



Opinion Science Podcast

Hosted by Andy Luttrell

Liking What Helps You with David Melnikoff

March 15th, 2021

Web: <http://opinionsciencepodcast.com/>

Twitter: [@OpinionSciPod](https://twitter.com/OpinionSciPod)

Facebook: [OpinionSciPod](https://www.facebook.com/OpinionSciPod)

Andy Luttrell:

How do you feel about donuts? If you're a good and decent person, you said you like them. I mean, imagine a donut. Your brain is just calling out to it, "I love you." Okay, now imagine you've been challenged to eat a dozen donuts in one sitting. You accept the challenge, obviously, because you know, free donuts. But once you're done, belly full of donuts, if I ask again, "How do you feel about donuts?" You might be like, "Eh. I'm not feeling any special yearning for them." After all, your need for donuts has been satisfied. As much as you said you liked them, you don't need them anymore.

You're listening to Opinion Science, the show about our opinions, where they come from, and how they change. I'm Andy Luttrell and today on the show we'll explore this idea that our likes and dislikes are tied to the goals we have in the moment. We like stuff that helps us reach our goals. When I'm hungry, donuts sound great, because they'll help me reduce the hunger. But when I'm full, the allure wears off. Our guest today takes this idea and runs with it. I was excited to talk to David Melnikoff. He's a postdoctoral scholar at Northeastern University and he studies how what we need in the moment can shape our evaluations of things.

In our conversation, we'll talk about what psychologists call attitudes, how those are connected to goals, and why some people sometimes actually prefer immoral people.

I sort of thought that to set the stage on the show, I call it Opinion Science. I often use the word opinion when plenty of us know that that is just a replacement for the word attitude, but attitude means something sort of different in common understanding, but I kind of think it's hard to talk about your work by subbing opinion in for attitude. So, I wondered if you would just sort of set the stage by saying what attitude means to psychologists? How have we often thought about that concept? And why is it important to study?

David Melnikoff:

Yeah, so the view that I subscribe to is this view that attitude really picks out three different things. This is... The jargon is the tripartite model of attitudes, but it really separates cognitive components of attitudes, affective components of attitudes, and even behavioral components. I think the idea that attitudes are behaviors is the most... the biggest stretch for a non-psychologist. And cognitive components of attitudes would be the more intuitive ones, so that is really an opinion. A cognitive

attitude could be something like a belief. So, stereotypes would be examples of kind of the cognitive dimension of attitudes. The belief that a group tends to have a certain trait, policy opinions, I think this policy is effective or I support it, or I think it's ineffective and I don't support it.

Affective attitudes are immediate hedonic responses to stimuli in your environment, or even more abstract things than concrete stimuli. You can have an affective response to an idea or an abstract concept. And sometimes your beliefs about an object don't necessarily match your affective response to it, so I can believe that something is good and actually feel negatively towards it. Or I can believe that something is bad and feel positively towards it. And so, of course because these things can dissociate, it's useful to distinguish cognitive and affective forms of attitudes.

Andy Luttrell:

But at the end of the day, both are about good and bad, right? So, the core of it would you say still is about positive negative?

David Melnikoff:

Yeah.

Andy Luttrell:

Okay.

David Melnikoff:

Especially for me. Yeah. I think if people want to go beyond positive and negative and say something like the belief that David is a psychologist, and call that an attitude, I think that would be fine. I just think we have to be careful about marking that as like what you're talking about, and so in this context of that, what attitude means, and that's fine. But for me, it's really about good or bad. You can mean that in a cognitive or an affective way. And the behavioral version of good or bad, good, and bad on the behavioral dimension usually corresponds to approach or avoid.

So, usually you expect someone to approach something that they think of as good and avoid something that they think of as bad. So, when you talk about a behavioral, a positive behavioral attitude, you're suggesting that someone has an approach tendency towards something. If they have a negative behavior attitude, you'd be suggesting an avoidance tendency towards that object.

Andy Luttrell:

Were attitudes always sort of in your field of vision as you got started in the world of psychology? Or was this one of those things where it's like this guy that I want to work with, or this person I want to work with, seems cool, and then turns out they study this thing called attitudes?

