when it turns out there are rampant problems at lower levels of an organization.

We should be going after hedge fund managers who set up incentive structures that reward certain kinds of pathological behavior. We should be going after the commanders who create circumstances in their own military units where there are going to be enticements and no constraints on people acting in ways that constitute atrocities.

Does that dismiss responsibility for the lower ranking?

I don't think of responsibility as zero sum. I think both people — commander and soldier, manager and employee — can be responsible. There will be cases where it might well turn out that individuals are in conditions that are so damaging to their ability to self-govern, where the conditions are so deplorably difficult for self-regulation that they are not actually responsible for what they do.

To see why social context might undermine responsibility, consider someone who is subject to strong social deprivation, and is then put in a circumstance where the only way out is to cheat on an exam.

Contrast that case with another person subject to no social deprivation who has a wealth of advantages. There is no special advantage to passing this exam. It's just one opportunity of many.

It may be that both agents have, in one sense, the same ability to resist the temptation to cheat, but it also matters that the possibility of cheating has a wildly different significance for each of them because of their larger contexts. Other things equal, there is less culpability when one's context makes cheating overwhelmingly attractive.

How do you pat someone on the back for that? Should we reward someone for not caving into temptation?

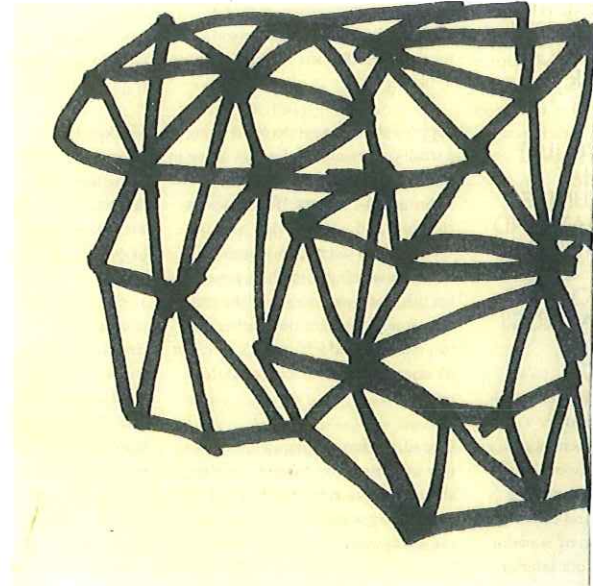
This is a great question. Praise and reward are surprisingly mysterious things. A lot of philosophers, and I count myself among them, tend to focus a lot on blame, to the exclusion of praise. I don't think that it's just that we're pathological, sadistic people, looking to talk about misery.

I think some of it is the human disposition to being sensitive to what people call cheater detection. We are really attuned to violations of social norms.

FAILURES TO CAPTURE CHEATING ARE MORE COSTLY TO US THAN FAILURES TO OFFER DUE PRAISE.

Praise is funny. Teaching kids about morality is a special case, so let's put that to the side. In ordinary life, praise is the kind of thing that's not offered whenever people have met the bare requirements of their obligations.

Instead, we tend to reserve it for cases when someone has exceeded some reasonable demand in a way we approve of. So praise is appropriate not when somebody does what they were supposed to do. Praise is appropriate when someone does more than what



they were supposed to do. If that's right, then failing to cheat is not a particularly praiseworthy act. It might be praiseworthy if the temptation to cheat was massive and you don't cheat — as in the deprivation case — but the

praise would presumably be mitigated, because there is still this norm against cheating.

THE OTHER INTERESTING THING ABOUT PRAISE IS THAT, BY AND LARGE, WE ARE NOT VERY GOOD AT IT.

FIRST OF ALL, WE JUST DON'T HAVE THAT MANY OPPORTUNITIES. ASIDE FROM THAT, AND ESPECIALLY AMONG ADULTS, PRAISE IS VERY DIFFICULT TO DO WITHOUT COMING ACROSS AS A JACKASS.

Why? Well because there is the standing of peers. It looks weird for a subordinate to praise a social superior unless the person is just boot-licking, at which point that undermines the integrity of the praise. And to praise someone from a position of superior authority reminds us of our inferior position, which often isn't pleasurable to adults, especially those raised with egalitarian instincts.

In many ways it is easier for peers to praise each other, but it's difficult to do it in a way that doesn't come off as condescending. So between adults, praise is actually very difficult to do, unless people know each other well enough to know whether the praise is genuine or if it's instead ill-motivated

praise.

Versus blame.

Blame is the easiest thing in the world for us to do. We can all do it!

Why is that?

I think it's rooted in an interest in cheater detection. In a lot of cases, cheater detection ends up being very important to us. Violations of the social contract can carry really big costs, so that's part of the reason why we are especially alert to opportunities to blame.

