Circular Definitions of 'Good' and the Good of Circular Definitions

Andrés G. Garcia[1] [ID]

Abstract

I defend the view that circular definitions can be useful and illuminating by focusing on the fitting-attitudes analysis of value. This definition states that an item has value if and only if it is a fitting target of attitudes. Good items are the fitting targets of positive attitudes, and bad items are the fitting targets of negative ones. I shall argue that a circular version of this definition, defended by Rabinowicz and Rønnow-Rasmussen (2006), is preferable to its non-circular counterpart and upholds reasonable standards of acceptability. The standards I will be discussing come from Humberstone (1997), who claims that definitions cannot be informative as long as they are *inferentially circular*.

Keywords Value · Circularity · Definition · Analysis · Fitting Attitudes · Value-making

1 Introduction[1]

Conceptual analysis can be understood broadly as the task of clarifying a target concept by offering a general account of its application conditions. The result is often meant to act like a definition, capturing *a priori* connections between the target and other concepts in our everyday thoughts and practices. One standard example states that a person is a bachelor if and only if that person is an unmarried man. A requirement often imposed on such definitions is that they avoid circularity, meaning they may not use a target concept when specifying its application conditions. An example of a circular definition states that a person is a bachelor if and only if they *identify as a bachelor*. I will argue against the requirement and

---

✉ Andrés G. Garcia
andres.garcia@fil.lu.se

[1] Dept. of Philosophy, Lund University, Lund, Sweden

defend the view that circular definitions can be useful and illuminating by focusing on an example from analytic axiology.[2]

The fitting-attitudes analysis (hereafter referred to as "FA") states that an item has value if and only if it is the fitting target of attitudes. Good objects are the fitting target of positive attitudes, and bad objects are the fitting target of negative ones. I shall argue that a circular version of the definition, first proposed by Rabinowicz and Rønnow-Rasmussen (2006), retains some of the benefits of FA while avoiding some of its biggest problems. The account (hereafter referred to as "CFA") states that an object has value if and only if the object is a fitting target of attitudes in virtue of factors that make the object valuable. Good objects are fitting targets of positive attitudes in virtue of factors that make them good, and bad objects are fitting targets of negative attitudes in virtue of factors that make them bad. I will argue that, contrary to appearances, this definition is informative and upholds reasonable standards of acceptability.

2 The Circular Definition of Value

The circular definition of value represents a compromise between two storied traditions from the history of analytic axiology. One was spearheaded by G. E. Moore at the start of the 20th century when he insisted on the irreducibility of value. Moore famously stated, "If I am asked, 'What is good?' my answer is that good is good, and that is the end of the matter." He continued: "Or if I am asked 'How is good to be defined?' my answer is that it cannot be defined, and that is all I have to say about it" (1993/1903, p. 58). One rival to this view insists that the concept of value *can* be defined by appealing to a combination of attitudinal and other normative concepts. Brentano laid the foundations of the approach during the latter part of the 19th century when he stated that the good is "that which is worthy of love, that which can be loved with a love that is correct" (1889/1969, p. 18).[3]

The popularity of the reductive approach was spurred on a century later by "the normativity revolution" (Reisner 2015, p. 189), which resulted in philosophers becoming interested in aspects of normativity aside from value and morality: Ought, correctness, fittingness, propriety, and reasons in general, became much more of a focus than they had been before.[4] Meanwhile, philosophers also began treating some such concepts as constituting the bedrock of normativity, with reasons, in particular, emerging as the most popular contender.[5] Raz captured the emerging paradigm by stating that the "normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons" (1999, p. 67).

---

[2] Philosophers tend to assume that circular definitions are not informative (Antonelli 2000, pp. 1–3; Gupta 2000, p. 123; Burgess 2008, p. 215; Keefe 2002, p. 275), but some of them take a more nuanced view. Humberstone (1997) develops a criterion on which circular definitions can be informative if they avoid "inferential circularity." The circular definition of bachelorhood upholds the criterion, for while the concept of a bachelorhood reoccurs within the specifications of its application conditions, it does so in a protective intentional context. I shall say more about this later.

[3] Another form of reductionism not discussed here is *analytic naturalism*, which attempts to define value in naturalistic terms. For a prominent example, see, e.g., Jackson (1998).

[4] One of the philosophers most responsible for this is Scanlon (1998), who resurrected the approach of Brentano through his own "buck-passing account." I will here treat the buck-passing account as a version of FA that interprets claims of fittingness in terms of reasons. Scanlon also had a particular metaphysical view of the precise manner in which value is grounded in reason-giving properties that I will leave aside here.

[5] The concept of *ought* was another. See, e.g., Broome (2013, pp. 49–57).

Skorupski similarly said that the "sole normative ingredient in any normative concept is the concept of a reason" (2010, p. 2).[6] As a result, many normative categories were understood in terms of reasons, resulting in the following version of Brentano's original approach becoming widely accepted:

FA: x has value $=_{df}$ there are reasons to direct attitudes toward x.