David Melnikoff:

Attitudes were always... Yeah, so my story, so I was always interested in attitudes. Originally, I was actually... So, as an undergrad I did single cell recordings in mice, in mouse hippocampus, so I was interested in memory. And as an undergrad, I sort of learned about this thing called Hebbian learning, and that's that neurons that fire together wire together, and that's how memories form.

I'm not saying that that is just how memories form, but that's kind of... As an undergrad, that's what I learned.

And so, I thought that I was studying associations between, that the way to study memory and learning is with these single cell recording techniques. And then I read this paper about this thing called the IAT, which was presented as kind of this way of tapping the same sort of associations that I was looking at in mice, but these were super important associations in humans. These were associations between racial groups and negative evaluations, and essentially I kind of thought that this was getting at the same thing, just at a higher level, and I don't have to take mice and extract their brains, and use a microtome... I mean, there's a lot of work that goes into... Not that there's not a lot of work with doing an IAT experiment, but there's a lot more work with doing single cell recordings. And you know, comparing racial attitudes to like associations between objects and smells or something like that, for me it was no contest. I had to go study these kinds of associations and use these tools.

But yeah, I mean, I thought that I was doing essentially what I was doing in the wet lab in a dry lab, because the IAT was fundamentally getting at the same kind of mental structure.

Andy Luttrell:

Given the choice between what you were doing and what you ended up doing, I would also choose what you ended up doing. And it kind of teases at... Well, so that contrast makes sense, where the learning and memory part sets us up to think of like, "Oh, the things that I've come to like, and value are the things that I've learned over time are good. And the things that I've come to have a distaste for, and reject are the things that I've come to learn are bad."

You changed the game from that perspective to something quite different, which is that like kind of regardless of what you've come to learn, other stuff matters for whether you see this thing as good or bad. And actually, so maybe before we even get there, I wanted to ask if you find it useful to distinguish between attitudes and evaluations. So, from my perspective, it tends to be helpful to say like there's an attitude, and sort of the imagery that I get is like a bookshelf in your head of all the stuff that you've ever encountered in your life, and there's a little tag on each thing, and the tag says good or bad, right?

And sometimes that string is tied really tight to the thing on the bookshelf. You go, "Oh, coffee? Love it. Tightly tie the good string to that cup of coffee in my brain." But other stuff, you go, "Ah, maybe there's a tag on it, but it's not tied that closely." But in the moment, yeah, those things that live in my brain, those I might call my attitudes. But in the moment when you say, "Do you want a cup of coffee right now?" In this moment, how I feel about coffee might be a little bit different than the thing that lives in the bookshelf in my head, and that I would call the evaluation. And so, I wonder whether you would say that you see those two things as different, because sometimes you talk about attitude change, and I wonder whether that means like, "Oh, how I see this in the moment is changing," versus like, "No, the bookshelf is changing. I'm changing out the tag on the bookshelf."

David Melnikoff:

I like to be a little... So, I totally share your intuition that there's the bookshelf in your brain, on the shelves are a bunch of attitudes. Evaluations reflect the in the moment affective response or cognitive response, which for one reason or another may not be exactly what you pulled off the shelf. So, somehow there's some noise in the process between... I don't know how far this metaphor goes, but pulling the book off the shelf, and opening it, and then saying good or bad.

And I like to be a little provocative and use both evaluation and attitude to mean the same thing, which is just the in the moment affective response or cognitive response, and to just get rid of the bookshelf all together. I'm skeptical that there's any kind of bookshelf-like structure when it comes to evaluation, and therefore setting aside one concept, like one attitude concept here, and one evaluation concept here doesn't make a lot of sense to me, given the data. And I think that we can just use both of these terms.

I would say either get rid of the word attitude, like if we're really committed to attitude being books on a shelf in our brain, then just get rid of it. But if we want to keep it, then let's let it mean the same thing as evaluation.

Andy Luttrell:

So, I mean, okay, so that sets the stage, and so you say that how did the data fit these two versions of the world? I think we could talk maybe about the more recent paper that you had come out on motivation and attitude change, where in particular I think the Hitler study very well encapsulates and provocatively encapsulates what you're suggesting. So, maybe you could sort of give a little bit of background on where this notion comes from, how you tested it, and what those results mean.

