

[This is a draft of a chapter of my forthcoming dissertation. You may cite it with permission. Find me at aaronmatthewfranklin (at) gmail (dot) com]

Transcendental Sentimentalism

Introduction

The notion that our capacity for sentiment is in some way tied up with our practice of offering moral evaluations has a degree of *prima facie* plausibility that is rare among philosophical theses. This may be for no other reason than we often feel very strongly about what is right and wrong, with these feelings often amounting to something over and above the sort of convictions we have about non-normative matters. Of course, if emotions are “in some way tied up with” our evaluative capacities, the interesting philosophical work lies in the specification of this relation. If the strong feelings we have about value were nothing but mere correlations or psychological epiphenomena, they would in no way help *explain* the nature of evaluative judgment. The moral sentimentalist rejects this picture; for them, the relationship between emotions and evaluative judgment is explanatorily significant. However, sentimentalist theories vary widely on the precise explanatory role they find for the sentiments. This paper will add to this variety by identifying a distinct form of sentimentalism I will call “Transcendental Sentimentalism.” Transcendental Sentimentalism (hereafter, “TS”) claims that having or having had a sentimental response to *x* is a necessary condition of the possibility of a person counting as having non-inferential evaluative knowledge of *x*.

The structure of this paper will be as follows: First, I introduce TS by unpacking the core elements of its target explanandum and the explanatory role it posits for the sentiments—namely, “non-inferential evaluative knowledge”, what it is to “count as having” this knowledge, and being a “necessary condition of the possibility” of this knowledge. A consequence of clarifying the nature of these components will be a differentiation of TS from other forms of sentimentalism presently on offer. Antti Kauppinen, esp. in his (2014) and (forthcoming), has led the way in clarifying and classifying contemporary sentimentalist views, taxonomizing the field with the following categories: explanatory sentimentalism, judgment sentimentalism, metaphysical sentimentalism, and epistemological sentimentalism. Showing that TS is a distinct form of sentimentalism will involve demonstrating that TS is either consistent with the rejection of or involves something in addition to these other forms of sentimentalism. This exercise is conducted primarily in the interest of introducing

and clarifying the positive commitments of an **independent TS**. That is, I aim to clarify what TS is on its own by making it clear what the transcendental sentimentalist need and need not commit to.

In the second part of this paper, I address the question of why anyone should believe TS. I offer a brief sketch of what I take to be promising strategies for demonstrating its plausibility. Explanatorily, TS makes for an attractive position because it retains the theoretical advantages of sentimentalism while avoiding some oft-discussed implausibilities burdening other forms of sentimentalism. This is an “explanatory loveliness” argument the likes of which are common in meta-ethics. A more promising strategy for directly establishing TS involves marshaling a transcendental argument in its support. If successful, this argument would establish TS by demonstrating that *not* having a sentimental response towards x results in a person’s *disqualification* (in the normative sense) from counting as having non-inferential evaluative knowledge of x.

The line of thought pursued in this paper is mainly exploratory rather than argumentative. I do not argue in favor of TS to the exclusion of those views to which it is opposed. Each part of this exploration will touch on points of philosophical contention that it is outside the scope of this particular project to resolve. Even in my sketch of argumentative strategies, there is much that must be filled in and premises which must be further established by the transcendental sentimentalist. The contribution I aim to make is the identification of a distinct way of conceiving the relationship between sentiment and evaluative thought and a strategy of rendering this conception plausible.

I. Non-inferential Evaluative Knowledge

Transcendental sentimentalism’s explanandum is non-inferential evaluative knowledge. By “evaluative knowledge” I mean the kind of knowledge that is about the non-instrumental value of some object, action, event, or other bearer of value. For the purposes of this paper, I will focus on the moral value of actions, but I am optimistic about TS’s ability to account for knowledge of other forms of non-instrumental value, especially those referred to by “thick” value concepts. By “non-inferential” knowledge I mean the sort of knowledge that, when it occurs, is formed immediately, without deliberation, upon the object of that knowledge being presented to an epistemic agent. A paradigmatic case of this kind of epistemic state is the knowledge sourced from perception; one doesn’t normally need to consider premises in syllogistic form in an effort to deduce the color of the objects of one’s

visual experience. A sighted person with the capacity to experience color simply *sees* that the fire-engine is red; in the seeing of this object (under normal circumstances), they come to know that it is red non-inferentially.

In the context of value judgments, non-inferential evaluative knowledge of the moral wrongness of some act involves the coming to know that is right or wrong on the basis of seeing it occur or upon the consideration of it in conception. For instance, if I asked you to consider the act of a psychopath torturing innocents for the purposes of experiencing momentary titillation, insofar as you are a capable moral judge, you will immediately conclude that this is an abhorrent, morally wrong act. Though there are undoubtedly moral principles on the basis of which you could infer the wrongness of this act, your immediate judgment did not result from your first considering these principles and finding an entailment appropriate for this case. That there is such a thing as non-inferential evaluative knowledge depends on the possibility of cases like these, in which the wrongness or rightness of some act is as transparent to us as the colors of the objects in our immediate environment.¹

Kauppinen uses the term “judgment sentimentalism” to refer to a constellation of views according to which judgments of moral value are in some way *constituted* by emotional or sentimental responses or are *about* these responses. TS claims that the occurrence of one kind of evaluative judgment is conditioned by sentimental response. However, being conditioned by sentiment does not require being constituted by sentiment. This is true of all conditioning and constitution generally. For example, though a steak’s being medium-rare is conditioned by exposure to some heat source, this medium-rare steak is not constituted by that heat source. Insofar as sentiments are not ordinary belief-states (which I will assume without argument for this paper), to say that moral judgments are *constituted* by sentimental responses is to commit to a type of non-cognitivism that understands the judgment “ ϕ -ing is morally wrong” as the expression of the non-cognitive attitude towards ϕ -ing which constitutes it. Transcendental Sentimentalism is plainly inconsistent with judgment sentimentalism of this variety because TS implies that there are judgments of value that express factive cognitive states, namely non-inferential evaluative knowledge.

This point does not foreclose upon the possibility of TS being the kind of judgment sentimentalism that takes evaluative judgments to be *about* sentimental responses. I will reject this

¹ Of course the physical or mental process of acquiring these pieces of knowledge may be a difference between these cases even if the truth of what they claim is alike in transparency.

possibility in the course of the discussion of how TS employs the term “being a necessary condition of” in part XX.