Definitions like this are interesting because of how useful they have been for the pursuit of a variety of theoretical tasks. This includes the task of defining monadic concepts such as *final value* (e.g., Chisholm 1981; Orsi 2015, p. 28; Rowland 2019, p. 7), *instrumental value* (e.g., Orsi 2015, p. 28; Rowland 2019, p. 13), *personal value* (e.g., Darwall 2002, p. 8; Rønnow-Rasmussen 2007; Orsi 2015, pp. 66–68; Rowland 2019, pp. 81–84), *kind-value* (e.g., Skorupski 2010, pp. 82–87; Garcia 2019; Rowland 2019, pp. 91–94), as well as comparatives like *betterness*, *worseness*, *equality*, and *parity* (e.g., Gert 2004; Rabinowicz 2008, 2012). FA has also been used to determine what object types can be value bearers (Rabinowicz and Rønnow-Rasmussen 1999, 2001; Zimmerman 2001: Ch. 3; Orsi 2015, pp. 35–39), what relations hold between values of wholes and values of parts (Dancy 2003, pp. 630–631), in what sense value depends on attitudes (Orsi and Garcia 2021, 2022; Rabinowicz and Rønnow-Rasmussen 2021), and whether valuable items are always valuable *in a respect* (Zimmerman 2001, pp. 20–26; Orsi 2015, pp. 58–61). The list goes on. Now, I cannot argue the point in detail here, but it seems that what makes FA useful in these contexts is not its reductive character but its establishment of a connection between value and fittingness—or reasons for attitudes. There are no grounds for thinking that a circular counterpart of the definition could not be applied in many of those contexts.

Such an account also appears less vulnerable to some difficult problems facing FA, among which is the infamous *wrong kind of reasons problem* (hereafter referred to as "WKR").[7] The problem is that there appear to be cases where value and reasons for attitudes come apart, so there are reasons to direct positive attitudes toward bad items, negative attitudes toward good items, and so on. In the classic case, an Evil Demon threatens a person with torture unless she admires his worthless saucer of mud for its own sake.[8] That she will be harmed unless she admires the saucer of mud gives her reasons to do so, but this does not imply that the saucer of mud is admirable for its own sake.

The standard solution, inspired by the remarks of Parfit (2001; 2011, App. A) and championed by Skorupski (2007), suggests that the Evil Demon's threatening ways do not provide reasons to admire his saucer of mud for its own sake. Instead, the threatened person is given reasons to *make it come about* that she admires the saucer of mud or *to want* to admire the saucer of mud. Hints for this can be found in the fact that her reasons are grounded in facts about the *usefulness* of her admiring the saucer of mud for its own sake rather than anything to do with the saucer of mud. The standard solution is ultimately unsatisfactory,

---

[6] Treating concepts as the bedrock of normativity has often meant treating them as primitive. For example, Scanlon (1998) identifies reasons as the bedrock of normativity. "A reason for something," he says, is "a consideration that counts in favor of it," and he adds that this concept "presupposes" that of a reason (1998, p. 17). However, the bedrock of normativity need not be primitive, for what is fundamental with respect to the normative domain need not be fundamental with respect to others.

[7] See Rabinowicz and Rønnow-Rasmussen (2004) for a discussion.

[8] The case was first thought up by Crisp (2000).

partly because it seems incorrect to say she is not given reasons to admire the saucer of mud for its own sake. After all, the Evil Demon only cares that she makes it come about that she admires it to fulfil his actual demand, which is that she admires the saucer of mud for its own sake.[9]

The apparent lesson of WKR is that the value of items only corresponds to reasons for attitudes that are explained by properties that make the items valuable in the first place.[10] That the Evil Demon is threatening does not seem to speak to the fact that the saucer of mud is admirable for its own sake, and this makes that factor seem irrelevant from the point of view of the definition. Adjusting FA accordingly may result in a circular definition of value that looks very like that of Rabinowicz and Rønnow-Rasmussen (2006, p. 119):

CFA: x has value $=_{df}$.

1. x has properties that make it valuable.
2. there are reasons to adopt attitudes toward x.
3. the properties mentioned within 1 explain why 2 is true[11]

The definition is circular because the target it aims to define, namely the concept of value, is also referenced in the specification of its application conditions, namely within the claim about value-makers.[12]

Given analytic philosophy's broader context and history, the aversion to circularity is not difficult to understand. For much of its early life, the tradition was as characterized by its commitments to the reductive ideal as to any particular methodology. This is relevant because a circular definition cannot hope to be reductive in the sense that it gives an exhaustive account of its target in terms of other concepts. However, even as analytic philosophy started moving away from its reductive commitments, many of its practitioners still saw circularity as something to be avoided (cf., Burgess 2008, p. 218).[13] It has only recently

---

[9] Again, see Rabinowicz and Rønnow-Rasmussen (2004) for a discussion.

[10] Note that everyone agrees that there is *some* connection between thick values (e.g., admirability, respectability, desirability) and reasons for positive attitudes (e.g., admiration, respect, desire). The challenge of distinguishing reasons that speak to the presence of value from those that do not is, therefore, a quite general one that needs to be answered irrespective of whether we accept FA. This is often overlooked by some of FA's most ardent critics.