Suppose there is a group of us living in a small community, and we are living in subsistence conditions. We barely make it through the winters. If someone is cheating on the food supply, that is very costly to our group, whereas if someone is doing something generous for the group, and we get to the end of the winter with three days rather than two days of food left... that's nice, but it's not a life or death kind of difference.

One of the funny things about us is that we seem to be disposed to blame all over the place. We need to be taught to gradually scale back which parts of the world we are in the business of blaming. You sometimes see a little kid who will fall down and will instantly turn around, and you can see the kid actively search for a blame target to account for the fact that they just fell down. The nearest person around them may end up being the target of the blame.

It's not uncommon. When you think of the way in which some animistic religions operate, these are worlds in which the culpability for all kinds of

natural phenomenon is getting passed off onto fictional agents. In those societies, we tend to ascribe agency and culpability all over the place and gradually get talked out of it as we learn. No, no, no, we learn to say, you are making a mistake in thinking the spirit of the wind is the explanation for why you got blown off course.

This is one way to get at the core of the free will question. The worry is that if we see the world as stripped of supernatural agency — in the natural world and ourselves — and it is all just physical causes, then where is the responsibility?

The key, I think, is to acknowledge that we are part of the physical causal fabric of the world, but to also recognize that we are the part capable of acting with malicious intent. That's the hook of praise and blame.

So the US has a philosophical mantra about pursuing your destiny, your happiness, and there is a free will association with that. How is it that a nation that is built so much on the idea that you are free, that you are the life force that drives your destiny and happiness, how is it that we are in this state right now of not taking responsibility?

Part of what you are picking up on is the remarkable dovetailing of distinct notions of freedom in the US. We have an individualistic conception of freedom combined with a fairly widespread commitment to the idea that we have political freedom here and the freedom of having alternative possibilities available to us, where it is open to us to do otherwise at any meaningful moment of decision. It's a super freedom-enthusiastic society in some sense.

The reason I am stuck is that your research

is defining free will in the context of moral responsibility. Free will is constrained by moral responsibility, and maybe that's why there is so much visible moral deterioration.

I agree with part of that. Free will has been stretched to cover too many things. People have meant different things by the term. There is a way in which I am identifying one domain of uses for it and being neutral about other uses of the term, even though there are reasonable, ordinary uses of the term that aren't related to moral responsibility. That's because we have one word that is picking out several different things.

As for a loss of a sense of responsibility, I am skeptical that your kind of characterization about our society is accurate. Plato is alleged to have said that folks have always been complaining about the loss of moral standards over time.

So either we are committed to the view that there was greater moral perfection in the distant past and we've been gradually degrading over time, or something else is going on. I think the something else going on is that moral norms shift over time, so things that looked like reprehensible behavior to one generation looks like perfectly reasonable behavior to the next generation.

As the norms shift, the things that count as appropriate shift, and any time there is any slippage at all, we interpret the slippage away from our moral commitments as moral deterioration. That's the sense in which I'm skeptical of whether or not there is some kind of global deterioration.

There has always been corruption. There have always been failures of

accountability. There has always been deplorable behavior. The difference is that we now live in a society where we can hear about the deplorable behavior of people happening thousands of miles away from us, and we hear about it instantaneously.

This signals to us, oh my gosh, there has been degradation in the sphere of moral concern in some way. The difference is that the number of people in the sphere is much larger and the ability to relay the information is stronger than it was in previous generations.

By contrast, 150 years ago you ordinarily didn't hear about the terrible behavior that was happening even 100 miles away, so it wasn't part of anybody's sense of how the moral universe was changing or being threatened. I have not been convinced of the global decline of accountability, the rule of law, and so on.

That's good news to me since you've been studying this for ten years. I guess I'm more pessimistic. It is my own judgment that in this country, you do what you want and who cares about the planet or other people. I'm glad that you don't find that this late-capitalism time that we are living in right now is so much more depressing than other times.

IT IS ARGUABLY MORE DEPRESSING THAN OTHER TIMES, BUT IT'S NOT BECAUSE OF DETERIORATION IN THE MORAL SPHERE, THAT PEOPLE ARE BEHAVING WORSE. IT IS DEPRESSING BECAUSE OUR CAPACITY FOR DESTRUCTION HAS GONE UP.

To put it differently, our circumstances are very different than they once were. A five year old in a padded room can only do so much damage. A five year old sitting behind a tank can do way more damage. A five year old with his finger on a nuclear weapon can do even more damage. In the world we have built for ourselves, we are no longer five year olds in a padded room. We are five year olds with our hands on tools that are more catastrophic than ever.

There are good reasons to be concerned about our future as a species and the shape of the planet. There is quite a bit to be worried about and depressed about, but it's not because we're worse people. *