II. “Counting as having” knowledge

According to Kauppinen (forthcoming), explanatory sentimentalism is the view that “sentimental reactions fundamentally explain our moral verdicts,” and “moral judgment is deep down driven by emotion.” The work of contemporary moral psychologists who advance accounts of the etiology and mechanisms of moral thought straightforwardly fits this description. Jonathan Haidt’s model of cognitive systems, for instance, understands immediate evaluative intuitions to be caused by non-reason governed affective responses. According to Haidt’s model, the occurrence of this class of moral thought is causally explained by the sentimental states of the thinker.² Prior to the advent of empirical psychology, explanatory sentimentalism is found in the work of early modern British Sentimentalists such as Frances Hutcheson, David Hume, and Adam Smith.³ Each of these thinkers advanced pictures of evaluative thought on which sentiment served as its mechanism or constituent.

Transcendental Sentimentalism, in holding that sentimental response is a necessary condition of the possibility of counting as having non-inferential evaluative knowledge, is consistent with the view that sentiments serve as a causal antecedent of moral thought. However, insofar as explanatory sentimentalism is an empirical thesis justified through the *a posteriori* observation of cases of correlations between affective response and evaluation, it does not entail TS. This is for at least two reasons. The first relates to their respective scopes; TS is a claim about non-inferential evaluative knowledge, so a complete account of the necessary conditions of having an evaluative *thought* will not necessarily exhaust or account for the necessary conditions of a particular subset of that sort of thought. That is, even if the causal story about evaluative thoughts that explanatory sentimentalism offers is correct, there are plausibly further constraints on being an evaluative thought of a particular kind. Relatedly, the second reason is that insofar as TS is a thesis about what conditions evaluative *knowledge*, empirical observation is an insufficient means for discovering the entire set of its necessary conditions. Consequently, the transcendental sentimentalist’s work differs from that of the empirical

² See Haidt (2001) and (2012)

³ See Hutcheson (1725) and (1728), Hume (1739) and (1751), and Smith (1759).

psychologists insofar as knowledge is a category of thought the necessary conditions of which cannot be discovered using purely non-normative methods.

According to the independent transcendental sentimentalism I will present, “counting as having non-inferential knowledge” is not a physical state of a subject that is a product of a causal chain, it is a normative description of the authoritative status of the knower’s reports of evaluative states of affairs. This is meant as an echo of Sellars’s famous dictum from “Empiricism and the Philosophy of Mind”, “The essential point is that in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says.” (p.169). Even if the explanatory sentimentalist were to show that every case of non-inferential evaluative knowledge was preceded by a sentimental response, this would not be evidence of TS’s truth. This is because *being an episode of knowing*, according to an independent TS, is a property that is not in the category of that which can be the product of causal forces. More clarification of this idea of knowledge will be offered in subsequent sections.

In the interest of seeing how TS requires substantiation beyond that which is appealed to by the explanatory sentimentalist, consider another example of a context in which sentiments play an explanatory role: motivated reasoning. Plausibly, a person is engaged in motivated reasoning when their desire or motivation to form a certain conclusion influences their deliberation. Cases of motivated reasoning are sometimes explained by the reasoner sincerely wanting something to be true or, alternatively, having a strong aversion to a particular conclusion. Because wantings and aversions are plausibly caused by, constituted by, or instances of sentimental response, the role of the sentiments in cases of motivated reasoning is functionally similar to the role the explanatory sentimentalist finds for them in the context of moral thought. One concern of the motivated reasoning theorist, as it is with the explanatory sentimentalist, is to discover, using empirical means, the precise role that sentiments play in causing the occurrence of their target psychological explanandum. Of course, psychological investigations, like all empirical investigations, should be informed by the conceptual distinctions and analyses uncovered by philosophy. However, the difference between these kinds of explanations and TS is that the latter says that having sentiments or emotions about the object of evaluation is an epistemic good and necessary for evaluative *knowledge*, not merely the occurrence of

thoughts about value.⁴ Consequently, to be a necessary condition of the possibility of counting as having non-inferential knowledge is to play a role in the accounting for a species of *knowing*, understood as a psychological state's positive normative status. This is a separate role from that of being a component of a model aimed at explaining the production of psychological phenomena.

This example is complicated by the fact that accusing someone of engaging in “motivated reasoning” is naturally understood as an indictment of the quality of their reasoning.⁵ For a conclusion to be the product of motivated reasoning, it might be thought, is for that conclusion to count as having a *negative* normative status. A less extreme interpretation might be that a conclusion is disqualified from counting as authoritative on account of its formation being influenced by motive. According to the latter interpretation, this motivated conclusion would *lack* normative status of any kind such that it could be made relevant in the space of reasons. However, we have come to a point dialectically where little further productive discussion can be had without an explication of “normative status”.

III. Counting as having a “normative status”

In the preceding section I referred to a psychological state's positive “normative status” (in virtue of which it counts as being knowledge) in order to differentiate which of that state's conditions are accessible by empirical observation and which are not. The implication is that the normative status of a psychological state is not something which can be observed empirically nor are its conditions for counting as having this normative status. In this section I will elaborate on this point. In an effort to understand what it is for counting as having non-inferential evaluative knowledge to be a normative status I will discuss counting having a normative status generally.

First, it is worth flagging that I prefer to use the locution “x *counts* as having [normative status x]” (rather than “has [normative status x]” or “is [normative status x]”) because it makes more explicit the role of rules and principles in its attribution. To say of something that it “counts” is to say that it is

⁴ I am deliberately avoiding the claim that sentiments help “justify” evaluative beliefs or judgments for reasons that will be made clearer when I differentiate TS from epistemological sentimentalism.

