Accepted for Publication in *European Journal for Philosophy of Science*
Penultimate version
Please cite final version

The Problem of Explaining Shifting Targets

David Colaço

Ludwig-Maximilians-Universität München
Fakultät für Philosophie, Wissenschaftstheorie
und Religionswissenschaft
Munich Center for Mathematical Philosophy
Geschwister-Scholl-Platz 1
D-80539 München, Germany

Contact: david.colaco@lmu.de

ORCID: 0000-0001-7877-5664

The Problem of Explaining Shifting Targets

Abstract: Sometimes we try to explain one thing but end up explaining something else. In this paper, I address when this is a problem, which I call the "Problem of Explaining Shifting Targets" (PEST). In PEST, the explanatory target shifts to another, where the original and shifted targets are manifestly different. Nonetheless, the explaining is treated as for the original target or one equivalent to it, as the agent does not recognize the change. Through the analysis of three candidate cases of PEST, I address the conditions under which it occurs. Further, I address three factors—agents are interested in the target they start with, which then shifts, but they do not recognize it—that capture why PEST should not be all that surprising. PEST is thus an error that makes salient that we must reflect on the dynamic relations between explanation and its target. Because it is subtle, PEST is insidious. It involves appeals to good evidence, and there may be no clear marker that someone has erred. Likewise, it is consequential. Shifts can wreak havoc on a research program or lead to people defending targets that they should not.

Acknowledgements: Thanks to John Dougherty, Stephan Hartmann, Eric Hochstein, and two anonymous referees for their comments on an earlier draft of this paper. An early version of this paper was presented at the 2023 Cambridge-LMU workshop on Philosophy of Science. Thanks to those in attendance for their feedback. This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 528370787.

1. Introduction

Imagine that your friend tells you that consuming turmeric in one's diet prevents cancer. They are at risk for cancer—their parents died from it—so they are weighing preventative options. Instead of seeing a doctor, they opt for drinking turmeric tea. You, let's assume, are skeptical. Your friend does not relent and wants to convince you of their claim. They do a bit of research and find out that a chemical in turmeric inhibits inflammatory enzymes, tumor invasion, and tumor growth. They show you mechanistic diagrams from scientific articles. They explain that this is why turmeric prevents cancer.

Mulling over this explanation, you spot a complication. Perhaps turmeric has an effect, but this does not entail that consuming turmeric in one's diet would have this effect any more than drinking Selsun Blue shampoo would prevent one's dandruff. This would mean that your friend's initial claim is incorrect, at least without massaging what 'consuming turmeric in one's diet' means. You tell your friend this. You say that if they convince you of anything, it is that something in turmeric, somehow introduced to the body, has some anti-cancer effect. This is not what they initially claimed. Though it is interesting, you suggest that your friend should not rely on tea to prevent their cancer. If your friend is steadfast and continues with their tea-drinking plan, it seems that they make an error.

Let me add a wrinkle to this example: what if you had not stumbled upon the complication? You and your friend find the explanatory details (the *explanans*) compelling. However, the problem lies with what these details explain (the explanatory target, or *explanandum*). Had you not stumbled upon this complication, it might have swayed your judgement that consuming turmeric prevents cancer.

This toy example typifies a problem that is the focus of this paper, which I dub the "Problem of Explaining Shifting Targets":

> **Problem of Explaining Shifting Targets (PEST):** In explaining, the explanatory target shifts to another, where the original and shifted targets are manifestly different. Nonetheless, the explaining is treated as for the original target or one equivalent to it.

Thus, one might explain something yet not explain the intended target. PEST occurs when:

1) An agent starts by articulating an explanandum that they want to explain ($E^{Original}$).
2) When explaining $E^{Original}$ an agent shifts to a manifestly different explanandum ($E^{Shifted}$).
3) Despite this shift, an agent treats the explanans as explaining $E^{Original}$, not recognizing the manifest difference between $E^{Original}$ and $E^{Shifted}$.

Unlike recharacterizing phenomena (Craver & Darden 2013; Colaço 2020) or stabilization (Feest 2011; Sullivan 2016), this shift need not be recognized. This shift of the target can result from scientific communication to non-experts. However, it also can happen between and within scientific communities. Because it is subtle, PEST is insidious. It involves appeals to good evidence, and there may be no clear marker that an agent has erred. Likewise, it is consequential. Shifts can wreak havoc on a research program, as agents defend targets that they should not.

In *Section 2*, I unpack the three steps of PEST. In *Section 3*, I discuss three candidate cases of PEST. In *Section 4*, I address why cases of PEST should not be surprising. I also give recommendations on how to deal with PEST. I conclude by addressing two potential objections: what it takes for explananda to be manifestly different in step (2), and what I take recognition to be in step (3). With this, I also sketch out dangers of covert shifts in explanatory targets, highlighting non-epistemic consequences of these shifts.

2. PEST

Let me start with a quote about where explaining tends to begin:

> The starting point of enquiry into explanation… is the gap between knowing that something is the case and understanding why it is. When we ask a why-question—Why is the sky blue? Why does the same side of the moon always face the earth? Why were you late? —we almost always already have a reason to believe that the explanandum is true (Lipton 2001, 102).

Lipton captures that explaining begins with something that one wants to explain. One might have a reason to believe it, but one need not be confident. The explanandum can be articulated via a *characterization* of a phenomenon, or a description of its constitutive features and the conditions under which it occurs (Colaço 2020). An example is the phase transition of lead, characterized as "lead melts at 327.5 ± 0.1 degrees C" (Bogen & Woodward 1988, 309). This is not exhaustive, but

it describes both features (lead melts) and conditions (a temperature range).[1] It can also be articulated via pointing to an instance of a phenomenon. This is the phase transition elicited in a study or evinced by a dataset.

The idea that explaining requires articulating what one wants to explain is present in accounts of the nature of explanation. For instance, accounts of mechanistic explanation indicate that characterizing the explanandum "is a crucial step" in explaining it (Craver 2007, 128). This articulation can be phrased as a question. Once one has articulated $E^{Original}$, we can ask: "Why does $E^{Original}$ happen?" Though there are no general rules on what this articulation must include, one must include something. Further, we should expect that people are interested in what they articulate. This is why they want to explain it.

