

THE AUTHOR

## IRRATIONAL INTENTIONALITY

Insofar as we are rational creatures, minimally good practical reasons exert a force on our actions. But what is it about the nature of inference that makes practical reasons normatively forceful? There are three attempted answers. First, the force of practical reasons might be derived from their *instrumental* role -- on this explanation, the presumed acceptability of the end transmits its rational propriety to the choice of appropriate means. Second, it might issue from their *substantive* role just in case the ends themselves are open to critical self-diagnosis, i.e., where the rationality of ends themselves is derived from their ability to satisfy a universally valid rule. Third, it might derive from their *phenomenological* role, i.e., ends and means are both evaluable in terms of their qualitative fit with respect to the contents of judgment. The instrumental account is the most influential by far, and usually associated with Hume; the substantive account is influential in ethics and value theory, associated with Kant.<sup>1</sup> Yet neither account explains intuitively manifest failures of practical reason mereologically: i.e., by saying that the contents of the judgment entertained do not fit properly with the judgment considered as a whole, and especially, in connection with the way that the knower is oriented towards the contents of the thought.

Consider: suppose I say, "My expectation that Smith will win the election is true." This is a grammatically bizarre thing to say; for expectations are not generally *true*, so much as apt or inapt, and the grammatical infelicity is of a kind that reflects a failure to adopt the right orientation towards the nested proposition. Notwithstanding the strangeness, though, the sentence is meaningful -- if we say it is *nonsense*, the challenge will have to involve a sharp derogative use of that term which has long fallen from favor. If we wanted to say that I misused my words, the question is why this instance counts as a misuse; and if one wanted to say that this is the wrong choice of propositional attitudes, the question is what makes a choice of attitudes wrong. You might want to say it is because the attitude of expectation is the wrong means to the end, which is the representation of a state where Smith wins the election; but this seems to be imposing an artificial

separation between relata, since the judgment at issue is not the nested proposition (whatever is to the right of the ‘that’-clause) and the attitude that it connects to, so much as our reflective disfavor of the way that the clause and attitude fit together given that the whole proposition of which they are parts is not (in my view) nonsense. We need an account of the ill fit between the attitude parts and their whole, not the relation between means and ends, or between maxims and the categorical imperative. That ill fit is grasped by thinking about the ways we are going about thinking of things, i.e., this thought as a proposition-attitude pairing, and how that pairing is intuitively bizarre for reasons, not just because of a credulity towards grammar.

Such failures are cases of irrational intentionality, and the most interesting (I call them ‘*radical*’) occasions of irrational intentionality are those where a judgment is subverted owing to a misalignment between the contents of a judgment and the way the knower is oriented towards those contents. I shall argue that the phenomenological account can explain what is promising about substantive accounts, and that it is capable of doing so without needing to trace its success back to its tacit or explicit instrumental role.

The phrase ‘irrational intentionality’ does not occur very often in analytic philosophy – or elsewhere. Certainly, a seasoned philosopher might even be surprised to find out that there is *any such thing* as irrational intentionality. Yet one aim of this paper is to argue that many well-known failures of practical rationality do, at least sometimes, fall under the banner of irrational intentionality. After some methodological and conceptual remarks in §1-2, I consider four potential examples of irrational intentionality: irrational conations (e.g., conations that subvert ends in action) and irrational positions (e.g., positionally constitutive ignorance towards facts) in §3, irrational concepts (a *Gestalt* orientation towards disjunctive categorizations) in §4, and irrational attention (e.g., delirious paranoia towards objects) in §3-4. I suggest the failures of reason that typify these kinds of cases issue from the fact that they are radical cases of irrational intentionality. None of them need to be explained as failures of instrumental reasoning in order to understand what has gone wrong with them. The problem is that each of these phenomena can be understood as a mental state that is intrinsically out of order.

An account of irrational intentionality has four features: it is the study of a certain kind of failed reasoning, involves errors that are ostensibly avoidable, with relatively simple constituent reasons and judgments, and whose failure suggests reference to cognitive mechanisms that are introspectively identifiable. 1. The error of reasoning can be identified as a subversion of the strict supporting relationship between judgment and its supporting reasons (premise-conclusions relation), i.e., where the attributed rational fault makes a decisive difference to the degree to which the conclusion is worthy of assent. 2. We are only interested in

those cases where there is a hope of being able to judge otherwise, i.e., where misorientations to objects of judgment are not ‘built-in’ by our native cognitive architecture. 3. The study needs to focus on relatively simple judgments, since the more we appeal to higher-order reasons (e.g., volitional necessities), the more we are at risk of smuggling means-ends reasoning into judgment. 4. Intentionality obeys the need for coherence, so is grounded in the recognition and/or integration of patterns, which is the very thing that irrational intentionality distorts.

Throughout, I shall be concerned to guard against three major counterarguments from an instrumentalist. An instrumentalist could object that practical reasons are belief-desire pairs, and therefore must be means-ends pairings (§2); that positions, being higher-order beliefs, are sufficiently complex as to count as instrumental reasons (§2); and that any study of inferential consequences of a concept involves tacit means-ends reasoning, as illustrated in our capacity to distinguish between benign and malevolent disjunctive categorizations (§3). I reply that the belief-desire vocabulary represents only a small and unrepresentative sample of folk psychological states, and that a more perspicuous accounting must take low-contrast attitudes seriously; that, while *higher-order states* can be reasonably suspected of being derived from tacit instrumental reasoning, this is not a decisive point one might make against merely *second-order* ones; and that the difference between benign and malevolent disjunctive categorizations can be found in the failure of antecedent conditions of application involved in bare cognitive processing, without appealing to failed connection to consequences.

## 1

We possess rationality insofar as our judgments are appropriately supported by our reasons for judgment, whether ‘judgment’ is propositional or imperative. There are at least two ways of explaining that supporting relationship.<sup>ii</sup> On the **instrumentalist** conception, an actor behaves rationally only insofar as they arrive at an *appropriate choice of means* in light of some *predetermined ends*.<sup>iii</sup> The ends are presupposed, the only suitable question is whether an agent has voluntarily chosen the right means to that end after deliberation. The **substantivist** conception of rationality proposes that it is possible for *ends themselves* to be (ir)rational, extending the field of inquiry into a discussion of whether our ultimate values are appropriate measures of some domain of rational action or thought.<sup>iv</sup> The suitability of such intrinsically valuable ends-in-themselves (e.g., the good will for Kant) are judged in terms of rational diagnostic

tests (e.g., the categorical imperative), and determine corresponding judgments of irrationality (e.g., the willful subordination of right principles to mere expediency). The instrumentalist conception is concerned with the development and deployment of the whole gamut of experience to effectively secure one's goals, typically in a way that *minimizes losses and maximizes*. The substantivist conception widens the field to consider whether the ends themselves are rational to hold by their own lights, and central to theories of justice. Yet it is also possible to demarcate a third conception of irrationality, the **phenomenological** theory, which treats some passage of thought as irrational simply by virtue of being held in the ways they are and being what they are about.<sup>v</sup> This conception widens the field further still by answering the same questions as the first and second theories, while also taking material inferences (i.e., negative rational intuitions, manifest sense of error) at face value.<sup>vi</sup>

