Contents

Research Articles

Book Review

RESEARCH ARTICLE

Time Sensitivity and Acceptance of Testimony

Nader Alsamaani*

Abstract: Time sensitivity seems to affect our intuitive evaluation of the reasonable risk of fallibility in testimonies. All things being equal, we tend to be less demanding in accepting time sensitive testimonies as opposed to time insensitive testimonies. This paper considers this intuitive response to testimonies as a strategy of acceptance. It argues that the intuitive strategy, which takes time sensitivity into account, is epistemically superior to two adjacent strategies that do not: the undemanding strategy adopted by non-reductionists and the cautious strategy adopted by reductionists. The paper demonstrates that in adopting the intuitive strategy of acceptance, one is likely to form more true beliefs and fewer false beliefs. Also, in following the intuitive strategy, the listener will be fulfilling his epistemic duties more efficiently.

Keywords: Acceptance of testimony; time sensitivity; reductionism; non-reductionism.

1. Introduction

Testimony occupies a central place in everyone's epistemic sphere. We depend on testimony for a wide variety of beliefs, which might range from

* Assistant professor of philosophy; Qassim University
  ✎ Department of Cultural Studies, College of Sharia, Qassim University, P. O. Box: 6666, 51452 Saudi Arabia
  ✉ nsamaany@qu.edu.sa

directions to the nearby petrol station to medical breakthroughs; from knowing the birthday of a historical figure to knowing one's own birthday.

Epistemologists seem to agree on two main points with regard to testimony: (1) testimony has immense value and is indispensable for our epistemic system; and (2) testimony is fallible since attesters can be unreliable or insincere. Accordingly, the person who is concerned with merely maximizing their knowledge through testimony is likely to be susceptible to deception. On the other hand, if one seeks to avoid recourse to testimony altogether and maintain one's epistemic autonomy, one will have an extremely limited range of knowledge that might perhaps be insufficient for basic survival.[1] Thus, one needs to accept, as most people do, the fact that, when incorporating testimony into one's epistemic system, the epistemic agent makes himself (to a degree) vulnerable to fallibility. This is the price paid for the knowledge that he is able to obtain. As Richard Moran states: '[M]y ultimate destination is the truth about the world, but often I must pass through the beliefs of another person as my only (fallible) access to this truth' (Moran 2006, 278). However, the price (i.e. the risk of fallibility) needs to be *reasonable*, as we do not want to end up with a large number of false beliefs. When we are presented with a testimony, we intuitively evaluate the risk of fallibility associated with it in order to reach a verdict about whether the risk of fallibility is or is not *reasonable*. Accordingly, we can decide whether or not we are willing to accept the testimony.

In this paper, I will first demonstrate how time sensitivity has an impact on our acceptance of testimony by affecting our intuitive evaluation of the *reasonable* risk of fallibility in that testimony. All things being equal, we tend to be less demanding in our acceptance of time sensitive testimonies than we are with time insensitive testimonies. I will develop this intuitive response to testimonies into a strategy of acceptance. I will then argue that what I call the intuitive strategy of acceptance is epistemically superior to

---

[1]   In making a similar point about acceptance in general, Richard Feldman states: 'We can succeed in believing lots of truths by believing everything. […] But that hardly achieves any sort of epistemic excellence. On the other hand, by believing very little we surely manage to avoid error. But this excessive conservativism does not achieve epistemic excellence either. It is by attaining a suitable mix of the two goals that we will achieve epistemic excellence' (Feldman 1988, 244).

two adjacent strategies: the undemanding strategy adopted by non-reductionists, and the cautious strategy adopted by reductionists. I demonstrate that, in adopting the intuitive strategy of acceptance, which takes time sensitivity into account, one is likely to form more true beliefs and fewer false beliefs. Also, in following the intuitive strategy, the listener will be fulfilling his epistemic duties more efficiently.