However it is articulated, an explanandum has specificity. For instance, explaining why lead melts at 327.5 ± 0.1 degrees C is different from explaining why metals melt. More mundanely, explaining why you were late is different from explaining why someone was late. These details tie to its empirical consequences. They are also why it has contrastive foils: if we think of explanations as answering questions that involve contrasts (Garfinkel 1981), or "explanations that answer questions of the form 'why this rather than that?'" (Lipton 2004, 2), how we parse "this" from "that" is a proxy for this specificity. For example, if we accept that lead melts at 327.5 ± 0.1 degrees C, we expect it to melt at this temperature rather than starting to melt around 1085 degrees C, or the melting point of copper. This contrast matters if you have ever soldered electronics. The articulation of $E^{Original}$ substantiates step (1).

Explaining consists in supplying an explanans for an explanandum, where the former answers how or why the latter occurs. I use the term 'explaining' to capture two things in tandem: the formulation of an explanans, and the investigation that substantiates and tests its adequacy, supporting this formulation.[2] There is more to say about explanation, filling many tomes of

[1] It is a type-level description as opposed to one of an instance of lead melting, so it is unclear what it might mean for it to be exhaustive.
[2] One might argue that 'explaining' is too broad. I accept that investigations I include in explaining are not unique to it. The point I make is that they arise through explaining.

philosophy of science. These debates, beyond what I commit to in this section, are orthogonal to my interests in this paper. For the sake of convenience, I focus on a causal, mechanistic account of explanation (Craver 2007). However, PEST is not unique to this account.

Whether a formulated explanans explains an explanandum depends in part on the specificity of the articulated explanandum. If one asks why lead melts at 327.5 ± 0.1 degrees C, it would be unsatisfying to interpret this question as "why does something happen?" and answer it as such. Both might be fine questions that we could answer with an explanans, but these questions, and correspondingly these explanantia, differ. While specificity might not be explicit, it is not inaccessible. It requires that the explainer reflect on what they are trying to explain.

A "shift of the explanatory target" occurs when the agent shifts from $E^{Original}$ to another explanandum, $E^{Shifted}$, where $E^{Shifted}$ has distinct empirical consequences and contrastive foils when compared to $E^{Original}$. The two are *manifestly different*: one can provide evidence for $E^{Original}$ or $E^{Shifted}$ that does not serve as evidence for the other. This difference can be one of specificity: an articulation of $E^{Shifted}$ has more or fewer features or conditions than one of $E^{Original}$. It can also be one of reference: $E^{Shifted}$ picks out a different occurrence from $E^{Original}$.[3] Likewise, the shift can be between a characterization and a phenomenon that is elicited in a study or vice versa. What matters is that these differences can be detected, and an agent's interest in the explanandum is sensitive to these differences.

Step (2) is not an inevitable outcome of explaining, even if it often happens. Sometimes, we do not deviate when explaining. Fictionalizing a case in cognitive science, suppose that we want to explain why people have a hard time remembering long strings of numbers. We might begin by asking why people have a hard time keeping mind strings above a certain length, and through explaining, we find that human capacities are bottlenecked, making it hard to keep in mind larger strings. In this case, they need not change how they ask the explanatory question. $E^{Original}$ might suffice.

[3] Whether these are different will depend on one's theory of reference.

Nothing so far about shifting targets is a problem. For instance, Craver and Darden argue that "a purported phenomenon might be recharacterized or discarded entirely as one learns more about the underlying mechanisms" (2013, 62). On their account, a discrepancy between the explanandum phenomenon and its mechanism should result in the phenomenon's recharacterization. Colaço, by contrast, argues that explanatory insights are typically "merely suggestive": "they provide reason to suspend judgment and evaluate a phenomenon's characterization" (2020, 14). While these accounts disagree on the epistemology of the situation, both indicate that explaining, as I define it, cues researchers in to whether they should recharacterize the explanandum.

There is also literature on stabilization. On Feest's account, stabilization is a fit between "surface regularities," or data patterns, and "hidden regularities," which stand in an explanatory relation to these patterns (Feest 2011, 57). Through stabilization, "attempts to validate claims about hidden phenomena may result in the finding that the surface phenomena… were in fact quite inadequate for the purpose" (Feest 2011, 66). On this account, stabilization can reveal that researchers must rethink their explanatory target. Sullivan addresses the stabilization of constructs, which "originate with a concept that investigators associate with certain observations" (2016, 667). Stabilization consists in determining whether phenomena should be grouped together, which strategies instantiate the construct, and whether the construct should be revised (Sullivan 2016, 668). At this level of analysis, constructs are explananda, and stabilization might change them. Sullivan argues that constructs often are not stabilized in psychology and neuroscience. At the same time, "if taxonomic categories are to be revised as new information about mechanisms comes in, the categories are not supposed to be stable," revealing that explananda are not fixed prior to explaining them (Sullivan 2017, 131).

Recharacterization and stabilization involve steps (1) and (2), but they are how scientific progress can be made. However, except for Sullivan, which I return to in *Section 4*, these accounts are about recognized changes of an explanandum. Oftentimes, we rethink what we want to explain or change what explanatory question we want to answer. The contrast between these cases and PEST becomes salient with step (3). In PEST cases, the shift is not recognized, and the agent goes on thinking that the explanans is for $E^{Original}$. Further, the agent does not address the manifest difference between $E^{Original}$ and $E^{Shifted}$. Rather, they take the explanans to be for $E^{Original}$, not

considering a change, whether they are different, or whether they are of equal interest to the agent. If all three steps are fulfilled, it is a case of PEST.

In PEST, we can grant that the supplied explanans potentially explains $E^{Shifted}$.[4] PEST thus involves explaining. The problem is that agents take this explanans to explain $E^{Original}$, failing to recognize the differences between the two. Providing evidence for the explanans need not resolve this problem, nor will evaluating the explanans alone indicate that it is a PEST case. Rather, we must reflect on the shift from one to the other. We must ask: *what are we explaining, and is it what we are trying to explain?*

3. Candidate Cases of PEST

To address PEST, I review three cases, which I choose for three reasons. First, they have different agents: agents amongst scientific and non-scientific communities, agents within different scientific communities, and agents within a community, respectively. Second, they have details that can be debated, which I return to in *Section 4.3.* Third, they have negative consequences that reveal the damage PEST can cause for a research program. For each, I lay out an explanandum, a shift from this to a manifestly different explanandum, and a persisting idea that the explanans explains the original. To make my concerns salient, I grant that each case involves explaining of $E^{Shifted}$.