⁵ However, the epistemic permissibility or impermissibility of engaging in motivated reasoning is not a settled matter in normative epistemology. See Kunda (1990) for a discussion of motivated reasoning that does not entail an epistemic transgression.

an event, object, state, or other entity that is governed by rules or principles, often of a game.⁶ For example, to say of a basketball going through a hoop that “it counts!” is to say that this event occurs within the scope of the rule-governed game according to which this event results in at least one point for a team. This is a *normative description*, or game-relative description, of this event, as all attributions of normative status are. The polar opposite of the description of “counting as x” is “disqualified from counting as x”. To say of an event, object, or state that it “doesn’t count!” is to claim that it violates the rules or occurs outside the scope of that which is governed by the rules. Imagine now a stray basketball flying in from off court and going through the hoop during our game—this event counts for nothing because this wayward ball has no normative status in the context of our game. The ball that does count is typically called the “game-ball”. This is the object that at the beginning of the game it is agreed will be made subject to the rules governing a particular run of basketball. Pointing at this ball and saying “this is the game-ball!” is to make a normative description of it and to anoint it with its normative status.

We can think of other game-relative descriptions like “being a pawn” or “being a bishop” for a game of chess, which are normative statuses an object can have that entails it is subject to rules which dictate how a player is allowed to move it on the chessboard. Again, the point here is that insofar as these are normative statuses, their attribution is not something that is discoverable strictly through empirical observation of the object. There is nothing about the game-ball’s causal profile that is relatively different from the non-game-ball’s such that we can find it to be the one with the normative status. The difference lies in the former’s being made subject to rules, which of course is itself an event with a causal description, but, insofar as “rules” are being invoked, demands a normative description as well. The same is true of the anointing of any normative status. That is, a (perhaps implicit) grasp of the rules of the game is necessary to understand and articulate what is meant by describing something as “counting as” having a normative status.

What sort of “game”-relative description can we offer of counting as having non-inferential evaluative knowledge? Here is a plausible account: to count as having non-inferential evaluative knowledge of x is, in part, to be able to offer immediate authoritative judgments about the value of x upon being presented with x. If, for instance, you have non-inferential knowledge of the moral value

⁶ The analogy between knowledge-attributions and moves within a game will be pursued here because I find it immensely clarificatory, but caution is in order. The reader is promised that an explication of their disanalogies are forthcoming.

of a type of charitable act, you are able to offer reports of that act's value upon being presented with tokens of that act-type. Part of my recognizing you as authoritative with respect to that act-type's value is recognizing that I am licensed to infer the truth of the propositional content of your evaluative judgments about acts of that type. For lack of a better term, call these a subset of the "rules" of the space of reasons—the "game" in which we give and ask for justifications for our beliefs and actions. Your cognitive state having this normative status (i.e. being an instance of non-inferential evaluative knowledge) just is for it to be subject to these epistemic rules, and many others still. For evaluative knowledge, especially moral value, these rules also plausibly have to do with what behavioral dispositions one should form in light of their knowledge. The idea is that any person who has non-inferential evaluative knowledge really "counts" as having knowledge of this type because they not only can immediately produce a judgment but also because they are ready to act on it. Failure to form the latter dispositions is evidence of a person's lack of conviction with respect to their evaluative judgment, and we should not countenance such lack of conviction from a true non-inferential authority of evaluative matters.

The analog for the game-ball-object or pawn-object in the context of knowledge is a psychological state of the subject, understood as a natural occurrence. Though not without hesitation, I will call this a cognitive state.⁷ Just as a game-ball is a natural object with a particular normative status and a pawn is a natural object with a particular normative status, an episode of knowing is a natural cognitive state with a particular normative status.

We can restate TS with this idea built in:

"A necessary condition of the possibility of a natural cognitive state having the normative status of "being non-inferential evaluative knowledge" is that the person in this cognitive state also is having or has had a sentimental response towards the subject-matter of that cognitive state."

Again, the idea is that we may be able, with empirical observation, to discover the nature of cognitive states, but we cannot through our causal interactions with them alone observe when they have a normative status or the conditions of this status obtaining.

⁷ I do not wish there to be too much read into what is meant by "cognitive state", including any implication that such a state has representational content as one of its natural properties. The most I want to say is that cognitive states have a particular subject matter. The temptation is to say that they have aboutness or intentionality, but I want to leave open the possibility that this aboutness may be reducible to some other natural property or is itself a normative status.

The preceding account is meant to be a workable picture of the game-relative description of the normative status of “counting as having non-inferential evaluative knowledge” and the natural object that takes on this status. Defending the details of this account is not my concern in this paper; my present concern is demonstrating that this kind of account that is based on epistemic and practical principles is possible, even if its details are up for debate.

Here is the disanalogy which allows us to not regard epistemic or evaluative deliberation literally as a type of game: unlike in chess and basketball, epistemic rules or the rules of what is appropriate conduct for evaluative authorities are neither a product of convention or stipulation. Or at least if they are, it is not obvious that they are. There are no shortage of other plausible non-epistemic normative statuses that require reference to principles that are not obviously the product of convention or stipulation: counting as being an “innocent”, counting as “belonging” to a particular person, a relationship counting as a “friendship”, counting as being a “person”, counting as “being guilty” of a crime, etc.. None of these are uncontroversial cases, and some may be less obviously rule-relative statuses than the basketball or chess related statuses discussed above. One test for whether these kinds of cases are in fact *normative* statuses is whether the disqualification from counting as having them makes essential reference to moral, practical, or other kinds of normative principles. I will expand on this point in the section that follows.

In pointing out this disanalogy I do not want to concede too much. I believe that being a game-ball and being a chess-piece are normative statuses, literally speaking. The dimension along which they differ with the examples in the preceding paragraphs is robustness—the extent of their importance/relevance outside the context of a particular game. The core of what I’ve described in this section has relevance to the notion of both robust and obviously non-robust normative statuses. Apropos of this point and in the interest of tying it back to the overarching purpose of this section, consider William DeVries’s elucidation of Sellars’s position on attributions of knowledge and concept-possession. If what I have said is correct, this point applies to all attributions of normative status:

“Such descriptions presume a background of rules that define and constitute the positions, objects and moves concerned. In none of these cases is the game-relative description

analysable into or reducible to a description that makes no reference to the rules of the game: having a bishop is not analysable into or reducible to having a certain physical relation to an object of a certain physical kind.” (2005, p. 131)

Transcendental Sentimentalism is not explanatory sentimentalism for the reason expressed here.⁸ Its articulation is tied up with the normative principles/constraints governing attributions of non-inferential evaluative knowledge.

