

***Adducing Seminal Works of Korsgaard and Williams
as Precedents for D&J's Evaluative Sensibilities***

Robert M. Walker

1. On how (only) Sensibilities solve the Instability Problem of Affect

1.1: What are the Sentiments, the Sentimental Values, Judgments of Sentimental Value, and the Response-Dependency Thesis of Sentimentalism?

We will start from the following excerpt on [DS, pp. 589-90] which addresses the significance of the topic to be discussed:

“Sentimental values play a crucial role in the human mental economy which, we contend, renders skepticism about them moot; one might say that human nature forces us to take them seriously. People care deeply and (we think) ineliminably about what is shameful, disgusting, or worthy of pride [for instance]. We cannot do without these evaluative concepts, because of the universality of the sentiments and our common need to regulate them with standards. Of course individuals and societies differ vastly in their senses of humor, honor, purity, and the like. But the standards these values impose are profoundly important to us all, in deciding how to live. Doubts about their metaphysical status do not loosen their hold on our moral psychology.”

According to [DS, p.586], among the emotions, the ***sentiments*** form a narrow subclass consisting exclusively of *anthropologically universal* responses such as “amusement, anger, contempt, disgust, fear, guilt, pride, and shame.” ***Sentimental values/value concepts*** are values that our species describes—in the form of a certain kind of evaluative belief—with the aid of the sentiments; ***(sentimental value) projection*** is my operative term for the cumulative mental process by which members of our species descry (track/locate) sentimental values. By way of projection, our species describes the funny via amusement, the

shameful via shame, the pride-worthy via pride, and so on. The *universality* stipulation means two things: first, according to evolutionary precedent, virtually all human beings experience these emotions; and second, sentimental values—along with the process of projection and its chief effects, judgments of sentimental value—are *intersubjective*, occupying an ineliminable role in our species' folk psychology and mental economy.

The chief effect of projection is a kind of evaluative belief, which D'Arms and Jacobson call judgments of sentimental value. A ***judgment (of sentimental value)*** is an evaluative belief grounded in affective perception ("perceiving by feeling"), that is, in a person's responses. [DS, p.592] We differentiate affirmative- and negative judgments of sentimental value. Relative to some object, one may formulate an ***affirmative*** judgment about the funny (resp., the shameful, the fearsome, and so forth) that is somehow based on expressing amusement (resp. shame, fear, etcetera) at that object. The following example is an affirmative judgment: "The neutron pun is funny because it amused me." Relative to some object, one may formulate a ***negative*** judgment about the funny (resp., the shameful, the fearsome, and so forth) that is somehow based on the *absence* of amusement (resp. shame, fear, etcetera) directed at that object. The following example is a negative judgment: "The neutron pun is not funny because it did not amuse me." The presence (absence) of the suitable affective response is stipulated to be a *necessary* condition—though, importantly, **not** a *sufficient* condition—for formulating an affirmative (negative) judgment of sentimental value.¹

¹ This is intentional. According to a particularly jejune form of sentimentalism, if I say an object "counts as V" (or V is a feature of said object, or the object is V (by my lights)), then I mean simply that the object is an "S-solicitor," and an ***affirmative (resp., negative) V-judgment relative to an object O*** would simply look like: "O is (not) V because I felt (did not feel) S at O." In particularly straightforward cases, jejune sentimentalism matches up well

Whether jejune or robust, any form of sentimentalism hinges upon the response-dependency thesis. If we fix a sentimental value *V*, and *S* is the sentiment that describes that value, then the ***response-dependency thesis of sentimentalism*** asserts that we human beings crucially depend on select patterns of feeling *S* to track two related things:^{2 3}

(1) **Patterns of projection and judgment about *V*** (How do I (resp. some number of people) tend to project “This is *V*!” or “This is not *V*!” onto my conscious experience (their intersubjective experience) of select objects?); and

(2) **Individuals’ affective perspectives on *V*** (In accordance with affective perception, what do I (resp. some number of people) tend to count as *V*?).

Our species’ folk psychology subscribes to the response-dependency thesis: the last paragraph demonstrated this relative to item (1). Moving on to (2), we consider a response-dependency scenario that the reader should be quite familiar with: suppose one person (the asserter) makes an assertion concerning another person’s (the subject) take on *V* along the lines of “This object is so what that person would find *V*!”⁴ According to common

with our species’ folk psychology. However, some (toy- and interesting WKR) cases are too sophisticated for jejune sentimentalism to handle, a fact that D’Arms and Jacobson adduce in favor of a more robust form of sentimentalism—rational sentimentalism—that can accommodate such cases.

² Throughout the paper, in the most general case we will use the letter *V* to denote a sentimental value, *S* to denote the sentiment that describes *V*.

³ Tracking/locating should be understood not in the sense of using a metal detector, but instead in the “People will make of this what they will” sense, in the sense where Hamlet says, “[There] is nothing [inherently] either good or bad, but thinking makes it so.” The relevant value concept is already seated in your mind, the data triggers the mind’s use of it, and you read—or *project*—that concept onto your conscious experience of the subject (one feature of this experience may be an affective perception of the subject, perceiving one’s self feeling/directing an emotion at that subject); when people exercise folk psychology, it is in this respect that they are seeing the goodness or badness of something.

Here, if *S* is a sentiment and *V* is the value our species typically describes via *S*, we would say: “Nothing is inherently *V* (not *V*), but I can read *V* (not *V*) onto a particular thing after first (not) feeling *S* at it.” or “Nothing is inherently (not) *V*, but I can see (by my own lights) that it is (not) *V* after first (not) feeling *S* at it.”