[11] Rabinowicz and Rønnow-Rasmussen (2006, pp. 119–120) introduce circularity by understanding the right kind of reasons for attitudes as those that "invoke" their value-making properties. Bykvist (2023) also takes this to mean that they accept something like CFA. It is important to note that my formulation of CFA is silent on the mechanisms whereby 1 makes 2 true. For example, it leaves open that 1 makes 2 true by facts about the properties mentioned in 1 becoming the reasons referred to by 2.

[12] For some other circular versions of FA, see, e.g., Wiggins (1987, p. 195) and Orsi (2015, pp. 156–157). See also Bykvist (2023), who defends CFA because it escapes WKR. I also make the same point in an earlier work of mine. See (Garcia, 2018, Appendix).

[13] Quine (1951) is a striking example of an analytic philosopher who rejects reductionism while apparently seeing circularity as a vice, as is evidenced by his criticisms of the concept of *analyticity*. He seems to reject several definitions because they implicitly invoke the concept when specifying its application conditions, which is striking given Quine's anti-reductionism and willingness to accept holism about epistemic justification and the like. To be fair, he later suggested that circularity was never the issue: "Repudiation of the first dogma, analyticity, is insistence on empirical criteria for semantic concepts: for synonymy, meaning" (1991, p. 272).

become commonplace to see defences of specific instances of circularity in as varied fields as semantics, aesthetics, epistemology, and metaphysics.

The question remains whether there are grounds for the aversion independently of any commitments to reductionism, for it seems clear that circular definitions cannot be rejected simply for being *inaccurate*. Whatever else might be said about it, a definition stating that an object is valuable if and only if it has value *is* accurate. If circular definitions are to be rejected, they must be because they are considered *uninformative*.[14] Perhaps part of the reason why CFA may appear so is that philosophers want definitions to uncover the contents of their targets in a way that provides an identity between them and their conditions of application. Circular definitions provide equivalencies, which philosophers should *also* be concerned with clarifying, but in so doing, they are also tasked with saying more about *why* the equivalencies hold. From this perspective, circular definitions may appear somewhat incomplete.

The task of answering this challenge cannot be undertaken here. Still, some points of departure invite themselves and are worth mentioning, one of which represents a kind of *conceptual holism*. On this view, circular definitions provide a map to situate a target within a larger conceptual structure (cf., Rabinowicz and Rønnow-Rasmussen 2006, p. 120). The structure itself may be circular, so a full grasp of one concept requires a grasp of the whole. Normative concepts, such as value and reasons, form central parts of such a whole while still relying on other normative concepts, like ought, correctness, fittingness, and propriety. Grasping the content of a concept may be a matter of understanding its contribution within a larger conceptual structure and the inferences it allows us to make.[15] If this is right, then it may well be that no definition of an individual concept could ever hope to be accurate and complete.

Definitions can illuminate in at least two ways. Burgess (2008, p. 2017) points out that some definitions are *pedagogically illuminating*, meaning they can successfully impart conceptual competence to neophytes. If a person can learn what it means to have value by grasping a definition of the concept, then the definition is pedagogically illuminating. Burgess (ibid.) also points out that some definitions aim to be *philosophically illuminating* by highlighting the requirements that underpin or can be abstracted from an already established conceptual competence (cf., Humberstone 1997, p. 251). After making this distinction, he says that although "definitions might be both pedagogically and philosophically illuminating, we have no reason to expect this in general, and absolutely no right to demand it of any would-be definition" (Burgess, 2008, p. 217).[16]

The points about conceptual holism and the nature of philosophical illumination indicate that some demands placed by philosophers on definitions have been too strict. Of course, this brings to the fore difficult questions about what definitions deserve to be referred to as

---

[14] Gertken and Kiesewetter (2017, p. 3, n. 14) acknowledge that circular definitions can be informative but suggest that a reductive one is even more so, *ceteris paribus*. However, it is unclear whether all else is equal since CFA escapes WKR and displays a richer structure than FA. Whether this is enough to compensate for the loss of reduction is controversial, but it seems safe to suggest that CFA is not so much less informative that it warrants outright rejection. Cf., Bykvist (2023, p. 96).

[15] Cf., also the non-reductive method of conceptual analysis favoured by Strawson (1992), which is about outlining the connections between a system of otherwise primitive concepts.

[16] An analogy might be found in the formulation of grammatical rules, for such rules can clearly be illuminating even to those who speak perfectly (cf., Humberstone 1997, pp. 255–256). Grammatical rules may not impart linguistic competence but shine a light on the underpinnings of such competence.

"analyses" in the first place. I have so far talked of conceptual analysis as the *task* of clarifying a concept, but it seems clear that philosophers also use the term to refer to the definitions that are the intended results of that task.