David Melnikoff:

Yeah. So, to set the stage for the Hitler study, I think it's important to I guess briefly summarize the two things that I think of as contributing to these in the moment evaluations. Because that... I do think that there are at least two kinds of computations underlying people's in the moment evaluations, and the Hitler study really looked at one. And I think it's just important to be clear about that.

So, I think, and not just me, a lot of people think that when you first see an object, immediately upon seeing it, unintentionally you will appraise that object as instrumental or not instrumental to your current needs and desires. And there's magnitude involved, so how instrumental and non-instrumental is it? So, I'm using the word appraisal because this idea of instrumentality and goal conduciveness is tied to appraisal theories of emotion, but attitudes researchers have used it for a long time, dating back to Kurt Lewin, who really emphasized the concept of instrumentality in goal conduciveness.

In addition to appraising instrumentality, you appraise something called action valence. So, do I intend to act positively or negatively towards this object? Whereby positive action I mean approach or if it's a minded entity, or an entity that you think can feel, or that has desires itself, do you intend to help it? Whereas if you intend to harm that object or avoid it, that would correspond to a negative action.

So, immediately upon seeing an object, you appraise action valence, so I intend to help, harm, approach, or avoid, and is this instrumental or not instrumental to my current goals? And your automatic evaluative responses towards the object is basically a readout of instrumentality and action valence. The more positive the action valence and the more instrumental the object, the more positive the evaluation and vice versa for negative action valence and non-instrumentality.

So, in the Hitler study, what we did... So, I think it's pretty uncontroversial that instrumentality is important for evaluation. Action valence is a little weird, and so what we did in this paper was basically hold instrumentality constant while manipulating action valence. And the way we did that was in most of the experiments, and in the Hitler experiment, what we did was create a game where people intended to participate in this simulation or a game which involved being an attorney in a trial, and the evaluations people made were towards the defendant, and they made those evaluations while expecting to play the role of prosecuting attorney or defense attorney.

So, if you're the defense attorney, you have positive intentions towards the defendant. You want to help them. If you are a prosecuting attorney, you have negative intentions towards the defendant. You want to harm them. So, we asked two things. One, does intention valence in this experiment alter implicit automatic evaluative responses? And crucially, in all of our experiments we included... I think we called it like attention turn off manipulation, where you'd be assigned to the role, and then right before you'd automatically evaluate the defendant, we'd say, "Oh, just kidding. You're not gonna be playing the game." We didn't say just kidding, but we have a cover story, and we said, "You're not gonna be playing the game anymore."

And that was because we worried that by saying, "Oh, you're gonna be the prosecuting attorney. You're gonna be the defense attorney." People would take that as information that, oh, they assigned me to defense attorney because that person's good, or they assigned me to the prosecuting attorney role because that person's bad. And of course, if they did draw that inference from role assignment, that inference is still valid having said... The cover story was, "Your operating system doesn't support the game, so you can't play it." So, the inference still holds, so they would still believe whatever they believed before the cancellation after.

Okay, and so the basic finding is that if people were in the prosecuting attorney role, that made their automatic evaluations more negative. If they were assigned to the defense attorney role, that made their automatic evaluations more positive. And this effect was completely eliminated by telling people that they would not be performing the role, or the helpful or harmful action anymore. They never actually engaged in any helpful or harmful behavior, but what mattered is whether you intended to help or harm the target.

We wanted to see if this worked with even really extreme attitude objects, so in one experiment the defendant was Adolf Hitler. And just in case anyone didn't know, like we had this very vivid reminder of all of the atrocities of the Holocaust to start the experiment with, and all the atrocities were listed under a picture of Adolf Hitler. And so, after reminding people of all of this, we assigned people to be the prosecuting attorney or defense attorney, and I'm doing air quotes, but Hitler war crimes trial. I mean, admittedly this was kind of a weird situation, but it's just a game and they're gonna have to be prosecuting or defense attorney, and again, they don't do anything.

They've just been assigned to the role, defense attorney or prosecuting attorney, and then they find out that they do or do not actually have to go through with the role.

And what we found was that being assigned to the role of defense attorney increased implicit positivity towards Hitler, did so just as much as it did towards novel attitude objects, which we used in our other experiments, and in fact the control stimuli in this experiment were just some smiling white men, and the implicit evaluations were not significantly more negative towards Hitler than the control stimuli. And this completely went away, and as soon as we told people, "Oh, you don't have to go through with this anymore," the effect goes away.