⁴ Such assertions are *adjacent* to judgments about *V*.

sense, any witness who buys into this assertion **believes that** the asserter, if asked to explain herself, would adduce the subject's disposition to feel V at such an object. In the comic realm, I have in mind assertions such as "*Friends with Benefits* is my girlfriend's kind of romantic comedy." or "I'll e-mail my brother a link to this YouTube video. It's the kind of thing he'd get a kick out of (that is, find funny)!" Such assertions are so pervasive in everyday life that they show up in media that propose to simulate, study, and possibly poke fun at everyday situations, such as situational comedies. In particular, such pervasiveness confirms our claim about folk psychology relative to item (2).

1.2: The Problems: the Instability of Sentimental Response and the Opacity of Judgment

From above, we recall that sentimentalism hinges upon the response-dependency thesis, a thesis to which our species' folk psychology subscribes. However, folk psychology also recognizes that our sentimental responses are prone to vicissitudes; we call these vicissitudes **instabilities (of affect)**, and human beings are familiar with several of their causes. Repeated exposure is one such cause: after hearing a joke I deem funny enough times consecutively, I stop laughing at it; after watching *Saw*—a movie I deem horrifying based on feeling fear and disgust at its "horror" sequences upon first viewing—enough times consecutively, the "horror" sequences eventually stop soliciting that familiar mix of fear and disgust. Mood effects are another source of these instabilities: for instance, suppose that feeling really downtrodden while watching a comedy you have never seen before—the lion's share of its contents, we assume, match up with your affective perspective on the

funny—*prevents* you from expressing amusement. What if you then deem the comedy unfunny (where, by hypothesis, you would deem it pretty funny if not for feeling down)?

And yet, these instabilities do not always translate to instability in sentimental value projection, or more precisely, instability in one's judgments (of sentimental value). Revisiting the repetitive exposure cases, by way of projection I could—and almost certainly *would*—still claim that the joke is funny, that select subject matter portrayed in *Saw* merits fear or disgust. In particular, our judgments are instability-resistant relative to what D'Arms and Jacobson call transparent- and translucent instabilities (of affect). ***Transparent instabilities*** are the ones whose causes are “so obvious that you account for [the cause] in judgment without second thought.” [DS, p.592-3] Thus transparent instabilities never issue in mistaken judgments. As noted above, repetitive exposure is a familiar source of transparent instabilities. ***Translucent instabilities*** are the ones that do not issue in mistaken judgments when their causes are obvious, while correcting for mistaken judgments burdens us to sustain ***reflective distance*** (as Korsgaard calls it) long enough to recognize the instability's cause and withdraw that previous judgment. [DS, p.592-3] As noted earlier, mood effects are familiar sources of translucent instabilities. Finally, ***opaque instabilities*** are the ones whose causes can only lie *wholly outside* of the agent's awareness. [DS, p.592-3] In particular, while transparent- and translucent instabilities need not push the constraints of a human being's emotive-physiological resources, opaque instabilities will. Among mistaken judgments, human beings are especially prone to upholding those issuing from opaque instabilities, doing so by adducing some reason(s) of the wrong kind whose normative force is opaque

(not obvious).⁵ Therefore, we get from instabilities of affect to a problem which D'Arms and Jacobson call the *opacity of judgment* ([DS, p.592-3]): "[Since] not all instabilities of response are obvious, not all our judgments take account of them," resulting in instabilities of judgment and mistaken judgments.⁶

We now recognize that the response-dependency thesis sets the aspiring sentimentalist with *the instability problem of affect*, the opacity of judgment being a natural consequence. According to [DS, pp.588, 593], the instability problem challenges the sentimentalist to reconcile two commonplaces of folk psychology: the existence of instabilities of affect; and, the existence of *coherent, relatively stable* patterns of affective perception in virtue of which (a) people perceive sentimental values by feeling the sentiments, and (b) the projection process is (resp., affective perspectives on sentimental value are) understood to be relatively stable.⁷ In particular, despite the existence of

⁵ See [WKR, pp.1&3]: A *reason of the wrong kind/wrong kind of reason (WKR)* is, roughly, a "consideration broadly in favor of (or against) having an evaluative attitude [that] does *not* bear on whether the object is valuable in the respect relevant to that attitude." In example, a demon offering you a million-dollar incentive to direct anger at someone else is a WKR for deeming that person deserving of anger. To say that the normative force of a WKR is *opaque* simply means that it is not obvious *how* (in favor, against) that reason bears on having the relevant attitude. By contrast, the normative force (namely pro-anger) of the WKR from the above example is obvious.

⁶ From [DS, p.597]: We observe that the instability of affect is "a feature of our dispositions that [reveals nothing] about our values;" as such, these instabilities can only serve as WKRs for judgments of sentimental value, and these judgments in turn will be mistaken, *alienated* from the relevant *sensibility* (a notion glossed in subsection 1.3). In general, a transparent (resp., transparent/opaque) instability of affect can only serve as a WKR of transparent (resp., transparent/opaque) normative force for a judgment of sentimental value.

⁷ In order to avoid the contentious topic of objectivity (whether any values have metaphysical status in the external world independent of the perceptual experience of rational beings), I propose that we interpret any claim about seeing or sensing a value "out in the world" or "in some object" as a (an at most intersubjective) claim about sentimental value projection generally, and judgments of sentimental value specifically.

instabilities of affect, folk psychology tells *against* the inference that projection (in particular, judgments) and individuals' affective perspectives are prone to vicissitudes *to the same extent* that our dispositions to the sentiments are. In fact, one goal of the discussion on the opacity of judgment earlier was to frame projection as a relatively—but imperfectly—stable process.⁸ Therefore, as D'Arms and Jacobson demonstrate, the two traditional proposals in the literature for solving the instability problem inevitably fail.⁹ How can it be solved?