On first glance, the instrumental conception is able to make sense of the insights of the substantive account by arguing that the (ir)rationality of a judgment about an end-in-itself is only meaningful by *covert* means-ends reasoning, e.g., by reference to higher-order ends held in the background. (So, e.g., justice *seems* like it is intrinsically valuable only because a whole lot of our practical beliefs happen to be concerned with justice.) The converse argument is also possible, as a champion of the substantive conception of rationality can argue that the evaluation of the means-ends relationship is itself unintelligible without an intuitive sense of what makes for good inferences, period. The substantivist might add that their account of rationality has the advantage of transparency, since if an explicit end is bad then the irrationality of the means can be immediately inferred. Their account also sates the appetite for foundational self-justifying norms like the categorical imperative. A standoff between instrumentalist and substantivist interpretations of rationality can be maintained because both accounts trade on a common vernacular, the practical language of means and ends. The third conception agrees that the instrumentalist and substantivist conceptions describe some proper instances of (ir)rationality, but also admits of a class of cases that instrumentalism rejects -- namely, those where the terms of a judgment can be irrational simply by virtue of one's orientation towards them. For the intentionalist, 'means-ends' reasoning sometimes mis-describes the nature of the connection between judgment and its supporting reasons, owing to the failure of the contents to fit the reasoner's point of view.

In what follows I provisionally explore some of the central cases of this third conception of rationality: paradoxes, *akrasia*, critical paranoia, unstructured patterns, disjunctive categorizations, and constitutive ignorance. The method is phenomenological, in that I suspend belief in instrumental rationality for the sake of understanding intentional rationality. My purpose is not to endorse the

intentionalist view, or even to provide a comprehensive analysis of the intentionalist view, but only to encourage systematic attention to its nature, scope, and philosophical consequences.

The phenomenological method requires that we should bracket away certain sorts of considerations for the sake of analysis. For our purposes, I shall ask the reader to suspend belief in any opinions they might have about what counts as a sufficient reason to believe that a conclusion is satisfactorily grounded in its premises, insofar as these judgments of sufficiency overtly demand means-ends reasoning. This will entail suspending belief in the propriety of certain forms of inference -- e.g., it is almost certainly true that cost-benefit analysis and inference to the best explanation will typically make overt use of means-ends rationality, and almost certainly false that deductive or material inference do. So, in cases of abductive reasoning, ask: ‘can this passage of judgment be captured by deductive or material inference?’ If it cannot, then set the inference aside as unfit as a candidate of irrational intentionality. Once bound by these strictures we might be only able to cover a small patch of epistemic ground, but it is still ground worth covering.

## 2

There is a special kinship between the idea that irrational ends-in-themselves and that some judgments are irrational to hold by virtue of the ways we are orientated towards what they are about, as the former seem to be a special instance of the latter. If an end itself is irrational, then any attempt to hold a judgment which describes the proper means of achieving that end shall also be irrational, just by virtue of what the judgment is about. It is an ‘end-in-itself’ just in case it conforms to a kind of judgment which is always -- perhaps even necessarily -- rational to uphold and enforce. Kant’s articulation of the categorical imperative is one such judgment, as his three formulations of the categorical imperative issue from his conceptual scheme that is ostensibly embedded in our position as rational creatures (i.e., the Categories of Quantity).<sup>30</sup> For instance, his formula of humanity is oriented towards the plurality of objects and ends, meaning the “matter or content” of a judgment. So, taking Kant at his word, violations of the formula of humanity are recognisable as cases where a judgment that fails as rational just by virtue of being held and being what it is about.

The notion of irrational intentionality comes from a different place than the substantivist account of irrationality, and has different sorts of downstream consequences. It arises from considerations related to the interplay of logic and mind, the traditional areas of concern in discussions of theoretical rationality -- whatever influence it has on making sense of the rationality of intrinsic value is

inferred from it and not assumed beforehand. Its central assumption is that we can examine the rationality of judgments in the tribunal of experience by bracketing away (or temporarily suspending belief in) any of those inferences in a web of belief that do not make a manifest difference to the diagnosis of the rationality of a judgment in its context. Its primary domain of application is the psychology of mental representation and its behavioral and agentic effects, and only indirectly on ethical theory. Meanwhile, if we discover that something is substantively irrational, then we will have come to understand something about living a strict moral life. If on the other hand we find ourselves looking at some representations as irrational in themselves, our conclusions may or may not have much to do with ends unworthy of pursuit or endorsement, though you can be sure that a strictly moral life will be precluded by irrational intentionality.

An agent displays irrational intentionality insofar as their judgment is defective by virtue of the ways that the agent who holds the judgment is oriented towards what their judgment is about. In a shallow sense, this always already applies to both judgments of improper means to ends and judgments of improper ends themselves, since both are presumably 'about' the wrong sorts of things in the wrong sorts of ways. One might then think that instrumental rationality is just a special case of phenomenological rationality, potentially risking psychologism. Though an argument of that sort is possible, I will not assert anything quite so immodest in this essay, nor do I see much promise in following that line of thought -- what follows is meant to be an exploration of the phenomenological view, not a reduction or elimination of the instrumentalist one.

It is sensible to observe a subclass of cases of irrational intentionality that are appreciably irrational from both the intentionalist and instrumentalist points of view, a place of overlap. For example, suppose I am playing a game of softball, out in the field, and I believe the team at bat is unlikely to make any hits. So, out of boredom, I allow my attention to wander, and I become fixed on watching a cat in the bleachers. But, due to my inattention, I miss out on a pop fly, which lands by my feet; and as a result, my team loses the game. My distracted attention was an instance of irrational intentionality since the cat was not the rationally appropriate thing to attend to in the rich context of my plans. And yet, my misdirected attention was rationally defective because it was an impediment to my broader ends. Or, for another example: suppose that an editorial board of a magazine decides not to publish a controversial article because they worry that attention to the article would inflame further controversy. Given their ends (avoid controversy), the choice of means (deciding to publish) would be instrumentally irrational; yet it can only be appreciated as irrational once one accepts that a controversial subject is made controversial in part by paying attention to it. Call all of those cases which are equally amenable to description in terms of

intentionalist and instrumentalist conceptions of rationality '**moderate cases**'. Although moderate cases are not the focus of this essay, they should give a sense of the general scope of the intentionalist view.

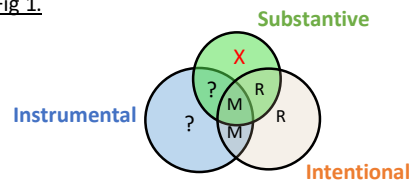
Though there is overlap between instrumental and intentionalist views, it would be a mistake to think of irrational intentionality as a trivial neologism. For the category also includes, but is not exhausted by, a special subclass of judgments that *seem rationally defective by virtue of their defective constituents and the ways they are coordinated, and not by their broader instrumental role*. Call these '**radical cases**'. This set includes, but is not exhausted by, those occasions where a claim is irrational by virtue of its terms (i.e., subject or predicate). Practical judgments that propose to contradict ends-in-themselves, or *irrational ends*, are one kind of radical case. So, e.g., "The best way to cause the most misery in the world is to create a nuclear holocaust" is an irrational judgment by virtue of the fact that the proposed end is substantially irrational -- the very idea that there is a 'best' way to cause mass misery is simply denied.