3.1.Turmeric, the cancer preventor?

In scientific reports and communications, people claim that the dietary consumption—that is, eating or drinking—of turmeric prevents cancer (see Gibson-Moore & Spiros 2017 for a review). These claims are presented by nutrition scientists who aim to account for correlations between turmeric consumption in a national diet, say of India versus the United States, and incidence of cancer. Some researchers state that turmeric might "prevent" cancer, though not all focus on diet (Gibson-Moore & Spiros 2017, 141-142). Uptake of this research emphasizes dietary consumption. One popular article discusses the "surprising health benefits of curry powder,"

[4] A potential explanation "satisfies all the conditions on actual explanation, with the possible exception of truth," likely due to inaccess to actual explanations (Lipton 2001, 5).

including that it "may have anti-cancer effects" (Kubala 2020). Another states that turmeric "can tackle and prevent major health problems like cancer" (Holden 2024). Both hedge these claims by mentioning that the science is ongoing.

Scientists offer explanantia involving curcumin,[5] a chemical of the turmeric rhizome, and its effect on the body. These models include: "antioxidant activity [e.g. induction of the reactive oxygen neutralising enzymes glutathione-S-transferase (GST) and quinine reductase]," "inhibition of proinflammatory enzymes [e.g. cyclooxygenase-2 (COX-2), lipoxygenase (LOK) and inducible nitric oxide synthase (iNOS)]," "inhibition of tumour invasion and anti-angiogenic activity (prevention of new blood vessel proliferation in tumour) (e.g. through downregulation of vascular endothelial growth factor)," and "induction of apoptosis (e.g. in human cell lines such as leukaemia, gastric, breast and colon cancer cells)" (Gibson-Moore & Spiros 2017, 143; see Kasi et al. 2016 for more detail). These schemata inform and are informed by in vitro studies. Both popular articles I cite mention curcumin and some of these schemata.

Skeptics question how much turmeric a human would need to consume to prevent cancer. Foodstuffs tend to contain little curcumin (Tayyem et al. 2006), suggesting that the amount of turmeric one would have to consume to induce this effect would be much higher than the amount people who consume it every day eat. They also question the bioavailability and retention of orally consumed curcumin. Eating or drinking turmeric does not free curcumin to become available to systems throughout the human body. For this reason, researchers question whether "an oral dose, in concentrations that have shown in vitro activity, will be effective in human clinical trials" (Gibson-Moore & Spiros 2017, 145). Skeptics are tentatively willing to accept that there is something to the idea that turmeric prevents cancer. However, their skepticism is directed at diet. This characterization, however, aligns with how it is presented in popular articles and some scientific reports. Further, if these studies are taken seriously, then these proponents do not lean on bad science.

[5] 'Curcumin' refers to a chemical, as opposed to a synonym for turmeric.

In this case, we first see scientific interest in studying a phenomenon characterized as dietary consumption of turmeric prevents cancer. This is $E^{Original}$ in this case. Its articulation as the consumption of turmeric fits step (1) of PEST. Researchers then study the causal basis of an effect like this and formulate mechanistic schemata of it. This fits my idea of explaining. Following these in vitro studies, and based on how the studies were performed, some researchers claim that their schema explains a cancer preventing effect of a chemical in turmeric when introduced to bodily tissue. This is the shifted explanandum, or $E^{Shifted}$ in this case, which is explained by the explanantia that are schematized.

Step (2) of PEST is fulfilled when $E^{Original}$ is a manifestly different explanandum when compared to $E^{Shifted}$. One reason to say that they are is that $E^{Original}$ is articulated in terms of dietary consumption, while $E^{Shifted}$ is not. $E^{Shifted}$ is less precise when compared to $E^{Original}$. Evidence could be collected, such as from studies on intravenous injection of curcumin, that would support $E^{Shifted}$ but not $E^{Original}$. They thus have distinct empirical consequences. Likewise, they have distinct contrastive foils. We could ask "why does eating turmeric prevent cancer rather than not prevent cancer?" for $E^{Original}$, while we cannot ask this question for $E^{Shifted}$. There is also a case to be made that they are not equally interesting. One motivation for $E^{Original}$ seems to be the correlation between dietary habits and cancer rates. This correlation is not accounted for if one drops the dietary condition. In addition, what motivates communications of this research is this condition. People are excited that a change in one's diet can prevent cancer. Further, growing concerns about oral dose effectivity highlight that this condition is a sticking point, which proves difficult to empirically corroborate. Scientists now seem aware that they should allay a change in judgment of the probability of dietary consumption of turmeric preventing cancer based on the in vitro studies alone. For these reasons, an assessment of the case should fall on them being manifestly different, fulfilling step (2) of PEST.

Step (3) of PEST is fulfilled when despite this shift, an agent treats the explanans as explaining $E^{Original}$, not recognizing the manifest difference between the two. Many agents continue to favor $E^{Original}$ but present the schema that better explain $E^{Shifted}$ as explanations of $E^{Original}$. Some scientists now recognize the differences between the two, downplaying the dietary condition. However, these

attempts seem lost amongst non-experts and some scientists as well, the agents in this case.[6] While this loss may in part be due to sloppy science journalism or misunderstanding of what researchers mean by 'consume,' the reader should not forget that scientific interest in this topic stemmed from dietary correlations and an interest in explaining them. Many experts and non-experts articulated this phenomenon and were interested in its explanation. Not recognizing the manifest differences of the dietary condition and the challenges that surround corroborating it, some agents have held onto $E^{Original}$. Hence, step (3) is fulfilled, and we have a candidate PEST case.

3.2.Aggressive rodents?

In the 20th century, scientists took greater interest in studying human sexually dimorphic behavior (see Longino 1983 for a review). In this research, a focus was put on answering why there (allegedly) is a difference between "so-called masculine behavior (aggressive, assertive, dominant, independent, creative) and so-called feminine behavior (passive, submissive, gentle, dependent, nurturing)" (Longino 1983, 12). Scientists interested in this question homed in on why "masculine" behavior is aggressive while "feminine" behavior is not. Interest in asking this question was informed by scientists' judgements that there is a sex-based difference between aggression, which reflects the values of these scientists, as Longino addresses in detail (1983).