1.3: The Solution: Posit the Existence of Disparate Evaluative Sensibilities

If the instability problem starts from factoids of our species' folk psychology, we must also locate the solution within folk psychology. If the central worry is that sentimental value projection might be a relatively unstable mental process, a conclusion which folk psychology tells *against*, we must adduce some commonplace of folk psychology—some *natural psychological kind*—to explain why projection is relatively stable. All that remains is to locate a suitable natural psychological kind. The obvious candidates are dispositions to the *sentiments* and *judgments* (of sentimental value); courtesy of the instability of affect and the opacity of judgment, respectively, adducing one or the other *indiscriminately* will not work.

That being said, barring the occasional instability of affect (resp., mistaken judgment), people tend to express *coherent* patterns of sentimental response (resp. patterns of judgment), “in virtue of which they can be understood as having **a distinctive perspective**

⁸ That discussion also tells against modeling an individual's affective perspectives solely on patterns in her sentimental responses or patterns in her judgments due to the respective instability issues.

⁹ See [DS, pp.90-93]: Proposal 1 is to wholly reject the “instability-resistance” of projection; Proposal 2 conflates our affective perspectives with our sincere judgments, *overestimating* the stability of said judgments to explain the stability of those perspectives.

[(coherent and *relatively stable*)] on various sentimental values, rather than just a welter of emotional responses [(resp. of judgments)].” [DS, p.593] For instance, we gain insight into a person’s affective perspective on *the funny/what counts as funny* from locating the “right” thematic patterns of her amusement, that is, the “right” thematic patterns in what makes her laugh. Of course, some degree of imprecision in such an investigation is inevitable; at some point, we can no longer deconstruct her take on the funny, ultimately chalking her take up to her sense of humor—that thing in her mind “behind-the-scenes” which we believe ultimately explains her perspective, though we *cannot* pin it down precisely.

Folk psychology recognizes that a similar story holds for any sentimental value V (S is the sentiment that describes V), that is, we ultimately chalk up a person’s affective take on V to *some* thing in her mind that “senses V” by issuing in coherent patterns of her feeling S at certain objects. [DS, p.597] D’Arms and Jacobson call such a thing an ***evaluative sensibility*** (or simply, a sensibility). In particular, we note that behind-the-scenes these sensibilities are what allow her to *perceive value through feeling*, and that sensibilities are understood to be coherent and relatively stable in folk psychology.¹⁰ Along with the sentimental values, sensibilities are natural psychological kinds: when we discuss a person’s *senses* of humor, honor, shame, and the like in everyday conversations, we are borrowing familiar colloquialisms for sensibilities from folk psychology; thus, to reiterate, folk psychology

¹⁰ From [DS, p. 594]: “Consider the claim that in order to see an object as having certain evaluative features, it is (or may be) necessary to feel some way about it. [...] ***Even those who possess the relevant sensibility often cannot articulate precisely what makes something funny, disgusting, or even contemptible. [...] Then people are prone to fall back upon the claim directly to perceive, with their feelings, the beauty, humor, disgustingness, or shamefulfulness of this but not that object.*** Whether or not we agree with their verdicts, we must recognize the ways in which the idiosyncrasies in their own patterns of response generate a distinctive perspective on these values.”

recognizes the existence of discrete, evaluative sensibilities that explain people's *coherent, relatively stable* patterns of affective perception. [DS, p.597] Earlier, I said that if the instability problem starts from folk psychology, the solution must do the same; we now see how that will work out. Proceeding in this vein, D'Arms and Jacobson contend that the only way sentimentalism will surmount the instability problem is by positing—as folk psychology already does—the existence of sensibilities “that may or may not be manifested in any particular response or judgment, but which ultimately [explain] what an agent finds shameful, disgusting, or funny—when other factors don't interfere with his judgment or response.”¹¹

I want to conclude the section with an excerpt from [DS, p.596], which gives a rough-and-ready gloss of what sensibilities are and summarizes how they—*and only they*—can help us reconcile the two heads of the instability problem, thereby solving it:

“What, then, is a sensibility? **A person's discrete sensibilities consist in facts about him (some idiosyncratic, some common to most human beings) that underlie and explain many of his dispositions to specific sentiments—though not all such [dispositions.]** Sensibilities issue in affective responses, in the first place, and it is by looking at the right responses that they are most clearly revealed. In the most straightforward sort of case, these responses provide a basis for evaluative judgments. **So ordinarily one's evaluative judgments arise out of one's sensibility by way of sentimental response.** However, not all such responses, let alone failures to respond, arise from the relevant sensibility. The simple cases we have discussed so far, based on repetition and mood effects, illustrate this point. **In what follows we will argue that the posit of more or less coherent sensibilities is necessary to solve the instability problem, while also allowing for a moral psychology that makes better sense of internal conflict, fundamental evaluative disagreement, and change of mind.”**

¹¹ This quote is from [DS, p.593]. On the same page, D'Arms and Jacobson forecast that they will attribute a certain kind of interpersonal, evaluative disagreement to differences in the relevant sensibility between parties, and “change of mind [to a sensibility's] alteration over time,” courtesy of “[repetition,] when combined with social influence and reflection.” [DS, pp.593&597] Beyond this footnote, I will not really attend to these issues.

In the next section, we will recap the highlights of Korsgaard's discussion of practical identities from Lecture 3 of *The Sources of Normativity*. Then, we juxtapose that notion with the notion of sensibilities, paying special attention to an important disanalogy between the two. In section 3, we will do much the same with Williams' minimally reflective ethical knowledge as discussed in *Ethics and the Limits of Philosophy*.