To see the scope of radical cases, let's consider a very different kind of example. Namely, cases of *irrational belief*, understood as mono-sentential paradoxes like "This sentence is false". It can hardly be disputed that this semantic paradox is irrational owing to its self-defeating characteristics, especially insofar as we are oriented towards the judgment as a potential belief. Indeed, since the paradox concerns a single sentence, the diagnosis is immediately available to attention. To hold that isolated judgment in mind as a suitable candidate for belief is already to have engaged in a mistake -- the merest attention to that judgment reveals that it is impossible to rationally believe, because its valence is apparently incapable of *staying put* under scrutiny. Yet the nature of that mistake is not easy to diagnose in a fashion that would advantage the substantivist account. For there is no distinctive sense in which the semantic paradox fails because it involves an explicitly irrational goal. For one thing, of course, it is not clear that all theoretical judgments can be properly phrased in terms of the logic of means and ends. For another thing, the central preoccupations of substantial accounts are in theories of justice, not logic. But putting those matters aside, it is still the case that the presumed end of every indicative sentence -- truth -- is left tacit. As a result, one of the main advantages of the substantivist theory, transparency, is not available. Nor is it a distinctive case of a failure of instrumental rationality, since unlike poly-sentential paradoxes, the judgment's failure is distinctively *self*-defeating. Ostensibly, the rational defect relates to the way it is held and about it is about, which is apparent regardless of any further opinions we might have.

The tripartite relationship between the theories can be observed in Fig. 1, below. 'R' stands for the conceptual space that includes radical cases, 'M' stands

for the space that includes 'moderate cases'. 'X' means a vacant conceptual space – I suggest, as a secondary or motivating concern of the essay, that substantial rationality is most distinctive when conceived a radical instance of intentional rationality, and obscure otherwise. A question mark is placed in those considerations that have been bracketed.

Fig 1.



While there are some advantages in appealing to a rich conception of goodness in the evaluation of inferential propriety, the account sometimes seems guilty of over-explanation of the obvious. If a student asserts the semantic paradox in the course of discussion section, or an oxymoron or manifest absurdity like "Rain isn't wet", their instructor need not challenge them in ideological fashion by proposing an overall worldview where the student has failed to participate in the good life. On some occasions, the instructor needs only to propose a challenge in order to see the assertion withdrawn. And while this can sometimes be done in an illegitimate or epistemically unjust way (as in Lewis's "incredulous stare"), the withdrawal is sometimes sufficient and justified by the student's own lights, e.g., insofar as they cannot appeal either to intuitions that conform to good-faith interpretations of their biography, or to beliefs justified by evidence and informed by theory.

Those modern-day value theorists who endorse substantive conceptions of rationality (e.g., Rawls, Nozick) have nothing to lose by expanding their attention to radical cases of irrational intentionality. For one thing, latter-day conversations over what qualifies as an end-in-itself, and what test to use in diagnosing it, may already seem to have a groundless or 'skyhook' quality to them. One could attempt to ground one's conception of an end-in-itself in terms of the web of belief or wide reflective equilibrium, potentially resulting in accusations that they are engaged in disguised means-ends reasoning. In so doing, though, the substantive theorist would lose precious ground to the instrumentalist one. For another thing, the special focus on ends that belongs to the substantive theorist can be explained by their preoccupation with particular sorts of philosophical questions and problems, namely those related to theories of justice. If a conception of rationality can address a wider range of theoretical questions, while also accommodating the purposes of a theory of justice, it seems reasonable to entertain it.



While mono-sentential paradoxes (*irrational beliefs*) and false claims about intrinsic ends (*irrational ends*) are part of the class of radical cases, they do not constitute the whole. Discussion will address four other sorts of cases: *irrational choices*, *irrational positions*, *irrational categorisations*, and *irrational attention* in ‘rolling commentary’ style. Conceptual tools necessary for the analysis will be presented during the course of their use in examining the cases. And though every case under discussion is exotic, most should also be familiar; there is by now a rich literature that examines irrational choices (akrasia), conations (aliefs and phobias), and concepts (Borgesian categories), just to choose three examples. The main point of this paper is to defend the idea that these philosophical ‘x-files’ all occupy the same cabinet drawer. (Incidentally, it must be noted that these kinds are not mutually exclusive descriptions of cases. So, for example, claims that someone occupies an irrational position are so likely to be used as a cudgel for the purpose of epistemic abuse that they might also count as irrational ends. Forewarned is forearmed.)

### 3

Even when limited to the examination of radical cases, a full articulation of the concept of irrational intentionality depends on analysis of the two component ideas, *rationality* and *intentionality*. This is not an easy task since both concepts resist reduction. If there is anything fundamental to consciousness, then intentionality is a good candidate. Putting aside meditative states, there is little point in denying that *rational consciousness is always consciousness of something*. To put an analytical gloss on it, a judgment is directly about the conditions under which it is worthy of assent (whether it be conditions for truth or execution), and indirectly about its terms (conditions for satisfaction). The terms of a judgment can be about objects, categorizations, and ends (to choose a small but important sample). Yet, consciousness about something is also always accomplished by way of *an orientation towards what one is thinking about*, and partly projected upon what one thinks about, and through which it is conceived. The ‘orientation’ of a judgment is, as it were, a residue of the knowing self that has been left on what the judgment is about. So, for example, one might observe that consciousness of every object is consciousness of an object *of attention*. Similarly, every classification is conceived as a *Gestalt*, every end is an end through a *conation*, and every fact is a fact conceived through a *position*. (See Fig.2.) In all of these cases, insofar as we are conceiving of the judgment itself, the italicized orientations are generally accessible to meta-cognition, while the terms characterize what is thought about. At least during waking hours and lucid

dreams, this two-part aspect of rational consciousness is available to apprehension on demand.

| <b>Fig.2.<br/>Intentionality.</b>                 | <b>Representation</b>         | <b>Orientation</b>       |
|---|-------------------------------|--------------------------|
| Term<br>(satisfaction<br>conditions)              | Object                        | <i>Attention</i>         |
|   | Categorization                | <i>Gestalt / pattern</i> |
|   | End                           | <i>Conation</i>          |
| Judgment<br>(truth or<br>execution<br>conditions) | Facts /<br>What is to be done | <i>Position</i>          |