Andy Luttrell:

So, just to recap, just to make sure that it's clear, so even when people ordinarily would say, "This person's a monster and undeniably so," if you are tasked in this moment with helping this person, even if you don't ever do anything to help this person, just knowing that it's sort of a goal you have automatically makes people evaluate this guy less negatively than they had before. To the point that they're not seeing him as really any different, you're saying, as sort of a range of other people.

David Melnikoff:

Yep. Yep.

Andy Luttrell:

And as soon as you go, "Oh, wait. Game's broken. You actually don't have to do this." People go right back to being like, "Yeah, no. This guy's a monster."

David Melnikoff:

Yes. Exactly. Exactly. And when we ask people to just tell us, is Hitler good or bad? I mean, there was no variant. Everyone used the far end of the scale. As in the most negative end of the scale. So, it wasn't that people for some reason... There was an interesting review where someone suggested that maybe given current events, people just don't think of Hitler as being that negative. The self-report measures seem to suggest that people do insist that he is in fact one of the most evil people in history.

Andy Luttrell:

Yeah, so the difference between the automatic and the explicit seems important, and I know that there's a little bit of inconsistency across the studies you did in that line of work I think, but by and large... Well, no. Not in that line of work. By and large you're finding that it's those automatic things, like the first... As soon as you see this face, your brain gives him a little more credit so long as you have the goal to help him. But ask me what my opinion and I can honestly and openly say, "No, I don't think this person is a good person." And so, these are showing up only on those first automatic reactions, is that right?

David Melnikoff:

So, you mentioned some inconsistency, and I would say that the key difference between where goals really seem to matter, it's not implicit or explicit. It's really affective or cognitive. So, in my reporting, are we measuring how people feel towards this person? Or are we measuring whether

people believe that the attitude object fits their definition of what a good or bad thing is? So, for people, like when people say good or bad, they usually mean morally good or morally bad. If I just ask you, “Is so-and-so a good person? Is so-and-so a bad person?” What they’re gonna tell you is is this person a moral person or is this person an immoral person, basically. And when we ask a question like that, are they good or bad, manipulating intention valence didn’t really matter. It maybe mattered a tiny bit, or it didn’t matter at all. And even manipulating instrumentality matters either anywhere from a little bit to not at all.

However, if we measure your automatic affective responses with an AMP, or if we measure your non-automatic, intentionally generated affective responses, or intentionally reported affective responses with feelings thermometers, then you’re capturing... So, a feelings thermometer would be a question like how warm or cold do you feel towards this person? Then you’re not really getting information about the goodness or badness of the person. You’re really... Almost all of the variance is accounted for by intention valence and instrumentality.

Yeah, so I can say the cognitive-affective distinction is really key here, less so than the implicit-explicit distinction. In terms of the dominance of motivational variables. Yeah.

Andy Luttrell:

It does seem like that there was some setup that it’s the automatic, or affective, or emotional, regardless, but is there... Why, I guess is my main question, like what is it about this goal that’s really shaping how I’m feeling in the moment? Even if I’m able to say, “No, I don’t think this is a good person, but as long as I need to help them, I feel good.” So, there’s like an evolutionarily adaptive explanation that seems appealing, where you go, “Well, we would need a system that encourages us to pursue goals.” Right? So, if we’re gonna be effective in the world in pursuing the goals that we have, we might need to shake predispositions that would have held us back from pursuing those goals. Is that kind of what it seems like is happening? Or does it seem like something else?

David Melnikoff:

That’s interesting. So, I don’t... I tend not to think of it, so I think everything you said is perfectly reasonable. I don’t come at it from an evolutionary point of view. Basically, what I think is happening is that people are constantly predicting how they’re going to feel. They’re constantly making affective predictions. And people feel... It feels good when you attain a goal. It feels bad when you don’t. So, when I see a stimulus, if that stimulus is instrumental to my goals, I’m anticipating that affective, that hedonic satisfaction of goal completion. If I see something that is counterproductive to my goals, I see it, or it doesn’t have to be, but you can hear it, however you perceive it, you will automatically anticipate that negative hedonic response of goal failure.