Scientists offered explanantia in terms of the influence of androgens and estrogens, what are known as sex hormones. "Differential distribution of these hormones between males and females," Longino notes, "has been cited as causing or influencing differences in behavior between the sexes" (1983, 12). In explaining, scientists performed rodent studies. They modulated the presence of these hormones as well as the animals' endogenous abilities to produce them. In these experiments, researchers focused on the increase or decrease of engagement of cage fighting behavior in these rodents, finding a correlation between hormone presence and this behavior.

[6] Interestingly, one popular article was revised, noting potential differences between the original and shifted target (Windsor 2024). Many, including the two I cite, remain the same.

Skeptics of this research do not direct their skepticism at the finding that exposing rodents to more androgens can increase fighting behavior in cages. Rather, they argue that "the hypothesis that human gender-related behaviors are hormonally determined or influenced is not supported by the evidence adduced for it" (Longino 1983; see Adkins 2019). That is, they are tentatively willing to accept that hormones can affect and thus explain this rodent behavior, but they are critical of the idea that this explanans explains human aggression as initially characterized. Many of the scientists doing the rodent research were not the same scientists who elaborated theories about how hormones explain human aggressive behavior and its connection to social roles. However, those who did elaborate theories treated this explanans as for human sexually dimorphic behavior.

In this case, we first see interest in studying a phenomenon characterized as human sexually dimorphic behavior and specifically human aggression. This is $E^{Original}$ in this case, and its articulation fits step (1) of PEST. Researchers then studied the causal basis of an effect like this and formulated a schema of it. This fits my idea of explaining. The investigations involved an operational measure of aggression in terms of fighting behavior, with rodents serving as human proxies. Following these rodent studies, and based on how the studies were performed, some researchers claim that their schema explains fighting behavior differences in rodents. This is the shifted explanandum, or $E^{Shifted}$ in this case, which is explained by the explanans that is schematized.

For step (2), this case involves an attempted operational measure of aggression and an extrapolation from rodents to humans, which, without tracing the processes of each, complicates assessments of the relation between the two (Steel 2007). One reason to say that the two explananda are manifestly different is that many more features and conditions are baked into $E^{Original}$ when compared to $E^{Shifted}$. Longino notes that "when appealed to in social explanations [aggressivity] includes not only combativeness but also such traits as assertiveness, independence, and intelligence" (1983, 13). This means that the concern is not just about extrapolation; it is about whether the explanation is for human aggression as initially characterized as opposed to aggression articulated in a separate way. The scientists might explain some human behavior, but this is not the target with which they started. The idea that these traits are linked has no direct support in this case, and the effect produced in rodents does not have any features that connect to these other

traits. There is also a case to be made that they are not equally interesting. $E^{Original}$ ties to how sex-based differences relate to social roles and norms, going beyond fighting behavior, while $E^{Shifted}$ does not. For these reasons, an assessment of the case should fall on them being manifestly different, fulfilling step (2) of PEST.

For step (3), scientists of the time theorized about sex-based social roles, presenting hormonal explanations that better explain $E^{Shifted}$ as explanations of $E^{Original}$. They looked to elicitations of a behavioral change in the studies that were part of their explaining, which they claimed was consistent with human aggression as characterized. Further, they did not focus on triaging their operational measures, connecting to other constitutive features of human aggression, or tracing processes that could warrant extrapolation. Their studies were fixated on their explaining, not addressing whether the two are manifestly different and whether their explaining was relevant to $E^{Original}$. For this reason, the problem in the case extends beyond just being one of operationalization or extrapolation; their responses (or lack there of) to these issues turned it into one of shifting explanatory targets. Skeptics, by contrast, collated concerns about whether the two are manifestly different. It would be unfair to say that skeptics did not recognize that the two might be manifestly different. Nonetheless, the proponents of the era, the agents in this case, did not recognize the difference. Hence, step (3) is fulfilled, and we have a candidate PEST case.

3.3. Can we transfer memories?

Memory transfer is one of the wilder cases in the history of science, becoming a paradigm of scientific failure (see Colaço 2018 for a review). The phenomenon was once characterized as the transfer of learned associations from one organism to another via the transfer of biological material. A proponent, Georges Ungar, investigated memory transfer in a series of studies. Without training, rodents hide in the dark. Ungar's studies associated darkness with electrical shock via shocking the rodents in the dark during training. Following this training, rodents exhibit dark-avoidance behavior, an association of the stimuli. After training the rodents, Ungar's team extracted chemicals from their brains and injected them into untrained mice, after which they were tested for dark-avoidance behavior. He argued that if they exhibited this behavior, he transferred a memory.

Further, he schematized a "mechanism of learning," which he called the "chemical connector hypothesis" (1968, 226). This schema consists in connector molecules that are transmitted between neurons. Ungar identified a polypeptide he thought was a connector on this schema, naming it "scotophobin," ancient Greek for "darkness fear" (Setlow 1997, 186). According to Ungar, the transfer of this molecule explains memory transfer. Ungar argued that scotophobin "was synthesized and distributed to a number of laboratories, which confirmed its dark-avoidance inducing effect" (1974, 599). He argued that studying the mechanism could "bypass the 'transfer' controversy" yet "almost automatically validate the bioassay approach," or administering biological material and measuring the behavior (1975, 171). Until his death in 1978, he argued that this schema and its component, scotophobin, explains memory transfer as he characterized it.

Ungar convinced many contemporaries (Colaço 2018, 37), but skeptics argued that his "conclusions are more likely false than true": one skeptic, representing the judgment of the group, suggested that the "synthesis of the pentadecapeptide [scotophobin] is essentially sound" (Stewart 1972, 209). This quote hints at the fact that no skeptics directed their challenges at Ungar's schema or the involvement of a polypeptide. Peptide transmission in the brain had a little support at the time, and this support grew in the late 1970s (Setlow 1997, 189). Skeptics were tentatively willing to accept that this schema potentially explains a behavioral effect.