A preliminary characterization of a few of these ideas may be helpful. ‘Conation’ means something like ‘effective motivation’, or the perceptible force of the judgment that is antecedent to the behavior in action. In this respect, a conation can be distinguished from doxastic (propositional or imperative) attitudes because conations are not necessarily ‘about’ anything, so much as they are causal explanations of how (e.g.) our ends are pursued, which could be retroactively described as reasons for action. I take it that a ‘position’ is a *complex* orientation through which we understand our judgments, but only manifest when directed towards *higher-order judgments*, i.e., those judgments that explicitly contain a reference to one or more doxastic attitudes. (My position relative to the sun is not a manifest issue in the sentence, “The sun is too bright”, but it is manifest in “I feel like the sun is too bright”.) Positions may or may not be a direct function of the will. For example, on one extreme, some especially resilient positions are often diagnosed as ‘schemes’, either somatic or conceptual. These schemes impose certain second-order attitudes upon us by alethic or effective necessity, and based on our innate structural endowments (e.g., for Kant, that every object is an object of intuition) -- but because of their pervasiveness they can only be discovered through painstaking research (e.g., into natural semantic universals). On the other extreme, positions that can be occupied on demand are more like epistemological ‘stances’, and harder to distinguish from taking on a second-order set of attitudes provisionally. In the middle are volitional necessities, understood as a function of things we care about, and which constitute our complex stories about ‘who we are’ from the internal point of view.<sup>viii</sup>

These rows are arranged in pairs: object-attention, categorization-Gestalt, and so on. But these pairings are characteristic, not necessary. So, for example, it is entirely possible to talk about the rationality of speaking about classifications conceived through positions (e.g., purport), or objects through conations

(affordances).<sup>ix</sup> Indeed, mono-sentential paradoxes have been characterized above as a case of contentful *flip-floppiness*, i.e., failure of a claim to maintain its contents insofar as we are paying it the merest attention.<sup>x</sup> I only single out these pairings for initial discussion because, when they are misaligned, they make for compelling and distinctive exemplars of radical cases while also appealing to phenomena that are relatively familiar in the literature.

Rationality involves a strict supporting relationship between judgments and a satisfactory accounting of the reason(s) for the judgment, insofar as those reasons are properly coordinated. (A *loose* relationship between judgment and reasons is merely *reasonable*, as opposed to rational. Though whether hard radical cases count as unreasonable or irrational is a matter of manifest interpretation – see below.) But then the notion of ‘reasons for judgment’ needs unpacking. The relationship between practical reasons and ends is a promising place to begin, since both reasons and ends are at least part of good answers to the question, ‘why did you do that?’<sup>xi</sup> Folk psychologists think that practical reasons are pairings between coordinated beliefs and desires.<sup>xii</sup> The ‘end’ or ‘goal’ is the term embedded in the aspirational part of the reason for the judgment. On first pass, then, practical rationality means having a strict supporting connection between judgments of what is to be done and coordinated attitudes, typically beliefs and desires. *A judgment is irrational insofar as the orientation towards the judgment subverts the strict connection between the judgment and its reasons* (i.e., the corresponding set of coordinated attitudes). Ultimately, the nature of that ‘strict’ connection is the subject of our investigation – we phenomenologists are trying to see to what extent we can earn it from the tribunal of experience. For the intentionalist, the nature of the strict connection is supposed to be adduced by observing the *manifest* lack of fit between the contents of a practical judgment (e.g., command and its terms) and the corresponding account of attitudes (e.g., coordinated beliefs and desires). At core, the idea that a strict supporting connection can be subverted necessarily depends on the proposal that one’s orientation in a judgment is capable of *making a decisive difference* to the relationship between the judgment and its supporting attitudes.

It is best to talk vaguely about reasons in terms of their representational and aspirational role, and not to concentrate on particular attitudes like beliefs, because the vocabulary of folk psychology is richer and vaguer than we sometimes give it credit. ‘Belief, intention, desire’ do not exhaust our lexicon of folk attitudes, as there is a standing reserve of antecedents to that-clauses -- e.g., imagining, guessing, assuming, intuiting, expecting, and accepting -- which all make a direct positive contribution to parts of the process of knowledge production.<sup>xiii</sup> Moreover, there is no antecedent reason to think non-standard

attitudes of this kind will simply reduce to the language of belief and desire. For one example of a clumsy attempt at reduction, consider expectations. An expectation is sometimes seen as a kind of belief, and belief essentially directed at truth. Yet they do not share the same meaning, since ‘Your belief is true’ makes perfect sense, while ‘Your expectation is true’ sounds like it was written by an author with a tin ear. Nevertheless, in the context of practical reason, expectations may play a belief-like role when answering “why did you do that?”, in that the expectation may be directed towards facts about opportunities for action. Considerations of this sort are behind the decision to discuss doxastic attitudes that are ostensibly fact-directed (such as intuition, expectation, and belief) in terms of ‘holdings’, as in, ‘holding that p’. Similar substitutions prompt the use of the conventional term, ‘valuing’, to describe standards of evaluation framed as an (anti-) preference – e.g., dreaming, wishing, hoping, needing, wanting, and demanding, and so on. I do not suggest these general terms *must* be useful on further analysis, but only that they seem appropriate to assume when doing philosophy after Hume.

Be that as it may, when I consult the tribunal of experience, I find that there are no guarantees that the difference between what is ‘valued’ and what is ‘held’ always merits a crisp demarcation. For while the contrast between these two sorts is pretty obvious when it comes to tokens of belief-desire pairs (e.g., during creative high-level planning, the contrast is rather stark), it is more of an ordeal to motivate these role assignments when talking about small-scale actions. So, e.g., it isn’t always obvious that the type of attitude, ‘accept’, covers tokens that have an aspirational role in inference, and that tokens of ‘recognize’ always play a representational role. If I judge that following the rules of baseball is ‘the thing to be done’ while I am playing baseball, the (ostensibly) aspirational reason I provide -- ‘I accept the rules of baseball’ – does not seem intuitively like it differs from the (ostensibly) representational reason, ‘I recognize the rules of baseball’. To be sure, there is presumably a truth-conditional difference to the assertions, but it is not clear that the distinction always conforms to their aspirational and representational roles. So, if during the process of rational deliberation on practical judgments we decide that we cannot infer an ‘ought’ from an ‘is’ *come what may*, as an institutional certification of proper thought about oughts, then either the aspirational role assignment to attitudes will be done *ad hoc*, or the practical judgment will be reinterpreted as a theoretical one. Presumably, an *ad hoc* connection between judgment and supporting attitudes would fall short of being a strict connection – at best, it seems better to say it’s a loose connection (i.e., merely reasonable, not rational). Meanwhile, in other contexts, the difference between ‘aspirational’ and ‘representational’ attitudes is quite

discernable and experientially motivated, where ‘accept’ seems less like acknowledgment and more like giving a mental ‘thumbs up’.