Applying the same logic to action valence, when I see an object, I don’t just... So, actually, let me start explaining this by talking not about affect, but just about motor responses. So, if you see a cup of coffee, for example, you don’t just see a cup of coffee. You also bring to mind motor programs, like grasping the cup of coffee, to prepare for action. You can think of affective responses as a kind of action. This is kind of like a visceral action or an interoceptive action. You are, because affect isn’t just something that happens to you. It happens that you can try to cultivate to achieve your goals.

You know, if I'm trying to beat someone in a boxing match, it's good that I feel negatively towards them. If I'm trying to make a friend, it's good that I feel positively towards them. So, just like I might automatically bring to mind appropriate actions, like grasping a cup, I automatically bring to mind appropriate affective responses given my current goals. Again, it's very goal centric. So, I might automatically bring to mind negative affect when my goal is to do harm to someone. And I might automatically bring to mind positive affect if my goal is to help someone.

Why it's functional to do something like that I think is pretty clear and that's kind of how I think of it.

Andy Luttrell:

It's sort of the feeling of achieving the goal is making me think in those trial lawyer scenarios. If you go, "My goal is to win this trial," you go like, "I feel good whenever I develop a case that's gonna help me win." So, my eye is on the prize, in other words.

David Melnikoff:

Exactly.

Andy Luttrell:

And I can sort of distract... In some ways, whatever I had thought of this person before doesn't matter, right? So long as I'm trying to pursue this one goal. And it's gonna feel so great when I win this case. So, anything that's gonna help me win feels good, and it reminds me... Just on Twitter yesterday, people were talking about this, and it was reminding me of when I teach prejudice in class, and you go, "Oh, what a great study this is." And you go, "Well, yes, the study did find that people have these awful feelings about each other." So, the finding is not great, but the study serves a purpose in teaching this. Or if you're developing a case and you're like, "I really want to prove that this firm or this company is discriminating," and you find evidence that like, "Oh, this is great!" You go, "Well, it's not great, but it's great," because it's helping me do this thing.

It seems to me like that is a very similar... That is like a natural experience that is like what you're talking about. Is that right?

David Melnikoff:

Exactly. And we have this... Well, I said I don't want to say bias, but I'm gonna say it, because I can't think of another word, but we have a bias towards assuming that... Say in an AMP, and just for if someone hasn't worked with an AMP, the way that AMP works is you see a picture of a stimulus. Usually, a picture could be a word, but in our case you would see a picture of Hitler appear on your screen really quickly, and then a neutral image would appear, and you have to rate the visual pleasantness of the neutral image. And the finding is that the response to the neutral image is influenced by your affective response to the prime image, and so the proportion of positive to negative responses to the target images are taken as a proxy for your affective responses to the primes.

Now, it's assumed that if you have a positive implicit evaluation of Hitler, that the positivity is about Hitler in a deep way, like you like Hitler. As opposed to Hitler's picture just reminding you of something that is affectively positive. And there's no reason to think... And the way it's set up, and this is true of any implicit attitude measure, there's no reason to think that the affective response or representation is about the prime. It was caused by the prime, but it doesn't necessarily need to be about the prime. And when we think about that, a lot of things about... When we think about it that way, I think a lot of these findings make a lot more sense.

So, why would it ever be the case that not only did the positive intention or goal make implicit evaluations of Hitler more positive, but did so just as much as it did for everyone else? So, just like a neutral image. Well, if you're just imagining that you're just adding the positivity of the goal to the mix. Then it makes... You should just expect this additive effect. It's not that the effect should be smaller with Hitler because he's Hitler and he's horrible. You should be able to predict the magnitude of the effect by just asking people how good you think you'd feel about achieving this goal? And that's gonna be... That should be what predicts the responses to the primes, because that's what they're thinking of.

And that thought is simply occasioned by the prime appearing on your screen.

Andy Luttrell:

The one thing that I still wonder is as a person who studies ambivalence a lot, a lot of this reminds me of this question of like, "Sure, I may be turning the dial down on my negativity because this person is gonna help me reach some goal." But at the end of the day, the attorney who goes, "I want to win," who still might go like, "But am I winning by defending someone that I don't actually think is that great?" We would ordinarily call that an ambivalent experience, right? You go, "I have some reasons to value this person and I have other reasons to think that this person is no good, or this object, this coffee, this whatever." Anything that's like motivationally relevant, it helps me reach a goal, but still, the bookcase, I'm gonna go back to it. Bookcase tag still is a negative one or a positive one.