However, skeptics questioned whether this behavior is associative. To his credit, Ungar controlled for sensitization in much of his work, ruling out a common non-associative phenomenon in his studies. Yet, skeptics raised concerns between what he showed and what he stated. Though not explicit when Ungar was alive, skeptics later argued that the elicited phenomenon should be characterized as an emotional stress reaction (Rose 1992, 195), which is not an association of darkness and shocks. Given that both supporters and skeptics in the debate committed to the idea that memory is associative, whether memory transfer occurs became a debate over whether one can transfer associations. The response of skeptics is that this explanans does not explain memory transfer as characterized. Though aware of skeptics' concerns, Ungar did not address them in his research. He thought he had ruled out non-associative effects like sensitization, and no one at the time had articulated emotional stress responses as a target, much less tried to link them to nascent ideas of neuropeptide signaling.

In this case, we first see interest in studying a phenomenon characterized as the transfer of learned associations from one organism to another. This is $E^{Original}$ in this case, and its articulation fits step (1) of PEST. Ungar studied its causal basis and formulated a mechanistic schema of it. This fits my idea of explaining. Following these scotophobin studies, and based on how the studies were performed, Ungar claimed that his schema explains the behavioral change that he had elicited. This behavioral change is the shifted explanandum, or $E^{Shifted}$ in this case, allegedly elicited in his studies, which is explained by the explanans that was schematized.

For step (2), one reason to say that the two explananda are manifestly different is that $E^{Original}$ is articulated in terms of learned associations, while $E^{Shifted}$ has no parallel articulation of constitutive features, focusing instead on behavioral changes after injection of scotophobin. At best, $E^{Shifted}$ seems less precise in its articulation when compared to $E^{Original}$. At worst, $E^{Shifted}$ refers to a different phenomenon than to $E^{Original}$. In research on memory, a perennial concern is mistaking non-associative behavioral effects for memory. This is why non-associative effects were persistent confounding factors to be controlled for in memory studies (Colaço 2018). Thus, articulating $E^{Shifted}$ picks out these non-associative effects that researchers wanted to avoid. There is also a case to be made that they are not equally interesting. One motivation for the articulation of $E^{Original}$ is that it connected the research to memory. The idea that the study involves a mere behavioral change due to injection does not carry the same interest. For these reasons, an assessment of the case should fall on them being manifestly different, thus fulfilling step (2) of PEST.

For step (3), Ungar continued to favor $E^{Original}$ but present the mechanistic schema that better explain $E^{Shifted}$ as explanations of $E^{Original}$. He elicited a behavioral change in the studies that were part of his explaining, which he claimed was consistent with memory transfer as characterized. His skeptical peers seemed to recognize that there were differences between $E^{Original}$ and $E^{Shifted}$, though recognition and articulation of $E^{Shifted}$ and its connections to Ungar's explanans would only develop after his death. Ungar did not acknowledge the difference, nor did other proponents at the time. He shifted from asking "why does the transfer of associations occur?" to "why does the behavior I have elicited occur?". However, he, an agent in this case, did not recognize that the two

questions might have different answers. Hence, step (3) is fulfilled, and we have a candidate PEST case.

3.4.The Consequences of PEST

I pick these examples for my candidate PEST cases in part because of their striking negative consequences. For the turmeric case, it can lead to two consequences. First, like your hypothetical friend, they might drink tea instead of seeking medical aid. Second, public fixation on turmeric's anti-cancer effects might doom the research to the bin of pseudoscience. This would not be because of the methodological quality of the research per se, but rather because people associate the research with $E^{Original}$. Any insights about curcumin's effects on cancer might be lost with this dismissal.

The aggression case shows that PEST can create a situation in which scientists' values were taken to be supported, even when there was suspicion that their explanans does not explain the lump of human sex-based differences they espoused. This consequence amounts to a veneer of support for a target that we do not have a reason to believe. Further, it has since become a paradigmatic case in which gender bias led scientists to endorse a position whose theoretical assumptions have turned out to be suspect. This is not because of the quality of the rodent studies, but rather because people associate the research with $E^{Original}$.

The memory transfer case shows the long-term consequences of PEST. Ungar thought that the explanation he supplied explained memory transfer as initially characterized. He stuck to $E^{Original}$ until his death, likely because it is what interested him and others. Had he acknowledged a difference, he might have reoriented to a different but nonetheless novel phenomenon regarding stress reactions. He might be recognized as a pioneer, foreshadowing insights about neuropeptide signaling (Wilson 1986). Instead, his legacy is associated with a paradigmatic scientific failure.

These negative consequences show that while these shifts are salient in hindsight, the subtlety of PEST means that we must remain vigilant of whether a shift occurs. For a simple example with this flavor, I am reminded of Kahneman's anecdote of trying to explain why politicians are so often

adulterous, failing to recognize for some time that the question he should have been asking, and the question he ultimately found he could explain, is why politicians' affairs are more commonly reported (2011). If this shift is not recognized, it can pull people down problematic avenues of research, as the cases show.

4. PEST Controls?

With candidate PEST cases in hand, I hope I have shown the reader that this is a problem. In this section, I address why this problem should not be surprising, given how we account for explanation. I also provide ideas on how PEST cases might be preempted or dealt with when in progress. I conclude this section with replies to objections to my account.

4.1. Why PEST Should Not Surprise

The issue that lies at the core of PEST cases is a combination of three factors that, together, make PEST possible. These three factors tie to the three steps of PEST.

The first factor is that agents are interested in an explanandum that they start with, which is why they want to explain it. There are many things that we might want to study, so picking something as an explanandum reflects a choice. As the aggression case shows, it can also reflect values. Of course, this choice can be constrained by institutional or economic factors. Consider this factor to hold in cases where researchers are not coerced into studying an explanandum. This is sufficient for my purposes, and I argue that it holds often enough.

The second factor is that explananda are not fixed or static. They can change during explaining, with explaining reshaping the explanandum.[7] During explaining, researchers aim to substantiate and test their explanans. If researchers induce the explanandum in their investigations that inform how they should formulate the explanans, then they might determine something about the explanandum during this process. However, this reshaping cuts to the core of how explaining

[7] For anyone with an "ontic" disposition to explanation, one could substitute 'reshape an explanandum' for 'reorientate us to another explanandum' without my argument losing its thrust.

works. By devising and assessing an explanans, researchers acquire new pieces of evidence to answer these questions. Explaining provides new information: there has been a move to answering a question that they could not before.[8] With that, the results of these investigations can be different from studies on the explanandum alone. For instance, we might study the dietary effects of turmeric by correlating incidence rates. Once we start explaining, however, we move into in vitro studies on cancer cell cultures. This means that new information can come from studies with different methods.