It would be nice to be able to explain that contrast between high- and low-contrast contexts, if at all possible. One way of understanding the difference is by attributing a difference of *position*. When I am in an activated physiological state, it may be especially obvious to me that my attitudes towards practical judgment can be delineated according to their aspirational and representational roles. When you want things passionately, it is easier to say: ‘Ah, I desire this, and I believe I do not have it, and here is my plan to bridge the gap.’ Yet this obviousness is diminished and confounded in a passive or chronically depressed position, as far as the small-scale nuanced attitudes are concerned. When I am in a listless or depressed position, most of the positive parts of my experience can be found in nuanced or ambiguous states of expectation and acceptance – typically, experiences with the sublime. (My own experience presented itself as if it were fear of effecting desire, or the experience of desire as if it were terror, and even sometimes as an addiction to belief; and that when rousing action did come, it came as a surprise.) So, it seems that the position one occupies can make a difference to the rational status of nuanced attitudes to judgment, however we want to characterize that difference.

Does this necessarily seem like a case of irrational intentionality at all, i.e., where a strict connection is *subverted*? For the nuanced practical attitudes that I have regarding the rules of baseball are at least a reasonable basis for the corresponding judgment to obey the rules, even if they fall short of providing strict support. I do not mean anything special by the idea of ‘subversion’, and use the term with its expected effects in mind. These are twofold. 1. ‘Subversion’ of rationality need not imply that the person is bereft of practical reason. It is still possible to be in a state of *irrational* intentionality while still being *reasonable*, just in case the position only subverts the strict connection while preserving a loose one. (If all I have going for me are the nuanced attitudes, I will likely play a lacklustre game of baseball. But at least I am playing in a way that responds to appropriate reasons.) 2. Even so, if your position subverts a strict connection between attitudes and judgment, there are *no guarantees* that even the loose connection will hold, all other things equal. For a position to ‘subvert’ a connection between a judgment and its attitudes is just to make it *weaker than* some other kind of supporting relation.

Be that as it may, clinical cases are not necessarily going to help us to think about substantial reason, since substantial accounts of rationality are interested in *finding fault where it is due*, and clinical cases are, presumably, faultless. For while occupying the passive position decisively weakens the judgment-attitudes relation, clinical depression is not a second-order choice. Assuming that clinical

depression and other disorders of praxis are properly characterized as weakness of will, they have to be acknowledged as especially severe, or *global*, weaknesses of will. In that respect, such cases are unlike the sort of core cases of weakness of the rational will (*akrasia*) that are typical to the literature in practical rationality, i.e., *local* cases of free, sane, intentional action.<sup>54</sup> For example, an otherwise healthy person's addiction to cigarettes gives them strong desire to smoke, and so gives them a strong supporting attitude towards smoking which subverts their other desire, which is to quit smoking. As a result, the smoker who gives in to the urge is a case of irrational intentionality. Yet that characterization does not hold quite as obviously for the case of the clinical depressive, because the lapsed quitter at least has some *hope available to think they could have done or believed otherwise*. So, the hope of an ability to judge otherwise is a salient feature of many cases where a strict supporting connection has been subverted, and where we do not feel the need to abstain from finding fault. For diagnostic purposes, we might even say that the hope of an ability to do otherwise is a necessary condition for tractable cases of subversion (compared with incorrigible cases, e.g., delusional or fantastic reasoning). Yet, since this hope does not need to be contingent on the agent's failed deployment of anything like a 'cost-benefit' analysis in the moment, it is visible under the phenomenological blindfold. So, it is a radical case.

Here is another persuasive instance of a tractable case of irrational intentionality: taking up a position of *constitutive ignorance*, i.e., a position which furnishes epistemically defective states (false beliefs, inappropriate designs, and the like) as a matter of necessity, but where the force of necessity can be pretty well understood without ever assuming that they are in possession of sophisticated instrumental reasons for doubt.<sup>55</sup> The most interesting examples of constitutive ignorance are cases of ignorance brought on by perverse *volitional* necessities. For example, hard-core members of the Flat-Earth Society define the conditions of group membership in terms of a false belief – that is to say, a straight-forward delusion about the form of the Earth -- in such a way that, insofar as they are rational and sincere believers, any attempt to challenge the legitimacy of the group will have all the effects of an existential challenge to it. And if members are members of it as a matter of volitional necessity, then it may have all the force of a challenge to them as individual persons. This precludes them from objective knowledge of Terran topography. And yet we are not precluded from finding fault, since they were the ones who chose to identify with false things at some point in their lives.

But, notice, that cases of constitutive ignorance are more than just cases of *willful* ignorance.<sup>56</sup> When it comes to cases of constitutive ignorance, the very conditions for a person's narrative identity require them to ignore contrary

evidence owing to what they care about, and to do so indefinitely. These false narratives are important to them as distinctive persons because they provide meaningful content to one's sense of integrity, insofar as they are conceived of as being bearers of meaning and value. Someone does not come to care about something *just because* they decided to, and they do not extricate themselves from what they care about *just because* they decide not to. In both of these senses, we say that volitional necessities are genuine cases of 'necessity'. By contrast, merely willful ignorance is reversible on demand, perhaps just as well described as taking on a new *position* as in adopting new second-order aspirations in order to see what happens to your first-order conative states. At any rate, the important point is that true believers in the Flat Earth Society are ignorant because their position makes delusion necessary, with respect to claims about the roundness of the Earth. Their belief is irrational because of how they are oriented towards the content of their belief, and only so because of the position they have put themselves in.

Discussion of positions makes one think immediately of adopting epistemically peculiar stances. Consider, for example, radical embodied in *weird art*. Specifically, let's consider Salvador Dalí's remarks on surrealism, taken from his essay '*Conquest of the Irrational*'. In that polemic, he encourages us to see things through the position of someone in an artificially induced state of delirium or paranoia. In his wonderful paintings, "images of concrete irrationality are thus authentically unknown images", with familiar melting clocks, exotic shapes, and so on. The main point of the "paranoid-critical" method is "the systematic-interpretative organization of surrealist experimental sensational material" that is "scattered and narcissistic". Dalí tells us that the surrealists artificially place themselves into a paranoid and delirious position, for the purposes of intentionally distorting what they are thinking about. As a result, the surrealist notices during the course of a day a variety of perverse happenings, "revealing a minimum of irrational intentionality, or, just the opposite, a minimum of suspect phenomenal nullity". By the former, I take him to mean irrational objects and events, as far as they relate to judgments about what is properly perceived.<sup>xviii</sup> (By the latter, I suppose he means to refer to the perceptible residue left over by radical doubt, though he apparently thinks that the 'minimum of suspect phenomenological nullities' are more like mundane embodied happenings than the higher operations of the *Cogito*.)

We should take Dalí at his word when he says that the surrealist method produces awareness of irrational intentionality, but it is also worth trying to figure out what is so irrational about what he is thinking about. The surrealist was not constitutively ignorant. Dalí thought the surrealist project was telling us

something important about reason and rationality, insofar as it relates to rational intuition. A central aim of the method, as I understand it, was to force the reader to acknowledge that the objective irrationality of a thought does not necessarily make the thought *impossible for a rational person to think*. So, it was not irrational because it produced irrational beliefs -- the surrealist was not a Lovecraftian villain who wished to push us into a permanent state of psychosis. The surrealist only wished us to understand that it is possible to pay attention to things in an irrational way -- i.e., to observe that we are able to orient ourselves towards objects and events in a way that makes them seem unnecessarily alien. And, far from being a volitional necessity, the surrealist's method is adopted at will. For that reason, it is hard to consider it an irrational position. I would instead like to call these cases of 'irrational attention', and return to them later in the paper for a diagnosis of what makes them distinctive.