And we know that people can report those feelings of ambivalence. If I were to guess, I would think that your reaction is that that is all one step beyond that first affective reaction. But I assume maybe you've thought about this notion of ambivalence and how it plays out in the work that you do.

David Melnikoff:

Yes. I do think it's one step beyond. I mean, there's no question that people experience ambivalence, are frequently ambivalent. All I would say is that I don't think the ambivalence is reflective of a conflict between the contents of a book on a shelf in your brain and the current situation. I think current situations are often ambivalent. That is, I have multiple goals, I have multiple intentions, and that can lead me to feel ambivalent. So, say for example the thing that's instrumental to my current goal is that this person is a really nasty person, and so I feel positively towards them because they're nasty. You know, I mean I could... That kind of just fits a kind of a definition of an ambivalent attitude. I may not feel ambivalent, but I kind of know that there's something different about this response that I'm having as compared to responses where I feel positively towards good people.

But I don't think you would need to posit anything deeper about the affective experience in that case. But I do think that you have deeply ambivalent affective responses to things. I would just say that those come from the multitude of goals that people pursue at any one time, and the fact that stimuli can be conducive to some and not to others.

Andy Luttrell:

Which kind of I think transitions us to the morality work that you've done, because that is also what strikes me as ambivalent, where you can be like, "Sure, someone who's nasty and who I'm willing to say is immoral is useful to me, and so serves on goal I have, but I might have other goals to be like a good, virtuous person, and this person conflicts with that goal." So, I wonder, so the work that you've done in this area is a clear response to prior work on what's been called morality dominance or goes maybe by other names. But if you could sort of set us up by kind of characterizing as you see it the claims that have been made before about what is important about moral character in people, and then sort of how you've considered why that might not always be the case.

David Melnikoff:

So, the previous claim is super simple. I think this is essentially a direct quote, but morality in others is always positive, and immorality in others is always negative. So, the more moral I think someone is, the more positively I'm gonna evaluate them. The more immoral I think they are, the more negatively I'm gonna evaluate them. Now, a lot of times people hear that, and they think, "Oh, so that claim is that people always like moral people and always dislike immoral people." No, that's not... I mean, to be fair... That's not the claim. That's obviously not true.

So, first of all, a lot of times when people say morality in this case, it's morality in the eye of the evaluator. So, clearly there are immoral people who are liked. Do the people who are doing the liking think that that person's immoral? Do they agree? And they might be confused, or they have just very different moral standards, but for whatever reason they might think this immoral person is actually moral, and therefore like them. So, it's not... The claim is not that immoral people are never liked. Clearly, they are.

It's clearly also the case that I can believe that someone is immoral, I can find that bad, but I can still like them overall because they have other qualities that are good. They're physically attractive. They're whatever. They're competent. They're whatever it is. They can have other compensating qualities, which makes them positive on the whole, but immorality would still remain a negative feature. The morality dominance claim is that the more I think you possess qualities that I view as immoral, the more negative I'm going to see you as being.

So, what I think of as immoral is gonna have a negative influence on my evaluation of you. What I think of as being moral is gonna have a positive influence.

Andy Luttrell:

Meaning, so if I deem you to have immoral traits, I should dislike you, right?

David Melnikoff:

More than you would if you deemed me to have moral traits.

Andy Luttrell:

Have moral traits. Yeah.

David Melnikoff:

Exactly. Exactly. Which is different. So, it's well known for other trait dimensions that's not the case. So, with competence, someone who wants to help me, the more competent I think they are, the more I'll like them. Someone who wants to harm me, the more competent they are, the more I will dislike them. So, there the valence of competence is conditional. It depends on what else... on the situation. Competence itself can be good or bad. But morality dominance hypothesis is that morality is special. These other dimensions are conditional, but morality is not. Morality is always positive.

Andy Luttrell:

And so, why might we say that's not always the case?