2. Korsgaard

In this section, we will discuss Korsgaard's notion of a practical identity; as we proceed, we will attend to the (dis)analogies between practical identities and evaluative sensibilities.

2.1: The reflective (self-conscious) human mind, the Problem of the Normative, and its solution—Practical Identities.

We start with an excerpt from *The Sources of Normativity*, in which Korsgaard's intent is to explain where human beings' *obligations* ("reactions to threat of disfigurement or loss/crisis of identity") come from, and why obligation exists [SN, p.113]:

"The reflective structure of human consciousness sets us a problem. Reflective distance from our impulses makes it both possible and necessary to decide which ones we will act on: it forces us to act for reasons. At the same time [...it] forces us to have a conception of our own identity, a conception which identifies us with the source of those reasons. In this way, it makes us laws to ourselves. When an impulse—say a desire—presents itself to us, we ask whether it could be a reason. We answer that question by seeing whether the maxim of acting on it can be willed as a law by a being with the identity in question. If it can be willed as a law it is a reason, for it has an intrinsically normative structure. If it cannot be willed as a law, we must reject it, and in that case we get obligation."

In the first two sentences, Korsgaard summarizes the *problem of the normative*, which follows from accepting that [SN, p.593]: (1) human beings can exercise *reflective distance*

(and scrutiny) relative to impulses—that is, to perceptions and desires;¹² (2) the reflective mind needs *reasons* to believe and to act autonomously (that is, perceptions and desires, respectively, which *survive* reflective scrutiny and are “*ratified*”);¹³ and (3) without **some** regulative/legislative thing in place, sustained reflective scrutiny might disqualify *all* impulses from being reasons. Importantly, for Korsgaard, if the problem starts from reflective distance, then the solution must do the same. The next two sentences of the excerpt above pull back to the following [SN, p.100]:

“[Reflective distance] forces us to have a *conception* of ourselves. [...] When you deliberate it is as if there were something over and above all of your desires, [some thing] which is *you*, and which *chooses* which desire to act on. This means that the principle or law by which you determine your actions is one that you regard as being expressive of *yourself*. To identify with such a principle or way of choosing is to be [...] a law to yourself.”

Korsgaard posits that all human beings have such a self-conception, which she calls a ***practical identity***, and understands it as projecting value onto your life, actions, and beliefs. [SN, p.101] Only one practical identity is necessary/unshakeable—the moral identity (under which the affirmation “I am a human being” is practical and normative), while the others are contingent, picked up through socialization. [SN, pp.120-2] Through the process of socialization, the average person develops a cumulative identity as a syndrome of practical identities [SN, p.101]: “You are a human being, a woman or a man, an adherent of a certain religion, a member of an ethnic group, a member of a certain profession, someone’s lover or

¹² According to [SN, p.93], unlike the lower animals, we humans have the capacity to “turn our attention on to our perceptions and desires themselves, [...] think *about* them, [...] distance] ourselves from them, and [call] them into question.”

¹³ Following [SN, p.93]: Perceptions impel us to believe; distancing myself from a given perception allows me to interrogate the corresponding impulse, to ask: “Shall I believe? Is this perception really a *reason* to believe?” Desires impel us to act; distancing myself from a given desire allows me to interrogate the corresponding impulse, to ask: “Shall I act? Is this desire really a *reason* to act?”

friend, and so on.” More importantly, Korsgaard stipulates that a practical identity fills an essential *legislative* role in the reflective mind, reigning in and regulating reflective scrutiny, and ensuring that *several* impulses survive scrutiny [SN, pp.101-2]: the impulses it ratifies—according to maxims that are compatible with one’s adopted identity—become reasons, while its *reactions* to impulses that it subsequently vetoes become obligations; thus reasons embody reflective endorsement, while obligations embody reflective rejection; in short, any practical identity issues in reasons and obligations (reactions).

At this point, I want to draw attention to an important analogy between practical identities and sensibilities: perhaps more so than reasons, obligations have a position relative to practical identities similar to the one that affective responses have relative to evaluative sensibilities. Recall that by the response-dependency thesis, there is one sentiment (if not several) that we crucially depend on to gain insight into the contours of a given evaluative sensibility. Korsgaard believes similarly that obligations, not reasons, are especially important for detecting the boundaries/contours of a practical identity, essentially constituting those boundaries. Obligations serve as the lines a person dares not cross for fear of protest or retaliation from others (i.e., when others ask “How dare you do such a thing?” or “Who do you think you are?”), or for fear of crisis of identity or self-alienation (when a person says, “I couldn’t live with myself if I did that.” or “If I cross that line, I’ll never go back.”). [SN, pp.101-2]

More strikingly, Korsgaard—and perhaps, folk psychology—recognizes a phenomenon of ***instability of obligation***, although it plays out somewhat differently from cases of instability of affect. According to Korsgaard, instability of obligation manifests itself

when a person makes an exception for herself in adhering to (or expressing) the obligations of a given identity; in particular, any instability of obligation is obvious to the agent, so instead of transparent, translucent, and opaque, the shades of instability of obligation are small and large. ***Small (resp. large) instabilities of obligation*** amount to the following scenario [SN, pp.102&122]: Korsgaard states that it is possible to *occasionally*/temporarily **short-circuit** an identity, in order to violate a small (resp., large) obligation to pursue a large temptation, and then turn that identity back on, so long as the person does not make a habit out of such violations; unless the person opts to cast off that identity altogether—i.e., because living up to its demands no longer matters to her, such a habit will only lead to crisis of identity and/or self-alienation; importantly, Korsgaard contends that “this [process] has a destabilizing effect on the obligation,” since the agent treats the obligation as flexible. Of course, in light of the problem of the normative, the temptation in question must first be ratified by another identity. Therefore, any instability of obligation always pulls back to a conflict of identities—either contingent versus contingent, or contingent versus moral. Additionally, Korsgaard juxtaposes the instability of obligation with the relative stability of a contingent identity [SN, p.102]: “The problem here does not come from the fragility of identity, but rather from its stability. It can take a few knocks, and we know it. The agent I am talking about now violates the law that she is to herself, making an exception of the moment or the case, which she knows she can get away with.” Thus to take the disanalogy between practical identities and sensibilities a few steps further:

- (1) Practical identities accommodate an instability problem instead of solving it.
- (2) The operation of short-circuiting one identity in favor of another is always transparent to the agent (and hence any instability of obligation is transparent), while the similar

phenomenon for sensibilities is opaque (non-obvious), and hence the resulting instabilities of affect are opaque too.¹⁴

(3) When activated, a practical identity is transparent—lying over and above your impulses, according to Korsgaard, while a sensibility is at best a marginally translucent thing (if not opaque)—lying behind and underneath certain patterns of affective response. In particular, when active, practical identities can be pinned down relatively precisely, while sensibilities cannot; practical identities have a clearly defined legislative role in the human mental economy, while sensibilities' corresponding (regulative) role is subtle, employing affective perception (resp., reasons of fit) as a lobbying mechanism to issue in judgments of sentimental value (resp., sentimental responses).

Climactic/large instabilities of obligation occur in potentially tense scenarios that test the constraints of one's available resources for adhering to one of several conflicting obligations. For example, in the church-burning sequence from *The Patriot*, one of the British soldiers hesitates after being ordered to burn down the church with all of the villagers locked inside. Describing what is going on with that soldier in terms of conflicting obligations, Korsgaard would say: "A good soldier obeys orders, but a good human being does not massacre the innocent." [SN, p.102] Thus we have a clash between his identity as a soldier and his moral identity, and the obligation he chooses to uphold—the obligation he diverts his resources and attention to—pulls back to the deeper facet of his cumulative identity; meanwhile the silenced obligation destabilizes in terms of its hold on the soldier in future deliberation. On [SN, p.122], Korsgaard describes such scenarios as the "stuff of drama, and perfectly familiar to us [all;]" you can only uphold one obligation, and the one you choose pulls back to the *more deep-seated* identity; in particular, in such scenarios one tends

¹⁴ [DS, p.596]: D'Arms and Jacobson interpret the silencing phenomenon "**as a frequent effect of conflict between distinct sensibilities and sentiments,**" while any instances of single-mindedness a person musters "**in the face of conflicting values [is] due to the propensity to have certain evaluative sensibilities short-circuited or overwhelmed by others—for better or worse.**" Courtesy of silencing, weaker reasons can circumvent stronger ones (in disanalogy with instability of obligation scenarios), and so the sentiment one ultimately expresses—and the value one describes in the moment—might be less salient.

towards a deeper understanding of who he is in the moment rather than self-alienation. In general, Korsgaard recognizes that the average person arranges practical identities into a hierarchal lattice according to the general strength of the obligations they give rise to; this process is at worst translucent, never fully opaque. Consequently, it is extremely difficult to act, and to believe, in ways that alienate you from the most deep-seated facets of your practical identity. [SN, p.102]

This starkly contrasts with scenarios where opaque instabilities of affect manifest themselves through *emotional diversion*, where behind the scenes there is intrapsychic conflict between several of a person's sensibilities. Emotions can be mutually incompatible [DS, p.600]: often enough, one's emotive-physiological resources are not suited to express two given sentiments simultaneously; examples of such pairings include fear being incompatible with amusement, and pride being incompatible with disgust. According to [DS, p.600], in such a scenario the sentiment expressed clues the agent into some of the object's features, but diverts—preempts—the agent from feeling the other sentiment at that object according to select other features; ergo, the agent cannot pay simultaneous respects to the two corresponding sentimental values which, by the agent's lights, can be projected onto—“detected in”—that object; meanwhile, behind the scenes the sentiment expressed pulls back to one sensibility that is preeminent, while the sentiment not expressed pulls back to an entirely different sensibility that the former sensibility has preempted, short-circuited, or *silenced*, without the agent's knowing.¹⁵

¹⁵ On [DS, p.600], D'Arms and Jacobson follow up with: “More generally, the presence of some psychologically incompatible affect should be taken as an obscuring factor, which suggests that **a failure to respond (with amusement, pride, or whatever) does not reveal a negative verdict (of lack of funniness or pride-worthiness) from the sensibility subject**”

In particular, for D'Arms and Jacobson, this manner of silencing is non-obvious—in disanalogy with short-circuiting a practical identity, and hence it would be erroneous to infer that the expressed sentiment pulls back to the more deep-seated sensibility; in fact, the subject of whether a “hierarchal lattice of sensibilities arranged according to affective responses” can be established is moot. It is also erroneous (and common among humans) to take the sentiment not expressed as a basis for a negative judgment of sentimental value, believing that it issues from the silenced sensibility—as such, this judgment is *alienated* from the preempted sensibility, and in assenting to it, the same can be said of the assenter. In general, D'Arms and Jacobson adduce intrapsychic conflict between sensibilities to explain this common phenomenon through which a person can be estranged from herself, or more precisely, from one of her sensibilities (see [DS, p.596]). We situate some diversionary response-mistaken judgment examples from [DS, p.603] and [WKR, pp.17-18], respectively:

(1) A friend makes a joke about your weight, a sensitive topic for you—a joke you would laugh at and deem funny had it been about someone else for whom your feelings are more or less neutral. You are so upset at the perceived sleight that your anger *diverts* the joke from making you laugh. Behind the scenes, your esteem sensibility has silenced your comic sensibility. Considering your anger and the absence of amusement it causes, your judgment (“Hey, that’s not funny!”) is alienated from your comic sensibility.