This information need not just inform how an explanans is formulated. If this new information comes from different methods, then it can provide new expectations about the target: what features and conditions characterize it, what studies elicit it, what data provide evidence for it, or what statistical patterns represent it (van Dongen et al. 2024). We can expect it to occur in situations like those of the novel studies, which are expectations of the conditions under which it occurs. These expectations can inform new ways the target might be discriminated from alternatives, making salient new, distinct contrastive foils in the process. But, building on *Section 2*, we identify and differentiate explananda via their empirical consequences and their contrastive foils. By consequence, this new information from explaining can influence how we identify and differentiate explananda from alternatives. In this sense, explaining can "reshape" the explanandum.

The third factor is that agents need not recognize when explaining reshapes an explanandum. They certainly can: Craver and Darden, Colaço, and Feest address recognized changes.[9] Sullivan, by contrast, highlights that we should be skeptical that researchers stabilize their constructs, but this account focuses on how researchers do not stabilize when they should. What PEST shows is that this change easily can occur through explaining yet also not be recognized by the agents invested in the target. My analysis vindicates this part of Sullivan's account. Nonetheless, we should not take these accounts to entail that shifts only occur when researchers intend them or recognize that

[8] I use the term 'information' rather than 'evidence' because it is unclear what it is evidence *for*.
[9] A shift might be determined from explaining. If we characterized copper as melting at 300 degrees C and find, when explaining, that the studies we perform require heating samples to 1000 degrees, this tells us that the characterization does not match what is elicited.

they occur. Further, the information acquired through explaining need not serve as evidence that rebuts or undercuts support for or belief in the original explanandum. If this information does not rebut or undercut, supporters of the original explanandum are not in any obvious way acting irrationally if they are steadfast.

When we put together these three factors, I wager that the fact that PEST can occur should not surprise the reader, even if it is not inevitable. The ease with which it can occur is laid bare when we unpack why something has gone wrong in my toy example in the *Introduction*. To be fair, one might point to many issues in this case, including that your friend should not try to defend that the phenomenon occurs by explaining it or try to defend a phenomenon whose connection to the explanans is not clear. Likewise, one might point out that they should consider the body of reasons to be skeptical of the phenomenon, regardless of what explaining they muster. To this list, I add another diagnosis. They should appreciate that what they aim to defend is not the same as what they end up explaining. They have shifted the target but do not recognize that what they have shifted to might be manifestly different to what they aim to defend and what interests them. This is PEST.

4.2.Dealing With PEST

Accounting for PEST informs a set of recommendations on how it might be addressed. The first recommendation is the most straightforward: agents should recognize that explananda can change, and one way they can change is through trying to explain them. However, this recommendation alone does not confer much in the way of a strategy.

A recommendation that is a strategy is that there should be a greater degree of tracking explananda over time. This involves keeping track of changes in how explananda are articulated as research progresses. This strategy could track stated characterizations of phenomena, but it equally could track communication with agents about them. As the cases show, changes in explananda are often not explicitly stated, which is part of the problem. Nonetheless, picking up cues about how agents talk about an explanandum can indicate whether there are any shifts in how the explanandum is articulated or investigated. Individuals in each case became suspicious of a shift in the explanatory

target, as there are skeptics for all three. This fact suggests that tracking changes in the explanandum is possible.

Another recommendation is that we can ask probing questions about what agents have in mind when they endorse or study an explanandum. That is, we can probe them on how they articulate what they study, and what changes would lead them to agree that the explanandum is manifestly different. One can ask questions of the following form: given the agent's articulation, if certain features or conditions were omitted or included, would the agent say it is manifestly different? Questions of this form would provide us with a sign of what agents commit to as constitutive of what they want to explain. If, for instance, agents in the turmeric case made clear that the dietary condition is not important to their characterization of turmeric's cancer preventing effect, this insight would help us determine whether there is a shift in the explanatory target that is manifestly different from where they started. In my cases, it is unclear what these probing questions would reveal, which is why I refer to them as 'candidate cases.'

A final recommendation is that we can collect evidence relevant to assessing the (in)equivalence of $E^{Original}$ and $E^{Shifted}$, if we suspect that there is a shift. If the two have distinct empirical consequences and contrastive foils, then a study could be performed that would support one but not the other. In principle, some tests could be performed that would help us better determine whether these explananda are inequivalent. This is not unlike the methods Sullivan includes as part of stabilization, including construct explication and assessment (2016, 668), or Colaço's claims that there are tests we can perform that help us determine the adequacy of a phenomenon's characterization (2020). While trying to collect this evidence is a tractable method in principle, evidence of the inequivalence of explananda can be hard to come by, which means that this evidence need not be forthcoming. If, for instance, an inequivalent phenomenon is elicited, it is often a challenging endeavor to identify it, let alone rule it out (Colaço 2024). It need not be the case that proponents of an original explanandum actively avoid testing whether it is equivalent to the shifted one. Though difficult, this strategy is possible, and it can be used to test whether explananda are manifestly different.

4.3.Objections and Replies

Before concluding this paper, let me address two potential objections related to how I account for PEST. The first is what it means for explananda to be manifestly different. One might object that in my cases, the explananda are not (or might not be) manifestly different, even if they are not equivalent. This raises the question of what, if anything, separates inequivalence from manifest difference.

Let me make this objection concrete with the turmeric case, as it is the case where I suspect the objection most naturally arises. Perhaps $E^{Shifted}$, that turmeric chemicals prevent cancer, is merely a more technical way of articulating $E^{Original}$, that consuming turmeric in one's diet has cancer preventing effects. $E^{Shifted}$ thus might reflect what nutrition scientists had in mind from the onset, as it is focused less on diet and more on a generic sense of 'consumption.' $E^{Shifted}$ does not rule out $E^{Original}$ on these readings. Further, there is no evidence that in vitro conditions are inequivalent to the real-world conditions of (say) drinking turmeric tea. $E^{Shifted}$ also remains interesting: even if diet is not a condition, it might still be a preventative measure for cancer, which is something that we want to prevent. On this reading, one might accept that $E^{Original}$ and $E^{Shifted}$ are not equivalent. However, one might reject that they are manifestly different.