David Melnikoff:

So, I'm just committed to that your affective responses at least to stimuli, so the affective dimension of your attitudes is a function of instrumentality and action valence. And it's just clearly the case that morality can be non-instrumental to your current goals, and of course you can have positive or negative intentions to a moral person, and someone you think is moral, and positive intentions towards someone you think is immoral. So, if I'm right that those are just the things that determine your affective responses to stimuli, there's nothing... There's no room for a morality dominance hypothesis. Morality is just like everything else. If it's instrumental, it's good. If it's not instrumental, it's bad. Affectively speaking.

Andy Luttrell:

So, that means nothing can be dominant. Would you go so far... I mean, other than instrumentality. There's no feature of any object, person, or issue that is dominantly... Dominant is sort of a weird word, because it might suggest that by default, averaging across all the experiences maybe most of the time, the dominant responses for it to be a valued thing or a devalued thing. But it's more of a black and white, universal claim, basically saying always this is good. And I think... Are you saying that this motivation lens, this goal lens would say, "Nothing is ever always good or bad." It will depend on whether it's good for you or bad for you.

David Melnikoff:

Exactly. And the other key, so this is the move that some people use when... Well, so let me just say that when someone who endorses something like a morality dominance hypothesis is confronted maybe with the importance of the motivational state, then the move is to say, "Well, usually it's positive in most situations." But that's a very different kind of claim. I mean, yes, it's usually positive, but that's because it's usually goal conducive. It's psychologically... What's happening is totally different if on the one hand you're positing that your mind is evolved to pick out moral features and immoral features, and to convert them into positive and negative responses, versus if there's nothing like that happening at all, your mind is designed to compute the instrumentality of the object regardless of whether it's moral or not, and your intentions toward

the object, regardless of whether it's moral or not, and then convert those appraisals into positive or negative responses.

So, yeah, it's certainly true that on average morality is good. But that tells you very little about the mechanism underlying evaluations of moral character. And I think that the actual mechanisms are fundamentally different than what is implied at the very least and often explicitly stipulated under views like the morality dominance hypothesis.

Andy Luttrell:

So, it sort of seems like you could sort of find a middle ground that kind of captures the spirit of both, kind of like what you're saying, which is like usually people have a goal for which moral others are conducive, and so by and large, most of the time people will prefer people that they deem moral over people they deem immoral.

David Melnikoff:

Absolutely. Yeah.

Andy Luttrell:

But it's just not the same as saying like, "Well, people always and forever have a goal for which moral others are conducive." And so, to sort of shatter that version of the claim, what is a circumstance where the goal makes what you want pretty different from maybe what they would say is ordinarily your goal?

David Melnikoff:

So, April Bailey and I came up with a series of such scenarios and in one scenario, which will be familiar because I've already talked about this lawyer game, but one scenario is you are a prosecuting or defense attorney, and you are selecting people to be on a jury. Now, if you're a prosecuting attorney and your goal is to get a guilty verdict, you might want people who are merciless, generally considered to be an immoral character trait. If you are a defense attorney, you want someone who's very merciful, perhaps, generally considered a moral character trait.

So, mercifulness versus mercilessness can be made more or less goal conducive depending on whether you want this person to provide a verdict of guilty or not guilty or have a very harsh sentence versus a very lenient sentence, something like that. And that's what we looked at. We created a situation like that. Participants evaluated novel target people who were simply described as merciful or merciless from the perspective of someone who was going to be selecting members of a jury either as a prosecuting attorney or a defense attorney, and yeah, and so we looked at whether people in fact, whether for example prosecuting attorneys in fact found the merciless person to be more conducive than the merciful person. And they did. And vice versa for the defense attorney.

And it was really important that we confirmed that people didn't just convince themselves that, "Oh, mercilessness is a moral thing in this case." So, we asked people like, "How moral are these individuals?" And regardless of role, everyone agreed that the merciful person is morally good, and the merciless person is morally bad. In addition to being warm, everyone said that they would rather be friends with the merciful person. They say they would rather be more similar to the

merciful person. But when it came to associating the moral and merciless person with positivity or negativity, and when it came to feeling thermometers, so explicit reports of feelings towards these people, what really mattered was whether the individual was goal conducive or not. There were positive implicit evaluations and explicit evaluations of the merciful person, but only if the merciful person was instrumental. When the merciless person was instrumental, evaluations of the merciless person were positive, and of the merciful person, they were negative.