2. Time sensitivity and testimony

Testimonies vary with regard to their time sensitivity. Some are bound to a tight timeframe: if one does not form a belief about the testimony within the particular timeframe, it will become epistemically valueless. Other testimonies are not time sensitive: one could delay forming the beliefs associated with them without any effect on their epistemic value to the listener. A testimony, though, could be time sensitive with respect to several considerations such as truth, justification/warrant, belief/acceptance, and probably others. In this paper, I am merely concerned with time sensitivity with respect to the truth of the testimony. To understand this point, let us consider the following example. Assume that you are late for a job interview and you get to the company's building and ask a stranger about the location of the interview. He tells you that it is taking place in the conference room and gives you directions to the place. You can either accept (or reject) the stranger's claim right away or withhold acceptance (for further investigation for instance). If you withhold your acceptance for a while, there will probably no longer be a *current* job interview. Thus, you miss the opportunity to have a belief altogether, since even if you decide to form a belief later, your belief will not be about the place of the *current* job interview; rather, it will be about something different—the place of a *past* job interview. The truth of the proposition 'the interview is taking place in the conference room' is bound to a specific timeframe (i.e., the real time of the interview). That is, outside of this timeframe, the proposition would not be true (i.e., there will not be an interview that is taking place *now* in the conference room).

In the above example, it seems that intuitively one is likely to be more lenient and accept the testimony. When we ask strangers about directions

to an office (or a gate), we tend to believe them readily. But does the element of time sensitivity have any impact on our intuitive undemanding acceptance of some testimonies? I believe the answer is affirmative. To test this hypothesis, let us compare two similar examples where the main difference between them is that one is time sensitive and the other is not:

1. A stranger identifies himself as a pediatrician to a mother and child waiting at a bus station and diagnoses the child with a non-urgent illness, prescribing a particular medicine.

2. As her child is choking and about to become unconscious, a stranger comes to the mother and identifies himself as a doctor. He tells the mother that her child needs an emergency tracheotomy and volunteers to perform it.

Let us assume that the stranger in both situations is the same person and that there are no clues that either undermine or strengthen his sincerity and reliability. All the mother has is the stranger's word. It seems that in the first example the mother would be warranted to withhold belief regarding both of the stranger's testimonies—that he is a doctor and that her child is sick and needs a particular medicine. She might bear in mind what the stranger says and decide to ask the family's GP the next time she visits, but it seems unwarranted to form a belief about her child's health based merely on a diagnosis made by a stranger.

On the other hand, in the second scenario, the mother would seem to be warranted in accepting the stranger's testimony that he is a doctor and that the child needs a critical procedure. This would be the expected and, perhaps, actual reaction of many people in a similar situation. It is quite common for people, in an emergency, to believe strangers' claims that they are doctors and to trust their claims about the situation of the ill (on airplanes, for instance).

Accordingly, the mother seems justified in both withholding belief about the first testimony and in accepting the second testimony, even though they are produced by the same person and contain fairly similar propositions. Apparently, what differentiates the two cases is the fact that, while the first testimony is fairly time insensitive, the second is very time sensitive. In the first example, the mother does not have to form a belief immediately (or

perhaps ever). She has the time, if she wishes to accept the testimony, to evaluate the sincerity and reliability of the attester. In the second case, however, the mother does not have the luxury of time to reflect carefully on the testimony and on the character of her interlocutor, as she has to accept (or reject) the testimony right away. Accordingly, even though the risk of fallibility (that the proposition is false) is fairly similar in both testimonies, the mother shows more tolerance towards the second than the first testimony. She seems to evaluate the risk of fallibility of the second (but not the first) testimony as *reasonable.* Her stance appears to be intuitively compelling. Since there was no difference between the risk of fallibility of the two testimonies, what makes the fallibility of the second testimony intuitively *reasonable* is apparently the fact that it is very time sensitive. It seems, therefore, that time sensitivity affects our intuitive evaluation of the risk of fallibility and of what is reasonable when accepting testimonies.