Relatedly, one might object that in the cases I present, the explaining relates to the original explananda. One could describe the cases as beginning with less specific characterizations of phenomena. In the first, this might be that turmeric has some effect on cancer. In the second, this might be that sex hormones have an effect on behavior that can be labeled as 'aggression.' In the third, this might be that something is transferred between organisms that has behavioral consequences. In each case, both $E^{Original}$ and $E^{Shifted}$ can be described via these characterizations. Likewise, they might note that 'manifestly different' does not mean 'completely unrelated.' In each case, $E^{Original}$ and $E^{Shifted}$ share characteristics. One might thus argue that it is plausible to think that any explanans for $E^{Shifted}$ then at least partially explains $E^{Original}$, as it explains the characteristics that are shared amongst them.

In response to this objection, I ask: do these comparatively unspecific characterizations capture what agents chose as their original explanandum? In the turmeric case, what seems of interest to many agents is that dietary choices can prevent cancer. To explain something that, in this case, does not obviously translate to a viable clinical intervention, let alone a diet change, seems unsatisfactory. In the aggression case, what was of interest to many agents was human sexually dimorphic behavior, inclusive of many traits thought to be related to gender and social roles. Merely labeling fighting behavior as 'aggression' does not warrant agents appeals to the explanans as explaining this richer set of traits. In the memory transfer case, what seems of interest to the agents was the transfer of memory. To explain something that, in this debate, does not meet the agreed-upon criteria for memory seems unsatisfactory, just how explaining why metals melt is an unsatisfactory answer to why lead melts at 327.5 ± 0.1 degrees C.

My response to this objection is thus that changes to explananda are manifestly different when they can be characterized as different phenomena, with different constitutive features, conditions of occurrence, or both, *and* they hold different interests for this agent. By consequence, changes of specificity are sufficient for manifestly different changes when the features or conditions that are included or excluded during the shift change whether and why the phenomenon is interesting. This reply affirms my recommendation for probing questions about an agent's commitment to the features and conditions articulated in a phenomenon's characterization. 'Manifest difference' captures the inequivalence of the explananda, affording different empirical consequences and contrastive foils. At the same time, it captures what agents would be willing and unwilling to endorse or give up on when articulating their explananda, given what initially interested them.

Returning to the turmeric example makes this response salient. The constitutive feature of turmeric *preventing* cancer is one that might be more or less precisely articulated, leading to explananda that are inequivalent but not manifestly different on my account. However, the articulation would be manifestly different if the feature were dropped entirely from the articulation, leaving us with something like "consuming turmeric does something related to cancer." This is because the articulation would be consistent with turmeric causing cancer, while the initial explanandum is not. The same can hold for the dietary condition. If agents are unwilling to give up on the dietary

condition in the initial articulation of the turmeric case, then dropping this condition counts as a manifest difference.

My response also speaks to the argument that if the explanans explains $E^{Shifted}$, and $E^{Shifted}$ shares characteristics with $E^{Original}$, then it partially explains $E^{Original}$ via explaining these shared characteristics. To begin, it is not clear that the explanantia in these cases are partial in the sense that adding to them would better explain $E^{Original}$ (Ruben 2015, 19). However, I posit that there is a difference in targets between explaining an explanandum as characterized and explaining some of these characteristics. Interest unites a set of characteristics as an explanandum, which helps confer an explanandum its specificity. This interest can change when characteristics are included or excluded, and it thus can change when only a subset of these characteristics are targeted. Interest can of course change, leading to a recharacterization. Yet, this point highlights that changing from explaining something as characterized and explaining a subset of its characteristics can amount to a change in explanatory target with a difference of interest, which makes the change result in a manifest difference and a shift.

The second objection ties to what it means for agents to not recognize a shift in the explanandum. I have chosen my cases in part because the agent or agents who commit PEST are different. However, one might object that in each of the cases I present, there were individuals who were at least suspicious that the explanans explained the original explanandum. Indeed, most scientific cases involve some degree of disagreement. It would therefore be too narrow to account for PEST as occurring only when there are no voices of dissent or skepticism. Nonetheless, this raises the question of what 'recognition' means in step (3).

In response to this objection, there are cases in which an agent does not notice or is not aware of the shift. This sort of scenario meets the most everyday sense of 'recognition.' However, the cases that I investigate are ones in which the agent might not be completely unaware of suspicions. They need not be unconceived alternatives. Rather, what matters for diagnosing recognition is that the agents do not address potential shifts or the potential differences between explananda. Likewise, shifted targets may not be articulated, even by skeptics. Failing to recognize, thus, can involve a passive or active disinterest in addressing alternative explananda revealed through explaining. Not

recognizing a shift may be due to a "preference bias" for results supportive of $E^{Original}$ over those supportive of $E^{Shifted}$ (Wilholt 2009). This bias shapes how results are interpreted. The key is that there is no defense given for why the agent is steadfast in treating the explanans as for $E^{Original}$. Rather, these agents stick with it, likely because it is what interests them and because it is what they think that they have explained. For recognition, all that matters for my account is that the agents do not acknowledge and address $E^{Shifted}$.

However, there is more to shifting targets. In this paper, I have provided a good-faith analysis of PEST by focusing my analysis on agents that do not recognize the shift. While I stand by this analysis, it would be remiss of me to avoid discussion of the possibility that an agent could shift targets *on purpose*, perhaps while feigning ignorance or simply not acknowledging it. One might bait-and-switch the target, implicitly shifting between a controversial, interesting target and a comparatively banal but explainable one. This would turn shifting targets into a kind of agnotology, inducing ignorance through methods like putting forward misleading scientific claims (Pinto 2017). Here, the misleading claim is the explanation while the ignorance relates to the fact that it does not explain $E^{Original}$. This tactic could be used to concoct the veneer of support for a target that we should not accept. It could be effective if people are questionable judges of explanations (Anderson & Sechler 1986).