And so, that was one of four experiments that took this approach to independently manipulating morality and instrumentality. Critically, we did it in the minds of our participants, so from their perspective, these were independent things and what mattered for their affective responses was the instrumentality. Crucially, in a second experiment, the trait that we manipulated was trustworthiness. So, this is the most important experiment in my mind, because even among the moral traits, trustworthiness is supposed to be the dominant of the dominant. So, this is the one that you read out from facial features in less than 100 milliseconds from Todorov's work. This is the one that infants seem to pick up on within months. And they don't just pick up on it, but they prefer trustworthy to untrustworthy.

So, if there was a candidate for like a really built-in stimulus response mechanism that just converts some feature into a certain affective response, you'd think trustworthiness would be it. And not only did instrumentality dominate, but when the untrustworthy person in this experiment was the instrumental one, they were liked as much as the trustworthy person when the trustworthy person was dominant. So, there was no effect of trustworthiness above and beyond the effect of instrumentality. Because what you might say is that both of these things are happening at once. There's an independent effective morality and then there's an independent effect of instrumentality, and those two act together to give you your evaluative response.

In that experiment, the result seems to suggest that you don't really need to appeal to any independent effect of morality. Just like all you needed to know in this experiment was whether the person was instrumental or not. And so, that's what really kind of convinced me that it's a reasonable... I guess what I'm arguing isn't necessarily... I mean, I could definitely be wrong and there's a lot of work that needs to be done, but I think it's a reasonable argument to make. I think it's something that we need to consider, that really all that matters for these affective responses is instrumentality and action valence.

Andy Luttrell:

You could sort of see in terms of the no special extra bonus of being a trustworthy person... To the point earlier about there can be multiple goals at a time, you can imagine someone who's just like, "I value morality over everything." Right? That's just... My goal, the most important goal to me is that we pursue virtuous ends. And for someone like that you'd say, "Well, sure. This dishonest person might help you win the game or whatever," but you don't care. You go, "That is not... I know you told me that's my goal. That's not my goal."

David Melnikoff:

Yep.

Andy Luttrell:

And if those co-occur, you could be like, “I get I want to win the game, so I’m gonna have to like this person, but I still...” and here’s where the ambivalence might come in. You go, “I still am so committed to valuing honesty that that is uncomfortable or that is a real knock against liking this person.”

David Melnikoff:

Yeah. I’m an adult human being and I can walk and chew gum. I can have more than one goal at once. You know, there’s no problem with that. And because that’s true, it’s not... Well, what I was gonna say is that because it’s true, obviously true that you can have multiple goals at once, if we were to have found... I think this is kind of what you were suggesting. If we were to have found that there was a trustworthiness bonus, we could have explained that by saying, “It’s still appealing only to goals.” Well, you had this other goal, and it was not instrumental to that, so en masse, the overall level of instrumentality was still higher for the trustworthy person when they were instrumental to the game goal than the untrustworthy person.

But that allows me a whole lot of degrees of freedom to always say, like no matter what we find, I can say, “Oh, well, there’s some goal in the background.” Which isn’t really fair. I mean, I just can’t... You can never falsify what I’m saying. But in principle, if what I’m saying is true, then you should be able to find a case. You should be able to construct a situation where there is no morality bonus whatsoever. It should be in principle possible to just completely eliminate it, whereas on the morality dominance view, that just really shouldn’t happen. There should always be some bonus.

And that’s why I view that study as really important, because if there is a bonus, then we can both say, “See? I’m right.”

Andy Luttrell:

Well, I want to be mindful of your time and just say thank you for talking with me about all this stuff. This was super cool. And hopefully someday I’ll run into you as a real person in the world.

David Melnikoff:

Oh my God. I would love that. I had a blast. Yeah. I mean, running into anyone would be amazing, but I’d love to chat more about this when we’re both at a conference or something. That would be great.

Andy Luttrell:

All right, that’ll do it for this episode of Opinion Science. Thank you so much to David Melnikoff for taking the time to talk about his work. You can check out the show notes for links to the research that we talked about. For more about this podcast, head on over to OpinionSciencePodcast.com and follow us on social media @OpinionSciPod. And heck, how about you rate and review the show online to give new listeners the confidence to check it out? All right, that’s all for now. See you in a couple weeks for more Opinion Science. Bye-bye!