IV. A “necessary condition of the possibility of” counting as...

I have claimed that according to TS sentimental response is a necessary condition of non-inferential evaluative knowledge in a way that’s distinct from its being a causal condition. For all the talk in the previous section about knowledge and its positive normative character it might be suspected TS understands sentiments to be an *epistemic* condition, i.e. a condition of the *justification* of our non-inferential evaluative knowledge. This is not necessarily the case. Epistemological sentimentalists, again following Kauppinen’s terminology, “believe that moral justification bottoms out in sentimental response of a certain kind” (2014, p. 4). For these theorists,⁹ having an emotion is not to be gripped by an irrational episode or to merely feel a certain way; emotions are means of experiencing aspects of the world which may be inaccessible via ordinary perception. *As experiences,¹⁰ they confer justification upon beliefs that are formed on the basis of what they represent. To be in a state of anger towards a token act ϕ , for instance, is to experience ϕ as having the property of being outrageous—a belief that this ϕ -ing is outrageous is thus prima facie justified on the basis of having this anger-experience. The analogous case is that my belief that the fire engine is red is prima facie justified on the basis of having a perception of the redness of the fire-engine.* Of course, a hurdle for epistemic sentimentalists is to specify under what conditions emotions accurately represent, as it is not difficult to come up with cases in which someone’s anger seems to be inappropriate because the object

⁸ DeVries makes a stronger claim than the transcendental sentimentalist needs to make in rejecting the *reducibility* of normative description along with its non-analyzability into non game-relative description. The transcendental sentimentalist can remain agnostic on the synthetic reducibility of norms.

⁹ See Prinz (2004), (2007); Goldie (2007); Roeser (2011); Tappolet (2011), (2016); and Kauppinen (2013).

¹⁰ Depending on the epistemic sentimentalist, these emotional “experiences” are cashed out either as literal perception, on analogy with perception, or non-perceptual intuitions. I use the neutral term “experience” to cover all of these possible cases.

of their anger is not, in fact, outrageous. Count my (pre-caFFEinated) feelings about morning traffic among these kinds of cases.

Transcendental sentimentalism is likely entailed by many forms of epistemological sentimentalism. That is, insofar as “being justified” is a property that all forms of knowledge essentially have, the epistemological sentimentalist will likely concur that justification-conferring sentiments are “necessary conditions of the possibility of” some evaluative knowledge. However, I believe the most plausible form of TS is held independently of, and accompanied by a rejection of, any justification-conferring role for sentiments. This is a difference from independent TS’s relationship to explanatory sentimentalism, a position with which it could align itself without theoretical cost. On independent TS, sentiments are not the sources of moral knowledge; they are psychological states which lack the epistemic power to confer justification. A consequence of this is that according to TS it is not a legitimate justification of a moral belief to make an avowal of one’s emotional state. For TS, this is not necessarily because emotions lack the proper propositional structure to serve as premises, but because they are irrelevant to the justificatory task. Such a view coheres with intuitive thoughts like, “Irrespective of how I might feel about the matter at hand, I should believe that murder is wrong and that it is wrong for me to murder.”

So far I have said that according to TS sentiments do more than serve as causal antecedents of evaluative thought, emphasizing their role in conditioning the normative status of counting as having non-inferential evaluative knowledge. I claimed this on the grounds that counting as having knowledge is a state of a subject that is not the effect of a causal chain of events, but rather a normative status that is not the subject-matter of empirical description. I have also said that, according to TS, sentiments need not play the normative role found for them by epistemic sentimentalists—that of conferring justification upon evaluative beliefs. What, then, is it for sentiments to be necessary conditions of the possibility of a kind of knowledge in a way distinct from that of a cause and that of a justification-conferring mental state?

The answer, too simply put, is that the sentiments’ role is that of a **transcendental condition**—i.e. a condition that figures into an *a priori* explanation of the possibility of a conditioned phenomena. Importantly, transcendental conditions are discoverable via armchair reasoning that takes

the conditioned as given and seeks to establish what is necessary for it to obtain. In the final section of this paper, I sketch a transcendental argument with this form aimed at establishing transcendental sentimentalism. In what immediately follows I illustrate the general nature of transcendental conditions by example, relying on ground already covered by focusing on transcendental conditions of counting as having certain normative statuses.

In the previous section I suggested that having knowledge was like being a pawn in a game of chess and like being the game-ball in a game of basketball. I claimed that these statuses were similar insofar as their attribution involves regarding whatever has that status as being subject to or being governed by some set of rules or principles—i.e. they are all alike in being normative statuses. However, there is another similarity among all these cases: there are necessary conditions of the possibility of counting as having these normative statuses. For an object to count as being a game-ball in basketball, it is necessary that this object is manipulable, that it has a circumference that is less than that of the game-hoops, that it is made of solid material etc.. Having the status of being the game-ball entails the possibility of being made subject to the rules of the game by its players. What I just enumerated are some of the physical conditions that must obtain for a natural object to be made subject to the rules of basketball. The point is that the object couldn't have been made subject to the rules of basketball had the ball been 80 inches in circumference and made of molasses. Similarly, the necessary conditions for a natural object to possibly count as a pawn in a game of chess, this object must be smaller than the individual tiles on the chess board, it must be moveable relative the board and the other pieces, etc.. If the object was larger than the chess board, its position in the game would be indeterminate, making it impossible for the rule-relative description of “being a pawn” to apply to it in a game of chess. These are transcendental conditions.¹¹

¹¹ It might be wondered: if transcendental conditions are all physical conditions, the kind which can be observed in the object with normative status, why isn't the discovery of these transcendental conditions just a matter of ordinary empirical investigation? So construed, transcendental sentimentalism would collapse back into a type of explanatory sentimentalism. The discovery of transcendental conditions can result from empirical observation, but importantly, it is not *necessary* that their discovery occurs in this way. Insofar as we can articulate the rules of a game, we can discover the transcendental conditions of objects counting as having a normative status in the context of that game. Imagine a game I am making up right now involving two trees and two people with one rope each. Each person is assigned a tree and the goal of this game is to be the first to tie the rope to the top branch of each tree, starting with the tree they are assigned. A transcendental condition of the possibility of a tree counting as a “game-tree” in this game is that it is climbable, for if it weren't the game couldn't be played. We're able to discover this condition by considering *in conception* the conditions necessary for this game's components to count in the game. This act of conception is possible because we are able to stipulate the rules of the game that make it possible for someone to play.