Reisner (2009) argues that FA does not bear the marks of a successful analysis. He says that the definition is meant to show that reasons for attitudes are metaphysically or conceptually prior to value, but he points out this does not fit well with the apparent *epistemic priority* of the latter. Reisner argues that to determine whether a reason for attitudes is of the right kind to speak to the presence of value, we rely on our evaluative intuitions about what is good and what makes things valuable in the first place. We judge that the Evil Demon's threatening ways do not give us reasons of the right kind, and we reach this judgment by observing that his saucer of mud is not admirable and that the Evil Demon's threatening ways do not make it so. CFA may seem to get away unscathed here since it denies that reasons for attitudes are conceptually prior to value, and it is silent about which of the two has metaphysical priority, but perhaps matters are not that simple. Reisner insists that for a definition to be considered a successful analysis, it *should* capture a flow of information going to its target concept from the specification of its application conditions.

Bykvist (2023) has also defended CFA and argued that it captures important conceptual truths about value while avoiding WKR.[17] However, he is sensitive to the issues raised by Reisner and so acknowledges that CFA may not deserve the relevant honorific:

> We can no longer say that the defining concepts are *prior* to the target concept, for one cannot grasp the defining concepts (which contain the concept of goodness) without first grasping the target concept of goodness. Nor can we say that we can determine the *extension* of 'x is good' by determining the extension of 'x has features that make x good and which provide S reason to favour x', for in order to determine the extension of the latter expression we need to know the extension of 'x is good'. In Humberstone's words, we have a case of *inferential circularity*. This should make us question whether circular FA should be seen as a philosophical analysis rather than just an informative conceptual truth. (Bykvist, 2023, p. 90)

Reisner and Bykvist suggest that we *ought* not to refer to any correct definitions resulting from analysing concepts as successful *analyses*. This is a conceptual ethics claim about how we *should* employ the concept of an analysis, and it is unclear why we should accept it. Of course, it is useful to be able to distinguish different kinds of definitions in a way that highlights their aims, potential weaknesses, and so on. Still, we can do this by using the appropriate qualifiers. For example, whenever we offer up a definition of a concept, we can state whether it is meant to constitute a *reductive analysis* (aiming for pedagogical *and* philosophical illumination) or a *non-reductive analysis* (aiming for just philosophical illumination), thus clarifying what sorts of obstacles it needs to overcome to be considered plausible. To me, it seems natural that clarifying a concept's analytic connections should be seen as the task of *analysing* it and that anything that results from this task be regarded as an *analysis*. Nevertheless, I am less interested in the conceptual ethics question to which I have just alluded and more interested in determining how much of a loss it would be to accept a

---

[17] I also argue this point in a brief and very sketchy appendix in an earlier work. See, e.g., Garcia (2018, Appendix).

circular definition of value in the first place. For this reason, I move on to consider the supposed inferential circularity of CFA.

3 Upholding the Ban on Inferential Circularity

Humberstone states that a definition is inferentially circular if and only if "any argument or inference from premises claiming the various conditions provided by the account to obtain, to the conclusion that the concept applies, is itself circular in whatever sense an inference or argument can be circular" (1997, p. 250). Burgess (2008) notes an ambiguity in talk about establishing inferential circularity since we may be interested in confirming or falsifying the conditions provided by definitions. Making the ambiguity explicit results in the following:

> BAN: A definition of the form 'x is F iff x is N' must not be inferentially circular in the sense that (a) to confirm that x is N you must first confirm that x is F, or (b) to falsify that x is N you must first falsify that x is F.

To illustrate these conditions, Burgess (2008, pp. 221–222) asks us to consider two circular definitions of the concept of being a cow:

> C1: x is a cow $=_{df}$ King Charles *knows* that x is a cow.
> C2: x is a cow $=_{df}$ King Charles *believes* that x is a cow.

Burgess (ibid., p. 222) notes that the latter escapes inferential circularity since we can confirm and falsify that King Charles *believes* Bessie is a cow without confirming and falsifying her bovinity, respectively. Burgess (ibid., p. 221) also notes that the former is inferentially circular because we cannot confirm that King Charles *knows* Bessie is a cow without confirming her bovinity.[18] We *can* falsify that King Charles knows that Bessie is a cow without falsifying her bovinity, e.g., by falsifying that he *believes* her to be a cow.

Burgess (ibid.) suggests that a definition is useless if it "could not teach us a word that was new to us" nor "introduce us to, or illuminate, an application procedure". The former definition helps illuminate an informative *falsification* procedure since it implies that to falsify that Bessie is a cow, we can drive to Buckingham Palace and ask King Charles what he believes and why. The definition does less well in illuminating an informative *confirmation* procedure; for it entails that to confirm that Bessie is a cow, we must visit the farm and, well, confirm that Bessie is a cow. This makes the definition inferentially circular.

The latter definition does better because belief differs from knowledge in lacking commitment to *truth* (Humberstone 1997, p. 253). [19] In other words, it specifies the application conditions of *bovinity* by invoking that very concept, but it does so alongside the concept of belief, which appears to *insulate* said conditions from facts about Bessie's actual bovinity.