(2) A daunting bully insults you, and “you might not become angry simply because you are terrified instead; [...] your fear diverted the insult from angering you.” Between your fear and the absence of anger it causes, your judgment (“Hey, that does not hurt my feelings!”) is alienated from your esteem sensibility.

2.2: Lost in Disanalogy: An Important Closing Message

to interference [(to being short-circuited)]. [...] The] failure to feel an emotion may not reveal that this emotion is unfitting, but merely that one’s affective resources are otherwise deployed. Thus the fact that one was not amused by something **isn’t always a negative verdict of one’s sense of humor**, and in general many dispositions to lack of response obscure rather than reveal one’s sensibilities. ”

A lot has been said to show that practical identities and sensibilities are quite disanalogous. However, I want to conclude the section by emphasizing *how* they resemble one another, and *why* I considered in the first place Korsgaard's discussion of the problem of the normative (and the resulting instability problem of obligation) and practical identities to be a precedent of D'Arms and Jacobson's discussion of the instability problem and evaluative sensibilities.

Barring obscuring factors and WKR's:

(a) Just as practical identities are responsible for us bridging the gap from perceptions to belief by upgrading select perceptions to "*reasons* to believe," we must posit sensibilities to explain how humans go from affective perceptions (perceiving through feeling) to judgments of sentimental value; sensibilities upgrade select perceptions to "*reasons* to believe."

(b) Just as practical identities are responsible for us bridging the gap from desires to action by upgrading select desires to "*reasons* to act," sensibilities explain how we go from RKR-candidate considerations to evaluative attitudes; sensibilities upgrade select of those considerations to "*reasons* to feel" or "*reasons* not to feel."

3. Williams

In the first part of this section, we will discuss Williams' take on ethical knowledge relative to applying thick concepts, and his belief that reflection can destroy (the "sensibilities" relative to thick concepts that underpin) that knowledge. I will direct the reader's attention to points of analogy between Williams' discussion and the corresponding discussion of sensibilities in [DS]: in particular, I contend that the two discussions resemble one another almost identically. In the second part of this section, I will attend to the single, crucial disanalogy between the two discussions, namely that D'Arms and Jacobson do not subscribe to William's *un-Socratic conclusion* in the case of knowledge of sentimental value.

3.1: *Thick (ethical) concepts, ethical knowledge of members of the hypertraditional society, and how that knowledge is destroyed by reflection*

We start from the following excerpt of Williams' *Ethics and the Limits of Philosophy*, in which Williams recaps and explains what he calls the **un-Socratic conclusion**, namely that reflection might destroy knowledge (instead of being the sole source of it) [ELP, p.167]:

"Earlier I said that reflection might destroy knowledge, because thick ethical concepts that were used in a less reflective state might be driven from use by reflection, while the more abstract and general ethical thoughts that would probably take their place would not satisfy the [three criteria] for propositional knowledge. To say that knowledge is destroyed in such a case is [to say... that] these people once had beliefs of a certain kind, which were in many cases pieces of knowledge; but now, because after reflection they can no longer use concepts essential to those beliefs, they can no longer form beliefs of that kind. A certain kind of knowledge with regard to particular situations, which used to guide them round their social world and helped to form it, is no longer available to them. Knowledge has been destroyed because a potentiality for a certain kind of knowledge has been destroyed; moreover, if they think about their earlier beliefs, they will now see them as the [ethnographer] saw them, as knowledge they do not share."

The role that thick concepts have for Williams mirrors the role that sentimental values have for D'Arms and Jacobson. In the first place, according to Williams thick concepts include things such as *coward*, *lie*, *brutality*, and *gratitude* that are projected onto worldly, social situations in a hit-or-miss fashion—that is, a person can apply them rightly or wrongly in a given scenario. [ELP, p.140] D'Arms and Jacobson have similar thoughts regarding sentimental value projection [DS, pp.611-12]:

"We are not claiming that sensibilities are guaranteed to get matters right, or even that they are statistically more likely to be correct than are sincere evaluative beliefs. Our point is rather that they have internal connections to sentimental values which theoretical reflection does not. This fact renders such reflection prone to forms of error and confusion from which sensibilities are immune. Human sensibilities are our most important and systematic connection to the sentimental values which occupy a central place in our lives."

Moreover, Williams recognizes that people can have evaluative perspectives on a thick concept, expressing coherent patterns in applying such concepts [ELP, p.141]: “How we ‘go on’ from one application of a concept to another is a function of the kind of interest that the concept represents, and we should not assume that we could see how people ‘go on’ if we did not share the evaluative perspective in which this kind of concept has its point.”

The striking resemblance between the two discussions extends to the role of perceptual metaphors, to perceiving when an evaluative concept applies (by an agent’s lights) to a scenario versus when not, and judging accordingly. Recall from Section 1 that according to folk psychology, people are commonly understood to perceive sentimental values by feeling. William has a similar perceptual metaphor in mind for members of his **“hypertraditional” homogenized society**, a society consisting of minimally reflective people using a narrow selection of thick concepts: “The members of the hypertraditional society apply their thick concepts, and in doing so they make various judgments. [...] They have, each, mastered these concepts, and they can perceive the personal and social happenings to which the concepts apply.” [ELP, p.143] In particular, the hypertraditional folks have a local value system that pulls back to them developing a shared syndrome of “sensibilities” relative to the thick concepts in their communal repertoire. The ethical propositional knowledge of the society consists of select judgments they believe, judgments which are true according to the facts, and “[one’s belief should ‘track the truth:’] granted the way that the people have gone about their inquiries, it must be no accident that the belief they have acquired is a true one, and if the truth on the subject had been otherwise, they would have acquired a different belief, true in those different circumstances.” [ELP, p.143] In

turn, D'Arms and Jacobson do not preclude the possibility of going from sensibilities to ethical knowledge of this kind: when D'Arms and Jacobson mention, for instance, how one person's sense of humor might be sharper than another's, and so statistically the former stands to make more accurate judgments about the funny than the latter—and hence, stands to have greater *knowledge* about the funny than the other person, they are *implicitly* discussing how humans stand to get from sensibilities to ethical knowledge. [DS, p.594]