As we will see in the transcendental argument I sketch later in this paper, counterfactuals are necessary for determining and articulating transcendental conditions. Were the ball or chess piece too big, no basketball or chess could be played, for example. However, like in the case of explicating “normative status”, these game examples are limited by the fact that they are determined by convention or stipulation. Luckily, there are less artificial examples to be mined. Consider the normative status of counting as a friend or a relationship counting as a friendship. Insofar as you are willing to grant me that there are principles governing friendship and there is such a thing as being normatively disqualified from being a friend to someone, it is sensible to ask what the transcendental conditions of counting as a friend are. Importantly, this is not the same as asking what the principles of friendship are or about the rules for how to be a good friend.¹² You may have no trouble knowing who your friends are and know when to strip someone of that status when they’ve wronged you, but it is a separate task entirely to specify precisely what transcendental conditions must obtain for friendship to occur.

Plausibly, games which are played according to conventional rules are discoverable by empirical observation. If you watched enough games of basketball, you could come to articulate these rules on the basis of this observation. This is possible because, plausibly, conventional rules are nothing over and above conditioned agreement on which behavior is sanctioned or prohibited in particular contexts. But these rules also exist in rule books for our consideration, and it is not necessarily by observation that these written rules are articulable.

The transcendental conditions of non-inferential evaluative knowledge are discoverable by us insofar as we are able to apprehend the epistemic rules governing the space of reasons. However, this act of apprehending epistemic rules is different from our choosing their content or observing patterns of human behavior. That is, these rules are neither stipulated nor mere conventions, or at least it would be a surprise to learn that they are given that we treat them differently from rules which are uncontroversially stipulated or conventional. We do not permit individuals to stipulate their own rules of rationality, nor should we think that epistemic principles are empirical generalizations of observed behavior. In support of this latter point, consider that an idea that is perfectly comprehensible is one of living in possible world in which circumstances are such that vast majority of people exhibit deliberative behavior that is prohibited by the principles of rationality (in fact, some of us may suspect we are currently living in a world like this).

Despite its being grounded in the apprehension of rules not discoverable by observation, transcendental sentimentalism can cohere with and productively learn from an empirical investigation into sentimental response. This is true for all transcendental conditions. For example, I claimed that there was a transcendental condition of “manipulability” upon counting as a game-ball in basketball. Absent any further empirical investigation into the physical nature of manipulability, we would know very little about the transcendental conditions of being a game-ball, other than perhaps what a thinly articulated functional property can reveal on its own. If the claim of the transcendental sentimentalist is correct, there is plenty they can learn about the nature of the transcendental condition of non-inferential evaluative knowledge from empirical psychology. This is because, like manipulability, sentimental response is a natural occurrence, presumably with an evolutionary origin and a particular physical manifestation. However, that sentimental response is a transcendental condition is discoverable only by considering the rules constitutive of the space of reasons that determine the normative status of cognitive states. The apprehension of these rules is not made possible by observation, nor by stipulation, but by normative deliberation.

¹² Like with many normative statuses, the principles or rules governing friendship are normally, at best, implicitly understood by the parties involved.

Luckily, friendship was a target of inquiry for one of our greatest philosophers. In the *Nicomachean Ethics*, Aristotle argues that “friendship”¹³ is, among other things, about the mutual recognition of the feeling of good will between two parties. On the basis of this observation and previous argumentation he says, “We conclude, therefore, that to be friends men must have good will for one another, must each wish for the good of the other... and must each be aware of one another’s good will.” (1156a4-5). Questions of interpretation inevitably arise when reading Aristotle, but a natural interpretation of this passage is that Aristotle is making explicit for us some (or perhaps the complete set) of the principles governing relationships that count as friendships. We might quibble with Aristotle on the details of his explication, but the point here is that it is not obvious that he is conveying the rules of application for an obviously conventional term or his own term of art. In the course of his discussion on friends, he seems to think something of real value is at stake, and that it is worth getting clear on it.

On the basis of Aristotle’s explication, we can easily reconstruct some of what he might take to be the transcendental conditions of counting as being a friend to someone: each party in this relationship must have the capacity for good will and must have the capacity to be aware of other’s good will towards them. Having both of these capacities is a necessary condition for the possibility of counting as a friend. If I lacked one of these capacities, or if I lost these capacities, I would be disqualified from counting as a friend. This is not a simple restatement of Aristotle’s principles of friendship; what is introduced in these transcendental conditions is the (admittedly thin) notion of “capacity” and the necessity of their obtaining. Of course, the more we discover about the psychological (i.e. causal) conditions of the human possession of these capacities, the better we can articulate the nature of friendship’s transcendental conditions. In lieu of “the capacity for good will and “the capacity to be aware of other’s good will towards them” we might substitute a more psychologically-informed rendering like “a capacity for emotional intelligence.” The *transcendentality* of these conditions lies in the *a priori* nature their initial discovery as necessary for the counting as having a normative status. Their discovery depended on Aristotle’s articulation of the principles of friendship that determine its genuine instances, not on the observation of psychological or sociological states of affairs. We could have put every self-reported “friendship” and the person’s involved under

¹³ There is some question as to whether “friend” or “friendship” in Aristotle’s use is the same as our contemporary concept of “friend”. I take the concepts to be sufficiently similar to work as an example in this context. See (CITE SEP ARTICLE).

the figurative microscope and still never have arrived at friendships' transcendental conditions if we had no grasp of the normative nature of this kind of relationship, i.e. if we had no grasp of the rules which govern the "friendship-game."

Again, the adequacy of Aristotle's account of friendship is not at issue. The point is that his principles of friendship appear to have a more robust standing than the rules of basketball or chess. Consequently, friendship itself does not seem to have importance only in the context of playing a game. However, like the rules of basketball and chess, on the basis of its principles we can uncover the transcendental conditions of the possibility of a relationship having this normative status.

For the transcendental sentimentalist the manipulability of the ball, the relative size of the chess-piece, and the capacity for emotional intelligence are all analogs for having a sentimental response. That is, without having a sentimental response towards some object, a person could not count as having non-inferential evaluative knowledge of that object. The former is a transcendental condition of the latter if TS is 1) true and 2) discoverable from the armchair on the basis of a grasp of some of the epistemic and practical principles governing attributions of non-inferential evaluative knowledge. Again, these principles have to do with how the game of attributing and justifying non-inferential evaluative knowledge should be played. For knowledge generally, these principles have to do with what we are licensed to infer and the justifications the knower is expected to be able to offer. For evaluative knowledge these principles also plausibly have to do with what behavioral dispositions one should form in light of their non-inferential judgment.