---

[18] To confirm that King Charles knows that Bessie is a cow, we need to both visit the farm to have a look at Bessie *and* visit Buckingham Palace to ask Charles what he believes and why. In this way, C2 also has the disadvantage of forcing us to spend more money on petrol.

[19] This means that the former definition affords a protective embedding to its conditions of application that the latter lacks. Humberstone (1997) explicates the notion of protective embedding by appealing to the semantic notion of *compositional independence*. For a closer look, see Keefe (2002).

This captures the intuitive sense in which the intentional concept of belief gives the concept of bovinity what Humberstone (1997, pp. 259–68) refers to as a "protective embedding."

Humberstone (1997: § 3, § 5, Appendix B) goes to great lengths to explain the details of his notion of protective embedding, but its intuitive motivation seems sufficient guidance in the present context. Very roughly, to say that a target concept is suitably insulated or has been afforded a protective embedding is to suggest that its occurrence in the specification of its application conditions does not entail that we have to establish the applicability of the target before establishing that its application conditions obtain. For further illustration, consider the dispositional account of colour defended by Burgess (2008). The account says that for something to be red, it has to appear red to suitable observers under normal circumstances. Proponents must explain what counts as suitable and normal to make the account work, but this problem should not detain us here. The question is whether the circularity displayed by the account renders it uninformative. Burgess (2008, pp. 225–229) does not think so since the concept of redness occurring within the specified application conditions is also suitably insulated. More specifically, the intentional concept of *appearing* gives the concept of redness its protective embedding. It ensures that to establish that something appears red to suitable observers under normal circumstances, we need not first establish that it is red.

The question is whether the circular definition of value suffers the disadvantage of inferential circularity, and so we must look to the application procedures entailed by CFA. In particular, since it is the notion of value-making that produces the circularity, it is this notion that risks throwing a wrench into the works: To confirm and falsify that a factor makes an item valuable, must we first confirm or falsify that the item has value? An affirmative answer entails that the application conditions specified by CFA are not adequately insulated from the reality of the item's value and that the definition is, therefore, inferentially circular.

Rabinowicz and Rønnow-Rasmussen (2006) concede that the circular definition is inferentially circular, as does Bykvist (2023), but since this issue depends on how we understand the relevant notion of value-making, we should consider it more carefully.

There are at least two ways of understanding value-making, one of which takes it to be a *success notion* and the other a *contributory notion*.[20] If we use the former understanding, then we are committed to saying that any factor that fails to make an item valuable cannot be a value-maker with respect to that item. However, this is not the case if we use the latter. To establish that certain properties *contribute* toward making an item valuable, we need not establish that the item has value. The idea is that contributory value-makers push items *in the direction of value* but can fail to get them all the way there. Adjusting CFA to make this more explicit results in something like the following:

CFA1: x has value $=_{df.}$

1. x has properties that *contribute* toward the value of x.
2. there are reasons to adopt attitudes toward x.
3. the properties mentioned within 1 explain why 2 is true.

Grasping what it means for properties to contribute toward value requires grasping what it would mean for those properties to succeed. This does not yet show that to confirm and falsify that 1–3 obtain, we must first confirm or falsify that the relevant item has any value. We

---

[20] See Beardsley (1973) who makes a similar distinction when he discusses the nature of *aesthetic value*.

can establish by simply looking at the individual properties of the item, whether these properties contribute to its value, but this does not require that we first look at the item as a whole and establish its evaluative status. In this way, the concept of *contribution* seems to provide the concept of value-making with the protective embedding that it needs, just like *belief* did for *bovinity* and *appearance* did for *redness* in the other examples mentioned above.

The obvious objection suggests that if a property contributes toward the value of an item, then the item has *contributive value* as far as its possession of the property is concerned. I take this to mean that the item must have value *in a respect*, which would make the definition equivalent to the following:

CFA2: x has value $=_{df}$.

1. x has properties that make it valuable in a contributive sense.
2. there are reasons to adopt attitudes toward x.
3. the properties mentioned within 1 explain why 2 is true.

The objection entails that all notions of value-making must be success notions after all, and so appears to be question-begging in the present context. It insists that if a property contributes toward an item's value, then the item *must* be valuable in some sense.[21] Nevertheless, perhaps the objection does highlight a challenge on my part to provide some independent reasons to accept a concept of contributive value-making in the first place. The matter is largely intuitive and can be strengthened by appeals to examples and analogies.

Take the claim that integration and complexity in an information system contribute to the presence of consciousness in that system, even though it may still fail to be conscious. Everyone presumably agrees that in saying this, we would not be committed to thinking that the system is conscious as far as its present level of integration and complexity is concerned. We might stipulate a technical concept of contributory consciousness to discuss the explanatory potential of its apparent integration and complexity, but we would not take this to denote a genuine type of consciousness. While the relation of value-making certainly differs from that of *mind-making*, the intuition is that something analogous can happen in the case of value. Just as complexity and integration can contribute toward the consciousness of a mindless system, factors can contribute to the value of valueless items. This may not be the standard sense of value-making, but it seems a genuine explanatory relation, and if it can be invoked when speaking about the mind, then it is unclear why it should be out of bounds when we talk about value. Recall the criterion that CFA is supposed to pass:

BAN: A definition of the form 'x is F iff x is N' must not be inferentially circular in the sense that (a) to confirm that x is N you must first confirm that x is F, or (b) to falsify that x is N you must first falsify that x is F.