Positions are orientations toward second-order judgments: constitutive ignorance is an irrational position, while the surrealist method is a rational position towards irrational events and objects. One naturally begins to wonder about expanding the list to further *higher-order states*, i.e., third- and fourth- order cases of irrational intentionality.<sup>viii</sup> For example, the documentarian Adam Curtis coined the term '*hypernormalization*' to describe an experience where a person takes on (or is forced to take on) a position that seems antecedently surreal or disorienting, and in doing so, become open to facts that threaten their constitutive ignorance. In the political sphere, hypernormalization is a complex, fourth-order collective intentional state, as it implies taking on a totally different position towards what powerful people want us to want, resulting in new (fleeting) perception of facts, much like the first experiences of the philosopher when they exit Plato's Cave. To be sure, such cases of 'rupture' are of enormous interest.<sup>ix</sup> Yet the higher up the chain of higher-order attitudes we climb, the more that it seems like we are at risk of introducing explicit means-ends inferences into our context of interpretation. They are best seen as moderate cases, not radical ones.

#### 4

I have characterized radical cases of irrational intentionality as cases where the strict supporting relationship between attitudes and judgment has been subverted, owing to manifest defects in the ways that the thinker is oriented towards the contents of judgment. Our main focus has been on subversive orientations in general, and on tractable examples in particular (i.e., cases where there is a hope of being able to judge otherwise). Three subversive orientations were discussed -- conations that subvert ends (*akrasia*), positions that subvert objects of attention



(paranoid-critical method), and positions that subvert knowledge (constitutive ignorance). Yet one issue has been neglected, namely, the irrationality of *categorizations*.

It is sometimes admitted that concepts can themselves be irrational. For example, the original legal category of ‘genocide’ contains technical limitations that make it difficult to use in such a way as would produce reliable inferential consequences, since the legal category does not include a great many instances of invidious mass killing (e.g., gendercide, urbanicide) on seemingly *ad hoc* grounds, since the technical concept – hereafter, ‘categorization’ -- only includes a narrowly circumscribed number of protected classes (e.g., religion, nationality, ethnicity, race). In this example, the legal category of ‘genocide’ can be anticipated to fail at serving the ends for which it was designed – namely, the intended destruction of *whole peoples*.<sup>xv</sup> Hence, this defect in the concept reflects a powerful, and interesting, instrumental explanation of the irrationality of concepts themselves -- namely, that they might sometimes be diagnosed as failures owing to their inappropriate inferential consequences, *and that that fact alone must reflect a failure of means-ends reasoning*. For vulgar pragmatism held that the whole question of successful inference depends on utility.<sup>xvi</sup> And if they are right, then the instrumental account has won the day -- for they will then have shown that *all* concepts are instrumental, irrespective of their theoretical or practical context. Yet, by way of contrast, I think it is just as plausible that we could call some concept ‘defective’ owing to something about the ways we are oriented towards its conditions of application. The categorization fails to have the kind of internal unity that plausibly maps onto a reliable location in conceptual space, and there is no need to describe this in terms of its instrumental role. It is phenomenologically unacceptable.

To flesh out the idea that a category can be antecedently unacceptable, we should consider a relatively clear case: disjunctive categories. Surrealist author Jorge Luis Borges is the namesake of those categorizations we call ‘*Borgesian categories*’. In one story, Borges’s fictitious narrator reported on a peculiar taxonomy of animals, which read as follows: “Those that belong to the emperor, embalmed ones, those that are trained, suckling pigs, mermaids (or Sirens), fabulous ones, stray dogs, those that are included in this classification, those that tremble as if they were mad, innumerable ones, those drawn with a very fine camel hair brush, et cetera, those that have just broken the flower vase, and those that, at a distance, resemble flies”. There are plenty of things wrong with the definition (i.e., it is circular and uninformative), but for my purposes the central characteristic is that its intension (definition) is *disjunctive*.<sup>xvii</sup> Intuitively, the semantic-lexical parts of the intension do not hang together in a coherent pattern, and are only unified by the implicit license to shift focus whenever one makes a

strategic use of ‘or’. On first glance, it seems reasonable to suspect that Borgesian categories are not going to be generally helpful to inquiry, based on antecedent formulation of the conditions of application alone.

Yet, admittedly, this is too quick. Recall, one analytic philosopher famously asked us to consider the strange properties of the color predicate, ‘grue’, where: “‘grue’ denotes *green* before time (t), and denotes *blue* at (t+1)”. This, too, is a disjunctive definition, albeit one conveniently separated into time-slices. Phenomenologically, ‘grue’ seems like irrational categorization because there is just something defective about the ways that we are oriented towards the denotation of the term. But the defects might be merely apparent. So, consider the case of ‘planet\*’, which has the following salient implication: “The extension of ‘planet (*simpliciter*)’ includes Pluto at time before 2006, and does not include it after 2006”. Here, ‘planet\*’ is formally identical to ‘grue’ in being a Borgesian category. Yet (as a matter of fact) ‘planet\*’ describes the actual linguistic conventions of the *International Astronomical Union*. Where ‘grue’ seemed like an irrational categorization, there is no mystery whatsoever to ‘planet\*’: the scientists decided that certain kinds of explanatory inference were worth caring about, which prompted a shift in extension. Given the rich contextual and instrumental environment, a case of irrational intentionality turned out to be wholly reasonable, and perhaps even rational. Diagnosis: the phenomenological blindfold caused us to misdiagnose a perfectly reasonable description of the facts as a case of irrational intentionality. If that’s true, then the intentionalist argument has apparently failed.

But this instrumentalist argument misses an important distinction. Suppose we agree that some disjunctive categories are *malevolent* (e.g., grue), while some are *benign* (e.g., planet\*). Then we have to motivate the difference without talking about its broader role in means-ends reasoning. On first pass, of course, we could make a distinction in terms of their explanatory role. Planet\* has a helpful explanatory role, not just in making sense of the universe, but also in making sense of our coordinated higher-order attitudes towards that universe, and our shifting sense of which inferences are worth caring about when we use the word “planet”. But one must not suppose that science is the sole standard for intelligibility; if one does, then they are left to wonder what to do with cases that are intelligible to reasonable people without being especially helpful to scientific explanation. The phenomenological difference lies in the peculiar sense we have, in considering the difference between grue and planet\*, that planet\* at least preserves a pattern. Consider *Heraclitus’s aphorism*, the proposal that we supposedly can never step in the same river twice, it being always a bit different with the passage of time. On some reflection, it should be clear that Heraclitean river(s) are a function of a disjunctive categorization practice: what we call ‘river’

in ordinary language is in fact a set of rivers, indexed to particular moments, and unified only as a Borgesian category. It is plain enough (to me) that the Borgesian river-manifold is not fit for ordinary belief (except, perhaps, in the philosophy seminar). It is a case of irrational intentionality, because the categorization is plainly *out of joint* with the apprehended *Gestalt*. Still, one hopefully understands the point of talking about the Heraclitean rivers -- it is not *absurd* or *nonsensical* to talk about or think about for its own sake, owing to its failure to fit an innate, absolute scheme in cognition. Indeed, the puzzle of the Heraclitean river has illuminated and motivated some genuinely interesting puzzles in language and mind. So, it is a benign categorization, because we know that we can depend on the demonstrative unity, pattern, or *Gestalt* of the denoted river-manifold to persist independently of how we categorize it. It is only irrational when forced to play a role in belief. So, the phenomenologist can make sense of the idea of a 'pattern' in planet\*, and captures an important nuance that instrumentalism does not need to care about.