V. A "necessary" condition...

For the transcendental sentimentalist, in what sense are these conditions "necessary"? On a kind of view like independent TS, a normative status like "being the game-ball" is not *analyzable* in terms of its transcendental conditions. Put another way, the transcendental conditions of counting as having a normative status are not *definitionally* necessary. Evidence of this is garnered by a simple test of conceptual confusion through contradiction. The claim "a game-ball in basketball must be manipulability and its circumference must fall within a certain range relative to the size of the hoop" is a proposition the denial of which does not commit one to an obvious logical contradiction. Alternatively, consider the assertion, "Bob's ball is this game's game-ball but it's not the ball we'll be

playing with during this game.” This claim involves the negation of the truism (albeit a conventional one) that a game-ball is the object that is played with in a game of basketball. Anyone, like our assertor, who sincerely makes a claim committed to the negation of this truism is apparently *conceptually* confused, just as they are if they assert “It was courageous for the sun to rise this morning” or “Redness has three edges but blueness has two.” Someone who sincerely makes a claim committed to the negation of a statement of the *transcendental* conditions of being a game-ball is surely incorrect, and if I am right this incorrectness is demonstrable using a priori argument, but they are not obviously committed to a *logical* or *conceptual* error.¹⁴

Insofar as there are transcendental conditions that are discoverable on the basis of stipulated or conventional rules, the necessity involved in transcendental claims is not always of an absolute or “for-all-time” variety. If the epistemic, practical, and evaluative principles governing counting as having non-inferential evaluative knowledge were absolute, then this would plausibly render their transcendental conditions absolutely necessary. Also relevant is the subject-matter of evaluative knowledge; if instances of having value were objective, absolute, and timeless, this would further constrain the conditions under which we could know them. However, the metaphysical status of these

¹⁴ For these reasons, transcendental sentimentalism is not the kind of judgment sentimentalism which commits to an analytic reduction of evaluative thought/judgment to sentimental response. This type of view, unlike non-cognitivism, may be made consistent with TS. For instance, neo-sentimentalist and fitting-attitude theories of value claim that judgments of what is of positive value are analyzable in terms of the appropriateness of pro-attitudes towards the bearer of value. For instance, to judge of something that it is desirable is to judge that it is appropriate or fitting for someone to desire it. Depending on the view, the non-inferential knowledge of what is an appropriate object of our attitudes may logically entail the occurrence of a sentimental response towards that object (i.e. TS). See XXX for examples of these type of views. Other analytic reduction programmes include those which claim that evaluative beliefs or sentences make reference to actual sentimental responses of the evaluator or the sentimental response of an ideal judge. Sentimentalists committed to the former kind of view include XXXX, Sentimentalists committed to the latter kind of view include XXXX. Transcendental sentimentalism is distinct from both of these camps.

These analyses of value are controversial. Happily, defending an analysis of value need not burden an independent TS. As an aside, I will mention that on the view I prefer, the thin concepts of “good” and “bad”, as well as the evaluative component of thick concepts (e.g. “courageous” and “cowardly”), resist analysis into other concepts, even though the non-inferential knowings whose content they figure in have numerous conditions, including sentimental response. However, the proper analysis of these terms is simply orthogonal to the claim that sentimental response conditions the possibility of a subject counting as having non-inferential evaluative knowledge.

principles or value itself is not of present concern;¹⁵ I have only suggested that they may be disanalogous with game-rules in virtue of being not *obviously* the product of convention or stipulation.

Transcendental claims of necessary conditions of possibility are *necessary* simply in virtue of being expressions of what *must* obtain for a conditioned object to count as having the status it has. Because I have said they are not definitionally true these claims are plausibly categorized as “synthetic” judgments. Because their truth is discoverable from the armchair and this discovery relies on a grasp of rules or principles that are not detectable via empirical observation, these claims are plausibly categorized as *a priori* judgments. Whether these synthetic a priori judgments are all relative to our communal or historical perspective is a question for another day.¹⁶

VI. Argumentative Strategies

Even if all that has been said so far is coherent, an obvious unanswered question remains: why think that *sentimental response* is a necessary condition of the possibility of counting as having non-inferential evaluative knowledge? In this section I will sketch two argumentative strategies aimed at establishing TS. The first of which will rely on the explanatory attractiveness of TS. The second of which will be a form of transcendental argument aimed at proving TS directly.

VI.i Explanatory Attractiveness

I began this chapter by stating that the interesting philosophical work of sentimentalism lies in its specifying the explanatory relationship between sentiments and evaluative thought. The variety of sentimentalisms mentioned so far represents the variety of possible ways of conceiving this relationship. The relative explanatory attractiveness of each of these variants can be measured against each other by comparing and weighing how much they explain vs. the implausibility of their

¹⁵ Because this not something that needs to be addressed in the course of articulating TS, TS is distinct from metaphysical sentimentalism. Whereas certain forms of judgment sentimentalism advance claims about the reducibility of value-concepts, metaphysical sentimentalists advance claims about the reducibility of value-properties. According to these theorists, facts about value are nothing over and above facts about sentimental responses. While many sentimentalists offer a package view about the conceptual and metaphysical reducibility of value to sentiment, they are logically distinct theses. Even if value isn't analyzable in terms of sentimental responses, its metaphysical constitution may be such that a synthetic reduction of the former to the latter is correct. An example of this type of view is Prinz's speaker subjectivism, on which an action is morally wrong “just in case there is an observer who has a sentiment of disapprobation toward it.” (2007, p. 92). Other metaphysical sentimentalists believe that value is grounded in sentiments that are not always actual but which are possible under ideal conditions.

¹⁶ I suspect they are not.

consequences. By giving each position credit for the former and docking them credit for the latter, we arrive at what David Enoch has termed the “plausibility points” of each.¹⁷

Generating a full ledger of the plausibility points of every sentimentalism is a task for a longer work, but surveying oft-discussed attractions and implausibilities of each variant suggests that TS acquits itself nicely by comparison. Like explanatory sentimentalism, TS can easily make sense of the close relationship between motivation and disposition to make moral judgments. It can also explain moral development and moral expertise in terms of the proper conditioning of our sentimental responses. Insofar as there is qualitative “what-it’s-likeness” associated with sentiment, TS can also account for the phenomenology of deeply-felt evaluation. The ability to cohere with plausible evolutionary etiologies of our evaluative tendencies also numbers among TS’s attractions.