Considering covert shifts of explananda may push step (3) beyond how some people would interpret it. In these situations, perhaps 'not recognizing' could be substituted for 'not overtly acknowledging.' Nevertheless, the possibility of shifts being done on purpose again highlights the importance of addressing shifting explanatory targets. The problems are not just epistemically problematic. They can have considerable non-epistemic consequences, such as being used to legitimize pseudoscience and achieve unscrupulous aims, such as supporting the sale of questionable commercial products. This very well might be a part of the turmeric case: companies want to sell their teas.

5. Conclusion

Explanatory targets can shift when we try to explain them. This is a problem when agents do not recognize that the shift has happened. In this paper, I have laid out this problem of explaining shifting targets with three candidate cases. I have highlighted that this problem should not be surprising, as agents are very often interested in what they want to explain, explanatory targets are not static, and agents need not recognize when explaining reshapes an explanandum. I refrain from speculating about the frequency of occurrences of PEST in science and its communication. Minimally, I argue that they happen, and when they happen, they cause trouble for a scientific research program.

This problem shows that, even when including the accounts I have reviewed in this paper, philosophical analyses of explanation do not provide clear answers to what an explanatory target should be or how we should assess it. We have an impoverished account of what explananda are, and correspondingly we have an impoverished account of when two explananda are the same or different. As a result, PEST can be difficult to address and prevent with our current philosophical resources. PEST shows that we must reflect upon the relation between explanans and explanandum, and scholars ought to take my recommendations for dealing with PEST seriously. My analysis in this paper supports the idea that we should always ask: what are we explaining, and is it what we are trying to explain? When we ask this two-part question, we acknowledge that explanatory targets can shift.

References:

Adkins, E. K. (2019). Genes, hormones, sex and gender. In *Sociobiology: Beyond Nature/nurture?* (pp. 385-415). Routledge.

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of personality and social psychology*, *50*(1), 24.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The philosophical review*, *97*(3), 303-352.

Colaço, D. (2018). Rip it up and start again: The rejection of a characterization of a phenomenon. *Studies in History and Philosophy of Science Part A*, *72*, 32-40.

Colaço, D. (2020). Recharacterizing scientific phenomena. *European Journal for Philosophy of Science*, *10*, 1-19.

Colaço, D. (2024). When remediating one artifact results in another: control, confounders, and correction. *History and Philosophy of the Life Sciences*, *46*(1), 5.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.

Feest, U. (2011). What exactly is stabilized when phenomena are stabilized?. *Synthese*, *182*(1), 57-71.

Garfinkel, A. (1982). *Forms of explanation: Rethinking the questions in social theory*. New Haven: Yale University Press.

Gibson-Moore, H., & Spiro, A. (2017). Can turmeric really prevent cancer?. *Nutrition Bulletin*, *42*(2).

Holden, J. (2024). Kitchen cupboard £1 spice that helps prevent cancer, lowers risk of heart disease and burns fat. *MyLondon.* https://www.mylondon.news/news/health/kitchen-cupboard-1-spice-helps-28662510. Accessed September 26th, 2024.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kasi, P. D., Tamilselvam, R., Skalicka-Woźniak, K., Nabavi, S. F., Daglia, M., Bishayee, A., ... & Nabavi, S. M. (2016). Molecular targets of curcumin for cancer therapy: an updated review. *Tumor Biology*, *37*, 13017-13028.

Kubala, J. (2020). 9 Surprising benefits of curry powder. *Healthline*. https://www.healthline.com/nutrition/curry-benefits. Accessed September 26th, 2024.

Lipton, P. (2001). Is explanation a guide to inference? A reply to Wesley C. Salmon. In *Explanation: Theoretical approaches and applications* (pp. 93-120). Dordrecht: Springer Netherlands.

Lipton, P. (2004). *Inference to the Best Explanation*. Second Edition. London: Routledge.

Longino, H. (1983). Beyond "bad science": Skeptical reflections on the value-freedom of scientific inquiry. *Science, Technology, & Human Values*, *8*(1), 7-17.

Pinto, M. F. (2017). To know or better not to: Agnotology and the social construction of ignorance in commercially driven research. *Science & Technology Studies*, *30*(2), 53-72.

Rose, S. (1992). *The making of memory: From molecules to mind*. London: Bantam Press.

Ruben, D. H. (2015). *Explaining explanation*. Routledge.

Setlow, B. (1997). Georges Ungar and memory transfer. *Journal of the History of the Neurosciences*, *6*(2), 181-192.

Stewart, W. W. (1972). Comments on the chemistry of scotophobin. *Nature*, *238*(5361), 202-209.

Steel, D. (2007). *Across the boundaries: Extrapolation in biology and social science*. Oxford University Press.
Sullivan, J. A. (2016). Construct stabilization and the unity of the mind-brain sciences. *Philosophy of Science*, *83*(5), 662-673.
Sullivan, J. A. (2017). Coordinated pluralism as a means to facilitate integrative taxonomies of cognition. *Philosophical Explorations*, *20*(2), 129-145.
Tayyem, R. F., Heath, D. D., Al-Delaimy, W. K., & Rock, C. L. (2006). Curcumin content of turmeric and curry powders. *Nutrition and cancer*, *55*(2), 126-131.
Ungar, G. (1968). Molecular mechanisms in learning. *Perspectives in biology and medicine*, *11*(2), 217-232.
Ungar, G. (1975). Molecular coding of memory. *Minireviews of the Neurosciences from Life Sciences*, 459-468.
Ungar, G. (1975). Is there a chemical memory trace?. *Israel Journal of Chemistry*, *14*(S1), 169-176.
van Dongen, N., van Bork, R., Finnemann, A., Haslbeck, J., van der Maas, H. L., Robinaugh, D. J., ... & Borsboom, D. (2024). Productive explanation: A framework for evaluating explanations in psychological science. *Psychological review*.
Wilson, D. (1986). Scotophobin resurrected as a neuropeptide. *Nature*, *320*(6060), 313-314.
Windsor, R. (2024). Turmeric and its antioxidant curcumin. *Verywellhealth. https://www.verywellhealth.com/turmeric-curcumin-benefits-7110668.* Accessed September 26th, 2024.
Wilholt, T. (2009). Bias and values in scientific research. *Studies in History and Philosophy of Science Part A*, *40*(1), 92-101.