---

[21] A similar objection states that to establish that certain properties contribute toward the value of an item, we must establish that the item has *potential value*. However, this appears even less of a problem since this concept does not fall entirely within the intended domain of circular definitions of value. These definitions are just meant to capture what it means for things to have value, and so we need an independent account of what it means for an item to be disposed to do *anything*, including being disposed to either *have value* or *make things valuable*.

To confirm and falsify that something merely contributes to an item's value, do we need to confirm or falsify that it has any sort of value? Suppose we are art critics tasked with judging the beauty of an abstract painting. While we find that the painting is crudely done, the interplay of its blue and orange colours stands out to us. The contrasting colours seem to be contributing toward the beauty of the painting, and yet, as far as we can see, the painting is not beautiful. There are no reasons to admire the painting, even with an eye toward the contrasting colours that pique our interest. If the artist were to react to our criticism by insisting that we are committed to finding the painting beautiful as far as its possession of the contrasting colours is concerned, then we would be right to dismiss him as desperate. The contrasting colours contribute toward the beauty of the painting as complexity and integration of information contribute toward the consciousness of a mindless system. If we are sympathetic and want to avoid hurting his feelings, we could stipulate a technical concept of contributory beauty to discuss the relevant explanatory relation, but we need not take this to denote a genuine value. After all, an interesting interplay of contrasting colours might be found almost everywhere around us, and while such properties may push things in the direction of beauty, they may only be successful in doing so in the right sort of contexts, within, for example, the right kinds of pictorial compositions.[22]

Suppose that we are asked to judge the tastiness of a fresh Kadhi soup. Tasting the soup, we find it is creamy and incorporates the right spices for its kind, but it is not prepared correctly and not tasty. Just as in the previous example, the cook would be desperate to react to our judgment by insisting that we are committed to finding his Kadhi soup tasty as far as its creamy textures and spices are concerned. The fact that the soup is creamy and incorporates some of the right spices pushes the soup toward tastiness, but only in the sense that contrasting colours may contribute to the beauty of a completely unremarkable work of art. Once again, if we want to avoid arguing with our hapless cook, we could stipulate a concept of contributory tastiness to speak about the explanatory relation just alluded to, but this would not commit us to judging that there are reasons to enjoy or admire the Kadhi soup. Some of the features that push the soup toward tastiness, like its creamy texture, may be found all around us, like a puddle of mud that we do not have any reasons to enjoy or admire, even if we happen to do so with an eye specifically toward its creaminess.

One response states that these examples only demonstrate that some properties are necessary but insufficient parts of overall conditions for value that are themselves sufficient but unnecessary. This is not controversial, nor does it undermine the general point that contributory value-makers must succeed in making items valuable in a respect. However, the intuition that is meant to be captured by these examples is not just the modal one that certain factors can only be value-making in combination with other factors. The intuition being pressed here is that certain factors *do* contribute to the value of objects even though those factors do not give rise to any value on their own.[23] There are no reasons to appreciate either the painting for its contrasting colours or the soup for its creaminess, even if both of these properties do succeed in pushing their respective bearers in the direction of value.

---

[22] I make no assumption here about the irreducibility of goodness or badness with respect to value relations. For example, if one adopts a view on which the good is simply that which is better than the neutral, then my claim is that just because an item is in possession of properties that *contribute* toward its value, this does not commit us to judging the item better than the neutral—even in the narrow sense of being better *in a respect*.

[23] Wygoda Cohen (2020), who distinguishes factors that only ground higher-order properties in combination with other factors and factors that ground by themselves but are still not sufficient to give rise to the higher-order properties *on their own*.

Admittedly, if the contrasting colours of a painting push it in the direction of beauty, then the state of the painting *having* those colours is likely to have value. Similarly, if the creamy textures and spices of the Khadi soup push it toward tastiness, then the state of the soup having these properties may have value. This seems right. That *something* has value due to an item's possession of contributory value-makers does not entail that the item itself must have value.

Another rejoinder insists there must be *something* good about the painting, for although its properties may not be enough to make it beautiful (even beauty *in a respect* may require more than a clever interplay of contrasts), they are enough to afford the painting an *aesthetic value*. The idea is that while everything beautiful has aesthetic value, not everything with aesthetic value is beautiful. Similarly, there must be *something* good about the soup, for although its properties may not be enough to make it tasty (even tastiness *in a respect* may require more than creaminess and certain species), they are enough to afford the painting a *culinary value*. Everything tasty has culinary value, but only some things with culinary value are tasty. If this is correct, then we are once again at risk of having to define some values in terms of others. For example, consider the following:

CFA3: x is beautiful$=_{df}$.

1. x has properties that make it aesthetically valuable.
2. there are reasons to appreciate x aesthetically.
3. the properties mentioned in 1 explain why 2 is true.