As discussed above, because TS is a distinct form of sentimentalism it need not require an expressivist semantics of evaluative judgments, need not require emotion to be epistemically on a par with perception, nor does it entail that moral judgments are merely subjectively true. It is the task of sentimentalist theorists to convince us that the perceived implausibility of each of the requirements of their respective position is not dispositive of its falsity. Though each requirement may end up proving true, absent any decisive proof, they cost the plausibility of the sentimentalisms of which they are a consequence. This cost of plausibility points must be made up for by each sentimentalism’s ability to productively explain evaluative phenomena. The conclusion of this form of argument is clear: TS, unburdened by such requirements, and capable of explaining what sentimentalism sets out to explain, is clearly in the black with respect to plausibility points relative to its sentimentalist competitors.

A complete argument of this variety would have to account for TS’s implausibilities, of which there are potentially many. It would also have to reckon with the defenses offered by sentimentalists of the supposed implausibilities mentioned above. Still, I hope this brief discussion is indicative of the viability of this strategy in making the case for TS’s truth.

VI.ii Transcendental Argument

A potentially more satisfying approach involves a transcendental argument aimed directly at establishing TS. Transcendental proofs are many different things to many different philosophers, but

¹⁷ Enoch (YEAR)

what is generally agreed upon is that they begin with a given phenomenon and attempt to derive conclusions about what conditions that phenomenon, commonly to an anti-skeptical end. In the case of TS, what is conditioned is non-inferential evaluative knowledge. For a transcendental argument like TS to get off the ground, the transcendental sentimentalist needs to have convinced us that this kind of knowledge is possible. Once this common ground is agreed to, the next task is establishing that sentimental response is necessary for it to actually obtain. This argument will not have the same anti-skeptical stakes as the classic transcendental proofs that are familiar from Kant or Korsgaard, but plausible reconstructions of the latter should share this argument's formal characteristics.

Drawing conclusions about necessary conditions for the possibility of counting as having a normative status is, if my previous examples work, a relatively intuitive task in many cases. Much of the preceding discussion relied on the fact that we have an implicit understanding of the transcendental conditions of counting as having a normative status in the context of games. My recommendation is to look to these examples once more in an attempt to make explicit what kind of implicit reasoning we are engaging in when we intuit these transcendental conditions. I claimed that a transcendental condition of the possibility of counting as a pawn in a chess game was that this piece was both moveable and smaller than the individual tiles on the chess board—in virtue of what is this claim correct? It's tempting to say “things just have to be that way for pawns!”, but while perhaps a viable retort in the context of setting up a game of chess, this is not a suitable premise for inclusion in an argument, since this would just be a restatement of the necessity that is in question. A way forward is revealed by asking “What would the situation look like if these conditions did not obtain?”. The answer for a would-be pawn is that it would be a large immovable object that takes up more than one tile of the chess board. Such an object, were we to try to treat it as if it were a pawn, would render the game unplayable. That is, it would make it impossible for the rules of chess to be followed. Insofar as this piece violates the rules of chess or, more accurately, cannot be made subject to its rules for how a pawn should behave, it is disqualified from counting as a pawn.

It is this notion of disqualification that should be employed in the transcendental argument aimed at establishing TS. An understanding of cases of disqualification from counting as having a normative status can be drawn on in an attempt to formulate conditions under which such disqualifications are prevented. The complete set of the necessary conditions under which all

disqualifications from having a normative status are prevented turn out to be the complete set of transcendental conditions of counting as having that normative status. This idea is made up of a number of moving parts which deserve careful tracking:

a_n: A case in which an object is disqualified from having the normative status N . (E.g. the case in which the would-be pawn is bigger than the chessboard tiles).

A: The complete set of cases in which an object is disqualified from having the normative status N . (E.g. all the possible ways in which a would-be pawn can be disqualified from counting as a pawn).

b_n: The condition under which the particular case of disqualification a_n is prevented. (E.g. the condition of the pawn-object being smaller than the individual tiles on the chessboard).

B: The complete set of conditions that, when obtaining together, prevent all of the disqualification cases in A from obtaining. (E.g. all of the conditions relating to the pawn-object that are necessary for it to be prevented from being disqualified from counting as a pawn).

B is the complete set of transcendental conditions for counting as having the normative status N . This complete set is potentially infinite, given that there are potentially infinite ways in which an object can be disqualified from counting as having a normative status. Happily, enumerating a complete set of transcendental conditions is not our goal, nor is it clear why anyone should have that goal. The important point it is necessary that b_n is a member of B . Let c_n be a condition that is not a member of B . It would be unnecessary that the c_n obtain for an object to count as having the normative status N because c_n 's obtaining does not prevent any cases of disqualification. An example of this kind of condition in our chess example is the condition of the pawn-object being made of wood. This is not a necessary condition of the possibility of counting as being a pawn in a game of chess—according to the

argument-form I'm suggesting, this is because *not* having been made of wood would not have disqualified this object from counting as a pawn.

Applying this strategy to TS requires being able to answer the following question: when are we disqualified from counting as having non-inferential evaluative knowledge of *x*? Producing an answer to this question will rely on our (sometimes implicit) grasp of epistemic principles. Let “*x*” stand for some existing object. One plausible entailment of epistemic principles is that when a person is completely unaware of *x*, it is impossible for them to have knowledge of any of *x*'s properties. Therefore, unawareness of *x* is disqualifying of counting as having knowledge of *x*'s properties. That is, even if they somehow guessed right about them, this unaware person cannot possibly be a reliable authority on the matter of *x*'s properties. For this disqualification to have been prevented, this person's awareness of *x* must have obtained. Using the terminology of the philosophy of mind, this awareness is plausibly constituted by a person being in a cognitive state which has the subject-matter “*x*”. Our conclusion is thus that being in a cognitive state with the subject-matter “*x*” is a necessary condition of the possibility of counting as having knowledge of *x*.