To be sure, instrumental or means-ends reasoning is probably indispensable when justifying our ways of thinking about physical structures or 'real patterns', of the sort that would distinguish between the epistemic status of centers of gravity and the smallest geographic circle which contains all of Daniel Dennett's lost socks. Astrophysics counts the inclusion of the former as a net gain, while no science makes much productive use of the latter. All the same, *real patterns* are not patterns *simpliciter*, and patterns *simpliciter* are just ways of talking about one of the most intimate parts of our experience without ontological commitments. So, e.g. we have to have hope in our ability to talk about patterns *in spite of* the noise (attributed structure across contexts), as opposed to patterns *in the noise* (attributed structure in context).<sup>xviii</sup> So, for example, a happy face momentarily glimpsed in TV static is a pattern, as it is one of those odd 'minimum suspect phenomenological nullities' that Dalí likely had in mind. Yet it differs in its qualities from other kinds of patterns which are identified through means-ends reasoning related to explanation and prediction: i.e., a repeatable illusion found in a stereogram, or repeated observations of celestial objects and events that behave on schedule (and mathematical *abstracta* concocted on that basis), all have a *Gestalt* sensibility. Yet none are like the malevolent *grue*, since it lacks a *pattern*. And that reason is sufficient to regard it as a case of irrational categorization. A separate question is whether the legal definition of genocide has similar *Gestalt* properties, though some of us think it does not.

I suggested at the outset that an account of irrational intentionality might be useful to the substantial theorist, who believes in ‘ends-in-themselves’. I have not tried to elaborate on this narrow slice of the idea of irrational ends, preferring first to make sense of the broader category of irrational intentionality. My hope is we could figure out how the idea supports or subverts theories in politics, law, or justice only afterwards. Even here, the findings are only partial. After all, a theory of irrational intentionality only provides us with a set of tools that provide a negative result, i.e., which allow us falsify assertions about the status of ends-in-themselves. So, for instance, we have not tried to weigh in with an account of intrinsic value that decisively settles the status of the rational good will (*eleutheronomy*) as opposed to happiness (*eudaimonia*). We have only provided a test of what ends-in-themselves *cannot* be -- namely, those cases where a practical judgment is subverted by the orientation that one takes towards it. For example, a moral theory that regarded ends-in-themselves through an orientation that resulted in carelessness, meaninglessness, corruption, and *akrasia*, is a theory that would be guilty of gross confusion – and they would be confused, not merely because of means-ends reasons, but also because they fail a test of intrinsic validity.

But we must be careful not to infer the wrong consequences from this idea. It is entirely possible that someone will have read all this and conclude that they have a new set of rhetorical tools that directly help to assign blame to parties to a deep moral disagreement -- that irrational intentionality is a mark of viciousness, delusion, and so on. But no, that is not the point. For, in the first instance, irrational intentionality is a diagnostic tool in philosophy of mind and a theory of knowledge. When we talk about tractable cases, our only hope is to identify a consideration that, in some broad sense, makes a decisive difference to the ways we might find fault with how people reason, insofar as we have understood their reasoning and situation. It is not necessarily to suggest that cases of irrationality are worthy of blame.

Perhaps the role that a theory of irrational intentionality has in moral inquiry might be better conceived as a way of understanding the role of *negative intuitions* in reasoning, especially practical reasoning. Huckleberry Finn was faced with one such negative intuition.<sup>xiv</sup> Raised in the racist South, Finn had bigoted instrumental opinions about the status of Black people in his society. Yet he also had a negative intuition towards racism when he was forced to reconcile the dogma with the existence of his friend, Jim. Ultimately, Finn found himself unable to act on his considered beliefs, and in that sense was *akratic*. That is also a case of irrational intentionality, insofar as his conations subverted his judgment that Jim was to be turned in, by virtue of what the judgment was about (Jim). Yet that subversion of judgment was praiseworthy, not blameworthy – Finn’s

conscience was reasons-responsive. There is a fault to be found, perhaps even a moral fault, but the fault was with some of the global aspects of his orientation to judgment, not a fault with the decision itself.

The upside to the decision to focus on tractable cases is that they are directed towards our *discursive powers*, i.e., our powers to set the terms of conversational engagement in a way that makes cooperation possible. For without that assumption -- i.e., the assumption that critical discursive engagement really is a going concern -- we are helpless but to deploy absolutist criteria for moral cooperation, developed in a first-personal fashion from transcendental truths, and deployed as scorekeepers in discussion of normative responsibility as if it were a zero-sum game. These are valid possibilities, perhaps, and might be rooted in the *synthetic a priori* propositions that can withstand the scrutiny of a mature cognitive science. But they are not happy places to begin investigation into a peculiar and difficult subject of how to make sense of ends-in-themselves.

I conclude with a suggestion. It seems to me that the presumption of good faith *is best motivated by* other worries I have about our understanding of ends-in-themselves. For whatever intrinsic values might be, they must be rooted in our experiential condition (i.e., not especially spooky) and potentially held in common (i.e., open to rational debate on the basis of shared evidence). Phenomenology helps us with both these requirements. It proposes that our attention is about things, and hence things in common -- thus, we are able to share the very same values or ends. But it strains plausibility to think we all experience things in the same sense -- hence, our values are a function of the ways we are oriented to them. If this seems like a productive aim for inquiry into the study of justice, morals, and the social sciences, then we cannot *presuppose* that intrinsic goods correspond to fixed, prior, categories that are a function of our first-person capacities, necessarily resulting in unique right answers. Nor are such first-personal right answers necessarily precluded.

| <b>About</b>                     | <b>Subversive orientation</b> | <b>Tractable examples</b>             |
|----------------------------------|-------------------------------|---------------------------------------|
| <i>Ends</i>                      | Conation*                     | Akrasia, or weakness of rational will |
|                                  | Attention                     | Carelessness                          |
|                                  | Gestalt                       | Meaninglessness                       |
|                                  | Position                      | Corruption                            |
| <i>Objects</i>                   | Conation                      | Abstracta                             |
|                                  | Attention*                    | Minimal patterns                      |
|                                  | Gestalt                       | Purely syntactic existentials         |
|                                  | Position                      | Paranoid-critical method              |
| <i>Categori-<br/>zations</i>     | Conation                      | Task-opaque concepts                  |
|                                  | Attention                     | Precepts                              |
|                                  | Gestalt*                      | Borgesian categories                  |
|                                  | Position                      | Undifferentiated holism               |
| <i>Beliefs (or<br/>commands)</i> | Conation                      | Motivated inference                   |
|                                  | Attention                     | Monosentential paradoxes              |
|                                  | Gestalt                       | Heracleian beliefs                    |
|                                  | Position*                     | Constitutive ignorance                |