The strategy is quite similar to the one I mentioned earlier, suggesting that contributory value-makers must at least succeed in making their bearers contributively valuable. One problem is that this latest strategy will not work for our most generic values, like overall value or *just plain goodness*. Consider the type of generic value judgment we might hold for entities like possible worlds. We might suggest that a world has properties that push it in the direction of value without the world being valuable. It needs to be clarified how the strategy employed in the beauty and tastiness cases should be applied here, for there is no more general type of value than overall value or just plain goodness to which we can appeal. The critic is left with the option to deny the intuition at play here, at which point the discussion devolves into mere table thumping.

It seems plausible to me to suggest that while the properties of an item can be positive contributors to its goodness, they may not be enough to make it good *in any way*. I also suspect this phenomenon may occur elsewhere and that there is an important but neglected program in analytic metaphysics that tries to come to grips with the general phenomenon of contributive explanations. I doubt hunting for additional examples can break the present stalemate. The next best thing is to try to explain away the intuition that if something contributes toward the value of an object, then some value must result. In fact, I have already given hints about what such an explanation might look like by suggesting that it is a good state *that* a painting has contrasting colours that could—if circumstances were only slightly different—make the painting beautiful. It is also good *that* the Khadi soup is creamy and includes spices of the right sort since the cook could then tweak his recipe and make tastier soups. While there may be something good or bad about the state that certain contributive

value-makers are instantiated in an item, this does not necessarily mean that the item itself must be good in any respect.

If it is true that CFA escapes the charge of inferential circularity, then this speaks well of the ban defended by Humberstone. After all, it certainly seems illuminating to learn that conceptual competence concerning value involves the confirmation and falsification procedures highlighted by the definition. Had CFA not survived the charge, then this might give us reasons for scepticism about the validity of Humberstone's ban. *Ceteris paribus*, accurate definitions are more interesting than those that are not accurate, and there are reasons to believe that CFA is more accurate than its non-circular counterpart. Indeed, some of those reasons go beyond any considerations concerning WKR. To make this point, I wish to draw some analogies between the present discussion and Burgess' aforementioned views on colour concepts. He states that an item has a certain colour if and only if it is perceived as having that colour by normal observers and in normal circumstances. He then observes that circular definitions of this sort "can be the best available precisely because they gesture toward the need for supplementation in a way non-circular definitions do not":

> If our early philosophical counselors were right, we need to experience redness to master the concept of redness. We need to experience redness in order to know what redness is. The dispositional theory of colour not only makes this need apparent, it locates the need in just the place where we would expect to find it—in our ability to know when something does, and when it does not, appear red. How could a definition that failed to do this be philosophically more perspicuous than one that succeeded? (Ibid, pp. 232–233)

A similar line of reasoning may apply to CFA, *mutatis mutandis*. One of the most distinctive aspects of value is its normative pull, which can be captured in the language of reasons. Still, intuitively, this leaves a residual evaluative aspect that cannot be captured in any other language—just as Moore insisted. This primitive aspect of value makes properties of objects stand out and endows them with a certain "evaluative gloss." That this evaluative gloss should be conveyable to neophytes through reductive definition may be just as impossible as imparting conceptual competence regarding colour concepts to those without colour vision. To acquire the whole concept of value, perhaps we must also be acquainted with its peculiar evaluative gloss. As Bykvist also notes (2023), this general kind of observation fits well with circular versions of FA.[24] While such analyses cannot relieve their target of all its apparent mystery, this should be counted among the main benefits of circular versions of FA since value *is* inescapably mysterious. The biggest problem with FA is not WKR or any similarly technical objection but the fact that it aims to do away with central features of the evaluative.

4 Concluding Remarks

I have argued that a purely contributive notion of value-making allows us to say that certain factors push objects in the direction of value without getting them all the way there. The idea is that factors can contribute toward the value of things without the objects being valuable in

---

[24] Cf., also Garcia (2018, Appendix).

any literal sense. If this is true, we can confirm that an object has value-makers without first confirming that the object has value, just as we can falsify that an object has value-makers without first falsifying that the object has value. This shows that the contributive notion of value-making can supply definitions that invoke it with a suitably protected embedding, which will help them avoid the charge of inferential circularity. If my argument is correct, the circular definition of value considered by Rabinowicz and Rønnow-Rasmussen (2006) fulfils reasonable standards of acceptability.