Williams distinguishes between the objectivist- and nonobjectivist models of ethical practice from the perspective of an insightful, sympathetic, but not totally identified ethnographer of the hypertraditional society. [ELP, p. 147] To the extent that the ethnographer sees the members as trying to find out the truth about value, to the extent that she exercises reflection to scrutinize the member's pre-theoretic judgments (seated in a localized value system), to scrutinize the society's thick concepts (and related "sensibilities"), and attend to the theoretical "landmines" of those judgments, she is following an "objectivist" model of ethical practice. [ELP, p. 147] On the nonobjectivist model, "she instead sees their judgments as part of their way of living, a cultural artifact they have come to inhabit (though they have not consciously built it). [... *She*] ***shall not be disposed to see the level of reflection as implicitly already there, and [she] shall not want to say that their judgments have, just as they stand, these implications*** (Korsgaard takes issue with this part!)." [ELP, p. 147] According to Williams, members of the society can possess ethical knowledge about the world in which they apply their thick concepts *under the nonobjectivist model only*; under the objectivist model, reflective scrutiny of judgments of value will almost certainly arrest the traditional (pre-theoretic) use of thick concepts, destroying the

knowledge a “hypertraditionalist” gains by applying those concepts. [ELP, p. 148] Moreover, Williams contends that reflection does not afford us with new knowledge—it only manages to “characteristically [disturb, unseat, or replace] those traditional concepts.” [ELP, p. 148] Thus Williams concludes that *reflection can destroy ethical knowledge*, and he spells out what this claim means in the excerpt we started from: reflection destroys a hypertraditionalist’s knowledge relative to a thick concept by *extirpating* her “sensibility” relative to that concept; meanwhile, the abstract/general theoretical convictions that do survive reflective scrutiny are ill-suited to replace the thick concepts that are eliminated from her repertoire and the knowledge that has been destroyed.

As it happens, D’Arms and Jacobson also believe that reflection cannot generate ethical knowledge, that reflection cannot be the source of norms that fill the essential role of sensibilities relative to sentimental value concepts. More precisely, they believe that post-reflective, theoretical norms and verdicts—at odds with a person’s established sensibility—relative to sentimental responses are ill-suited to fill the position of that sensibility; such norms and verdicts are always defended by citing purportedly decisive considerations either for or against having select sentimental responses, considerations that have no bearing on whether select objects one is disposed to feeling a given sentiment at *are* valuable in the relevant sense. [DS, pp.610-11] As such, mankind *cannot* internalize these value-irresponsive norms and verdicts into value-responsive sensibilities. [DS, pp.610-11]

The resemblance between the two discussions ends there, however. I suspect that two considerations lie at heart of the crucial disanalogy/divergence (to be discussed below) between the two discussions. First, I want to draw your attention to the following

observation of Korsgaard. In alignment with Korsgaard, I—and D’Arms and Jacobson, I suspect—consider the nonobjectivist view to be mistaken. Williams is wrong in arguing that thick concepts should be understood such that reflection is not inherent in them and their application: as Korsgaard emphasizes on [SN, p.117], thick concepts are *normative*, they are the kind of things that can govern only a *reflective*/reason-dependent mind, and therefore, their use is necessarily reflective (founded upon reason(s)). In the excerpt on page 1 from [DS, p.589], D’Arms and Jacobson assert that sentimental values are normative, and hence to some extent reflection is inherent in sentimental value concepts, and in sentimental value projection, along with any associated affective perceptions of value.¹⁶

The second consideration is one that I am putting forward. I posit that the following prospects of the reflective mind are essentially equivalent: subjecting *coherent patterns* (in how a thick concept is applied) to skepticism/reflective scrutiny, to the point that thick concepts might be subject to dismissal from an individual’s (or group’s) repertoire; making a *sensibility* relative to that concept highly malleable to reflection, to the point that it might be reflection-fragile or even reflection-eliminable. Starting from this equivalence, the following disanalogy arises between William’s discussion and that of [DS]: D’Arms and Jacobson expressly *reject* the second prospect for a person’s sensibilities relative to the sentimental values; by the equivalence, it then follows naturally that they should deem skepticism about sentimental values moot (see the excerpt from [DS, p.589] on p.1 above). More precisely, D’Arms and Jacobson hold that sensibilities are ***rigid*** in that the malleability of a person’s sensibility is constrained by the kind of things that activate it (and to what extent they can

¹⁶ In particular, D’Arms and Jacobson are correct in positing that sentiments can be—and to the fullest extent possible, should be— regulated by *reasons of fit*, giving added credence to Rational Sentimentalism.

activate it), and that reflection and socialization in particular play fairly restricted, albeit significant, roles in shaping that sensibility. [DS, pp.594-5] In particular, this means that a human with a normal, healthy moral psychology cannot exercise reflection to drive sentimental value concepts from use or to extirpate his sensibilities (as [DS] paints them) relative to sentimental values; in particular, the reflective mind *cannot* preempt (destroy, as Williams puts it) knowledge of sentimental value—that knowledge is crucially resilient in the reflective mind; therefore, the un-Socratic conclusion does carry through to the knowledge (members of) our species can have concerning the funny, the disgusting, the pride-worthy and other sentimental values.