Of course, mere awareness is not *sufficient* for counting as having knowledge. After all, there is no shortage of cases in which a person believes or is aware of something but should not be treated as an authority on it. I have previously implied that *knowing* is to be in a state conditioned by this awareness (i.e. to be in the cognitive state which constitutes it) and for this state to count as having a certain normative status. Analogously, *being a pawn* is to be an object that is moveable and of a certain size and for this object to count as having a certain normative status.

This case may sound trivial, but the point is to take note of its form which may be of use in the context of an argument for the necessary conditions of the possibility of non-inferential evaluative knowledge. I suggest that this useable form of transcendental argument is as follows:

- 1) Non-inferential evaluative knowledge is possible.
- 2) For a person to have knowledge of any kind is for them to be in a cognitive state with a certain normative status.
- 3) For a person to have non-inferential evaluative knowledge is for them to be in a cognitive state with a certain normative status.

- 4) There is a set of possible cases "A": for each member of A " a_n ", a_n obtaining is sufficient for a cognitive state's disqualification from counting as non-inferential evaluative knowledge.
- 5) a_1 is one such member of A: it is the case in which a person is disqualified from counting as having non-inferential evaluative knowledge of x because they do not form or reinforce any approbative or aversive dispositions upon being presented with x. (In other words, they are unmoved by x).
- 6) b_1 is the condition under which disqualifying case a_1 is preventing from obtaining. This is the condition that a person has or has had a sentimental response to x.
- 7) It is necessary that b_1 obtain for it to be the case that no member of A obtains.
- 8) b_1 is a necessary condition of the possibility of non-inferential evaluative knowledge.

Consider the following concrete case that tracks this line of reasoning: Sam cannot count as having non-inferential knowledge of the wrongness of stealing, for if he did, he would have formed the disposition to not steal the pie when faced with the temptation to steal it. We can stipulate that the evidence is decisive that he did not form such a disposition because he, in fact, stole the pie under conditions that would not have prevented this disposition from actualizing. This lack of knowledge must be explainable in some way by the occurrence of a condition that disqualified him from being a non-inferential knower with respect to the wrongness of pie-stealing. It is explainable by the fact that Sam lacks the proper conditioning to have a negative sentimental response to his intention to steal the pie. Not having a negative sentimental response disqualified Sam from counting as a non-inferential authority on the wrongness of pie-stealing and explains Sam's knowledge deficit. Having a negative sentimental response is a necessary condition of the possibility of counting as having non-inferential knowledge of the rightness and wrongness of pie-stealing. That is, this necessary condition obtaining (the sentimental response) is the means by which Sam could have been prevented from being disqualified from counting as having non-inferential evaluative knowledge.

That this section is only a sketch of this transcendental proof is a consequence of this paper's being exploratory rather than argumentative. For this proof to work, the transcendental sentimentalist must not only *explicate* each part of their position, as I have done in the first part of this paper, they

must *argue* for it. That such argument is necessary is evident from the controversial nature of many of the premises presented above. Premise 1 will be resisted by rationalists who reject intuitionism about value. Premise 2 will be resisted by those who think we can offer purely empirical descriptions of states of knowledge. Premise 5 will need to be established on the basis of the epistemic and practical principles governing non-inferential evaluative knowledge—debates over internalism and externalism about moral motivation will have application here.

The contributions that I hope to have made are 1) the illustration that this way of conceiving the explanatory relationship between sentiments and evaluation is distinct from other sentimentalisms on offer, i.e. the illustration of an independent TS; and 2) the demonstration that this transcendental sentimentalism should be treated as plausible in virtue of the fact that there are two promising lines of argument that could be marshaled in its favor.

Bibliography

- Aristotle (2012), *Aristotle's Nicomachean Ethics*, Robert C. Bartlett, and Susan D. Collins (eds/trans.), Chicago: The University of Chicago Press.
- DeVries, W. A. (2014). *Wilfrid Sellars*. Routledge.
- Goldie, P. (2007), "Seeing What Is the Kind Thing to Do: Perception and Emotion in Morality," *dialectica*, 61(3): 347–361.
- Haidt, J., (2001), "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review*, 108: 814–34.
- Haidt, J., (2012), *The Righteous Mind*, New York: Pantheon.
- Hume, D., (1739), *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd rev. edn., P. H. Nidditch. Oxford: Clarendon Press, 1978.
- Hume, D., (1751), *An Enquiry Concerning the Principles of Morals*, in H.D. Aiken (ed.), *Hume's Moral and Political Philosophy*, New York: Hafner Press, 1948.
- Hutcheson, F., 1725, *An Inquiry Into the Original of Our Ideas of Beauty and Virtue in Two*

- Treatises*, W. Leidhold (ed.), Indianapolis, IN: Liberty Fund, 2004.
- Hutcheson, F. , 1728, *An Essay on the Nature and Conduct of Passions, With Illustrations on the Moral Sense*, A. Garrett (ed.), Indianapolis, IN: Liberty Fund, 2002.
- Kauppinen, A. (2013). "A Humean theory of moral intuition." *Canadian Journal of Philosophy*, 43(3), 360-381.
- Kauppinen, A. (2014), "Moral Sentimentalism", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/moral-sentimentalism/>
- Kauppinen, A. (forthcoming), "What Is Sentimentalism? What Is Rationalism? Commentary on Joshua May," *Behavioral and Brain Sciences*
- Kunda, Z. (1990). "The case for motivated reasoning." *Psychological bulletin*, 108 3, 480-98 .
- Kraut, Richard, "Aristotle's Ethics", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/aristotle-ethics/>.
- Prinz, J., 2004, *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- Prinz, J., 2007, *The Emotional Construction of Morals*, New York: Oxford University Press.
- Roeser, S., 2011, *Moral Emotions and Intuitions*, Basingstoke: Palgrave Macmillan
- Sellars, W. (1956). "Empiricism and the Philosophy of Mind". *Minnesota studies in the philosophy of science*, 1(19), 253-329.
- Smith, A., (1759), *The Theory of Moral Sentiments*, K. Haakonssen (ed.), Cambridge: Cambridge University Press 2002.
- Tappolet, C., 2011, "Values and Emotions: Neo-Sentimentalism's Prospects," in Bagnoli (ed.) 2011: 117–134.
- Tappolet, C. (2016). *Emotions, value, and agency*. Oxford University Press.