Declarations

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Antonelli GA (2000) Virtuous circles: from fixed points to revision rules. In: Gupta A, Chapuis A (eds) Circularity, definition, and Truth. Indian Council of Philosophical Research, pp 1–27
Beardsley MC (1973) What is an aesthetic quality? Theoria 39(1–3):50–70
Brentano F (1889/1969) The origin of our knowledge of right and wrong. Routledge
Broome J (2013) Rationality through reasoning. Wiley
Burgess JA (2008) When is circularity in definitions benign? Philosophical Q 58(231):214–233
Bykvist K (2023) They smiled at the Good and Frowned at the bad': the fitting attitude analysis reconsidered. In: Garcia AG, Gunnemyr M, Werkmäster J (eds) Value, Morality & Social reality: essays dedicated to Dan Egonsson, Björn Petersson & Toni Rønnow-Rasmussen. Department of Philosophy, Lund University, Lund
Chisholm R (1981) Defining intrinsic value. Analysis 41(2):99–100
Crisp R (2000) Review of 'Value … and what follows' by Joel Kupperman. Philosophy 75(3):452–462
Dancy J (2003) Are there organic unities? Ethics 113(3):629–650
Darwall S (2002) Welfare and rational care. Princeton University Press, Princeton
Garcia AG (2018) *Between Values and the World: Studies in second-order value theory*. (Doctoral dissertation) Lund University
Garcia AG (2019) The finality and instrumentality of value in a way. Philosophia 47(3):681–692
Gert J (2004) Value and parity. Ethics 114(3):492–510
Gertken J, Kiesewetter B (2017) The right and the wrong kind of reasons. Philos Compass 12(5):e12412
Gupta A (2000) On circular concepts. In: Gupta A, Chapuis A (eds) Circularity, definition, and Truth. Indian Council of Philosophical Research, pp 123–153
Humberstone IL (1997) Two types of circularity. Philos Phenomenological Research: Q J 57(2):249–280
Jackson F (1998) From Metaphysics to Ethics: a defence of conceptual analysis. Clarendon, Oxford
Keefe R (2002) When does circularity matter? Proc Aristotelian Soc 102(3):253–270
Moore GE (1993/1903) Principia Ethica. Cambridge University Press
Orsi F (2015) Value theory. Bloomsbury Academic
Orsi F, Garcia AG (2021) The explanatory objection to the fitting attitude analysis of value. Philos Stud 178(4):1207–1221

Orsi F, Garcia AG (2022) The new explanatory objection against the fitting attitude account of value. Philosophia, 1–16

Parfit D (2001) Rationality and reasons. In: Egonsson D, Josefsson J, Petersson B, Rønnow-Rasmussen T (eds) Exploring practical philosophy: from action to Value. Ashgate, Aldershot, pp 17–39

Parfit D (2011) On what matters. Oxford University Press, Oxford

Quine WVO (1951) Two dogmas of empiricism. Philos Rev 60(1):20–43

Quine, WVO (1991) Two dogmas in retrospect. Canad J Philos 21(3):265–274.

Rabinowicz W (2008) Value relations. Theoria 74(1):18–49

Rabinowicz W (2012) Value relations revisited. Econ Philos 28(2):133–164

Rabinowicz W, Rønnow-Rasmussen T (1999) A distinction in value: intrinsic and for its own sake. Proc Aristotelian Soc 100(1):33–49

Rabinowicz W, Rønnow-Rasmussen, T. (2001). Tropic of Value. In: Carlsson E, Sliwinski R (eds.) Omnium-gatherum; Philosophical essays dedicated to Jan Österberg on the occasion of his sixtieth birthday. Uppsala Philosophical Studies, Vol. 50, pp. 263–277.

Rabinowicz W, Rønnow-Rasmussen T (2004) The strike of the demon: on fitting pro-attitudes and value. Ethics 114(3):391–423

Rabinowicz W, Rønnow-Rasmussen T (2006) Buck-passing and the right kind of reasons. Philosophical Q 56(222):114–120

Rabinowicz W, Rønnow-Rasmussen T (2021) Explaining value: on Orsi and Garcia's explanatory objection to the fitting-attitude analysis. Philos Stud 178(8):2473–2482

Raz J (1999) Engaging reason: on the theory of value and action. Oxford University Press, Oxford

Reisner A (2009) Abandoning the Buck passing analysis of final value. Ethical Theory Moral Pract 12:379–395

Reisner A (2015) Normative conflicts and the structure of normativity. In: Hirose I, Reisner A (eds) Weighing and reasoning: themes from the work of John Broome. Oxford University Press

Rønnow-Rasmussen T (2011) Personal value. Oxford University Press, Oxford

Rowland R (2019) The normative and the evaluative: the buck-passing account of value. Oxford University Press

Scanlon T (1998) What we owe to each other. Belknap Press of Harvard University

Skorupski J (2007) Buck-passing about Goodness. In: Tännsjö T, Rønnow-Rasmussen T, Petersson B, Josefsson J, Egonssson D (eds) Hommage à Wlodek: philosophical papers dedicated to Wlodek Rabinowicz. Department of Philosophy, Lund University

Skorupski J (2010) The domain of reasons. Oxford University Press, Oxford

Strawson P (1992) Analysis and metaphysics: an introduction to Philosophy. Oxford University Press, Oxford

Wiggins D (1987) A sensible subjectivism. Oxford University Press, Oxford

Wygoda Cohen S (2020) Not all partial grounds partly ground: some useful distinctions in the theory of grounding. Philos Phenomenol Res 100(1):75–92

Zimmerman MJ (2001) The nature of intrinsic value. Rowman & Littlefield