3.2: Exemplifying the Crucial Disanalogy

On [DS, pp.609-11], D'Arms and Jacobson consider a case (the Chris versus Fred scenario) where a person's post-reflective verdict on some matter can be at odds with her pre-established sensibility and the judgments that would be grounded in that sensibility.

They see this phenomenon having two consequences:

- (1) To a witness (including possibly herself), her perspective on the relevant sentimental value can appear incoherent, which pulls back to the conflict between her sensibility and her pending theory on the value;
- (2) Where judgments grounded in said sensibility are backed by reasons that are sensitive to the value at issue, and that can regulate her sentimental responses to the extent that coherent dispositions would appear, they contend that her reflective verdict/norm, and the reasons it is based on, are defective in *both* of the above respects.

We turn to the Chris vs. Fred scenario. Suppose that Chris and Fred have analogous esteem sensibilities relative to the shameful. Starting from some agreed upon list of qualities beyond a person's control (i.e., some third person's blameless stupidity) they have felt shame at—and are still disposed to feeling shame at, Fred deems the objects shameful, while Chris's

verdict (based on reflection) is that although he does direct shame at those objects in accordance with his sensibility, **it is always erroneous to feel shame at such objects, and it is not worthwhile or good to feel shame generally.** [DS, p.609] Chris's view (in bold) clashes with the coherent patterns of shame his esteem sensibility issues in relative to certain qualities beyond a person's control, the authors further stipulate that he recognizes the conflict, acknowledging it while discussing with Fred; and yet, he stands by his view for reasons he considers decisive.

In analyzing his post-reflective verdict/view about the shameful, we see that it is alienated from the patterns of shame that resemble his sensibility—hence it is not a judgment of sentimental value, and there is no precedent for believing that it can do the work of a sensibility or that it can be internalized as a sensibility; in particular, his sincerely avowed belief, a belief at odds with his sensibility, stands to mislead others—or even himself—about the nature of his reasons. [DS, p.609] His view is based on considerations of the moral propriety of feeling shame, considerations that are not suited to regulate his shame—or really anyone's shame, and are not sensitive to the value at issue (they do not concern whether the blameless qualities really are shameful). These considerations are WKR's for not directing shame at those qualities, and hence cannot support the negative judgment ("These qualities are not shameful."); his sensibility certainly speaks to the contrary. D'Arms and Jacobson contend that sensibilities cannot be replaced by theoretical norms because the considerations people (always) adduce for such norms are wholly insensitive to the values at issue and cannot regulate the sentiments that decry those values. An evaluative sensibility, however, is by definition a source of reasons that are sensitive to

the value at issue, that can regulate the corresponding sentiment, and that explain coherent patterns in feeling that sentiment. Folk psychology posits as much and human nature takes sentimental values and sensibilities seriously. In short, sensibilities are demonstrably effective in tracking values and governing human responses, while theoretical norms on sentimental value tend to be grounded in considerations that make them *ineffective* in both respects. Statistically speaking, foregoing one's established sensibility in favor of a theoretical norm in a scenario such as Chris's—without precedent for being confident that you, let alone mankind, can internalize it as a sensibility—would be the defective move; sensibilities are the better source for reasons up to those two tasks

Of course, nothing in the discussion up to now actually dismisses the un-Socratic conclusion in Chris's case; that is what we need to verify. If in following Williams' lead, we construe the esteem sensibility as a "perspectival" knowledge provider relative to the shameful, then the following excerpt discusses how reflection might render Chris's perspective on the shameful *incoherent* without preempting or extirpating his esteem sensibility [DS, p.609]:

"[Chris] has no coherent evaluative perspective on the shameful, [courtesy of] a rational tension in his view of the [matter]. His way of seeing a portion of the evaluative landscape, informed by his response, cannot be brought into accord with his theory of that value. Until he revises his theory or reforms his sensibility, there is simply no good answer to the question of what is shameful by Chris's lights."

The sensibility will remain essentially intact even in this "rare" case—Chris's case is rare since such reflection-sensibility conflicts are seldom lucid to people. [DS, pp.610-11] It is then obvious that the un-Socratic conclusion fails to hold in Chris's case; D'Arms and Jacobson's points in the above excerpt extend to showing that un-Socratic conclusion also

fails in any reflection-sensibility conflict case, and hence it fails to obtain in general. We have now accounted for the crucial disanalogy in full.

4. Concluding Remarks

The goal of this paper was to show that there are substantial and significant respects in which the practical identity discussion in *The Sources of Normativity*, and the un-Socratic conclusion discussion in *Ethics and the Limits of Philosophy* resemble the investigation into evaluative sensibilities in [DS] and, in fact, serve as precedents for that investigation. Interestingly, there are key disanalogies between each of the first two discussions and the investigation that [DS] pursues. In the first place, “instability of affect/intrapsychic conflict in sensibility” cases are devoid of the attractive niceties (i.e., transparency, disposed towards deeper self-understanding) involved in “instability of obligation/intrapsychic conflict in identity” cases. To that extent, the disanalogy favored practical identities as more attractive than sensibilities. By contrast, compared to theoretical norms on sentimental value, the stature of sensibilities (relative to sentimental values) looked quite good after showing that the un-Socratic conclusion does not obtain for them in the way Williams suspected it would for any sensibility relative to a thick concept. Consequently, humans with a hearty, moral psychology have reason to be *optimistic* about both the resiliency/rigidity of our sensibilities, and the prospect of gaining knowledge about the sentimental values that play a crucial role in navigating our social lives.