---

<sup>i</sup> The instrumental conception of reason is usually credited to Hume, but in fact he recognized two conceptions of unreason: "... 'tis only in two senses, that any affection can be call'd unreasonable. *First*, When a passion, such as hope or fear, grief or joy, despair or security, is founded on the supposition of the existence of objects, which really do not exist. *Secondly*, When in exerting any passion in action, we choose means insufficient for the design'd end, and deceive ourselves in our judgment of causes and effects." (emphasis preserved from original). The first is one case of irrational intentionality, in my sense. See:

Hume, David. 2000. *A Treatise of Human Nature*. (David Fate Norton and Mary J Norton, eds.) Oxford: Oxford, p. 267.

<sup>ii</sup> Weber, Max. 1978. *Economy and Society*. (Guenther Roth; Claus Wittich, eds.) Berkeley: University of California Press.

<sup>iii</sup> Kolodny & Brunero formulate means-ends reasoning as follows: means are (a) either actions or conations, (b) directed at the use of tools or resources, and in a fashion that is (c) either causal or constitutive of the end. They think that the ends are primarily actions, and not valued states sought to be achieved or honored. Both concepts are rendered in a way that is meant to capture the notion of *instrumental transmission*, associated with Joseph Raz (2005) -- i.e., the idea that if we assume that we have a reason to pursue some end, then we must also have a reason to pursue some means to achieve that end (preferably, the most fitting means to the end).

On the one hand, (c) might have some basis in ordinary speech, since, e.g., "I raise my hand for purposes of voting, and thereby voted" sounds like means-ends reasoning, though it contains only a single action described in two different ways.

But on the other hand, I cannot make much out of the notion of 'transmission' without a corresponding notion of 'transmission cost', or level of cognitive effort at attributing rationality to some action or end. And the effort involved in evaluating the rational propriety of raising one's hand for purposes of voting is just the same as the perceived cost of voting. See:

Kolodny, N. & John Brunero. 2018. "Instrumental Rationality", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2018/entries/rationality-instrumental/>>.

Raz, Joseph. 2005. "The Myth of Instrumental Rationality." *Journal of Ethics & Social Philosophy*. 1(1).

<sup>iv</sup> Nozick, R. 1993. *The Nature of Rationality*. Princeton: Princeton University Press.

Korsgaard, C. 1997. "The Normativity of Instrumental Reason" in G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press.

Wedgwood, R. 2011. "Instrumental Rationality." *Oxford Studies in Metaethics*, 6.

<sup>v</sup> Unfortunately, the adherents of phenomenological accounts of intrinsic value tend to make the mistake of inferring the existence of non-natural properties, when in fact the proper deployment of the phenomenological method would require us to *suspend belief*, for purposes of diagnosis. See:

Lemos, Noah. 1994. *Intrinsic Value: Concept and Warrant*. Cambridge: Cambridge.

Chisholm, Roderick. 2005. "Defining Intrinsic Value." In *Recent Work on Intrinsic Value*. Eds. Toni Ronnow-Rasmussen & Michael J. Zimmerman. Dordrecht: Springer.

<sup>vi</sup> Brandom, R. 1994. *Making it Explicit*. Cambridge: Harvard. p.97-204.

<sup>vii</sup> Kant, I. 2012. Second section. *Groundwork of the Metaphysics of Morals*. (eds. Mary Gregor & Jens Timmermann.) Cambridge: Cambridge.

<sup>viii</sup> Frankfurt, Harry. 1982. "The importance of what we care about." *Synthese*. 53(2) (Nov. 1982)

<sup>ix</sup> Croft, W & D. Alan Cruse. *Cognitive linguistics*. Cambridge: Cambridge.

<sup>x</sup> The conventional term is 'hysteresis', but it is an unfortunate term.

<sup>xi</sup> Anscombe, GEM. 1957. *Intention*. Cambridge: Harvard.

<sup>xii</sup> Davidson, D. 2001. "Actions, Reasons, and Causes." *Essays on Action and Events*. New York: Oxford.

<sup>xiii</sup> For example, if during the process of choosing research questions and projects you are unable or unwilling to exercise your theoretical imagination, then you will not have met the conditions of due diligence and minimal explanatory effort required for objectivity (in the epistemological sense).

<sup>xiv</sup> Mele, A. 2012. *Backsliding: Understanding Weakness of Will*. Oxford. P. 8-11.

---

Sandven, T. 1995. "Intentional Action and Pure Causality -- A Critical Discussion of Some Central Conceptual Distinctions in the Work of Jon Elster". *Philosophy of the social sciences*. 25(3).

Elster, J. 1985. *Making Sense of Marx*. Cambridge:Cambridge.

<sup>xv</sup> Alcott, LM. 2017. "Eurocentrism as an epistemology of ignorance." *Routledge Handbook of Epistemic Injustice*. (Ian James Kidd, José Medina, Gaile Pohlhaus, eds.) New York:Routledge.

<sup>xvi</sup> Indeed, the mere decision to be ignorant is not necessarily *irrational*. Far too many productive cognitive states require the willing and provisional suspension of belief, e.g., for the purpose of entertaining hypotheticals, for purposes of abstraction, and (indeed!) the phenomenological method itself. In these cases, it is entirely rational to will ignorance temporarily to reach a broader point, so long as the suspension of belief does not go on forever.

<sup>xvii</sup> Dali, Salvador. 1935. "Conquest of the Irrational." Archive.org. URL = <https://archive.org/details/DaliConquestIrrational/>

<sup>xviii</sup> Curry, Nicholas J. 2012. *The Emotions and Two Aesthetic Paradoxes*. Dissertation. University of Missouri St. Louis.

Curtis, Adam. 2016. "HyperNormalisation." BBC.

<sup>xix</sup> The idea of disorientation, and the experience of being 'disorientated', is given careful attention by queer phenomenology. On this approach, disorientation is the experience of confusion and ungroundedness that occurs when we switch between orientations. She suggests that some sense of disorientation may be a necessary feature in experimenting with new ways of living, in order to seek out lives worth living well.

Ahmed, Sara. 2006. *Queer phenomenology*. Duke. p.157-179.

<sup>xx</sup> Boghossian, P. 2010. "The Concept of Genocide." *Journal of Genocide Research*.

<sup>xxi</sup> Brandom, R. *ibid*, *supra* note vi. p. 123.

<sup>xxii</sup> Borges, JL. 1975. *Other Inquisitions, 1937-1952*. Austin:University of Texas Press.

<sup>xxiii</sup> Dennett, D. 1991. "Real Patterns." *The Journal of Philosophy*. 88(1)

<sup>xxiv</sup> Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. New York:Oxford. [ebook version]