3. The intuitive strategy of acceptance

Our intuitive response to testimonies (as represented by the examples of the mother(s) and the alleged doctor) could be developed into a strategy of acceptance. This strategy assumes that one is warranted in accepting a time sensitive testimony at face value unless there are reasons that might undermine the reliability or sincerity of the attester. Further, it maintains that one is warranted in withholding one's acceptance of time insensitive testimonies unless there are positive reasons for the sincerity and reliability of the attester. I call this the intuitive strategy of acceptance.

There are two other competing strategies of acceptance. The first is the undemanding strategy that advances that our response to testimonies must be one of unchallenging acceptance. It proposes that one is justified in accepting a testimony as long as there is no reason to believe that one's interlocutor is insincere or unreliable. The second is the cautious strategy that advances that our acceptance of testimony must be conditioned. One is justified in accepting a testimony only if one possesses reasons to suggest that one's interlocutor is sincere and reliable. Advocates of the first strategy, which is typically traced back to Reid, include non-reductionists like Coady (1992), Burge (1997), and Weiner (2003), among others. Advocates of the

second strategy, which is typically traced back to Hume, include reductionists like Audi (1997), Lackey (2003), and E. Fricker (1995), among others.

Unlike the intuitive strategy, both of these strategies do not seem to consider time sensitivity a relevant factor in accepting testimonies. Apparently, in both strategies, the reasonable risk of fallibility in testimonies is fixed and remains constant across all instances of testimony. The two strategies differ, however, in their evaluation of the *reasonable* risk of fallibility in testimonies. On the one hand, the undemanding strategy seems to maintain that initially the *reasonable* risk of fallibility in testimony is identical to the risk of fallibility in testimony in general. Hence, it advocates accepting a testimony at face value unless there are positive reasons that might raise the risk of fallibility with a particular testimony (i.e. that one's interlocutor is insincere or unreliable). On the other hand, the cautious strategy seems to maintain that the risk of fallibility in testimony in general is initially higher than the *reasonable* risk of fallibility. Hence, this strategy requires positive reasons in order to reduce the risk of fallibility to the reasonable level so that the testimony might be considered acceptable (i.e. by confirming sincerity and reliability).

Besides being intuitively appealing, I maintain that the intuitive strategy is also epistemically superior to both the undemanding and cautious strategies. I provide two arguments in support of my claim.

First, the intuitive strategy has an advantage over the undemanding and cautious strategies with respect to providing more true beliefs and fewer false beliefs through testimony. To demonstrate, I will test the three strategies on all of the possible instances of testimony in relation to time sensitivity and truth and falsity, which are: (1) A testimony that is true and time sensitive; (2) a testimony that is false and time insensitive; (3) a testimony that is true and time insensitive; (4) a testimony that is false and time sensitive (see Table 1).

For the sake of simplicity, I will assume that there is no immediately available reason that undermines or supports the sincerity and reliability of the interlocutor in each of the four testimonies. However, I will assume that the sincerity and reliability of the interlocutor could be known through further investigation (which also, I assume, takes time). In addition, I will assume that the essential factor in determining the truth or

Testimony	True/False	Time Sensitivity
Testimony 1	True	Time Sensitive
Testimony 2	False	Time Insensitive
Testimony 3	True	Time Insensitive
Testimony 4	False	Time Sensitive

Table 1

falsity of a testimony is the sincerity and reliability of the attester because the probability that the testimony is true is high if the attester is sincere and reliable and is low if he is not. I am aware that it is possible, however, that a testimony could be true even if the attester were neither sincere nor reliable, or false even if he were both sincere and reliable. Since the probabilities of such occurrences are quite low, I will omit them in my analysis below for the sake of simplicity.

Given the above, let us move to test the three strategies. In adopting the intuitive strategy of acceptance in handling the testimonies in table 1, with everything being equal, I will end up with three true beliefs and one false belief. Following this strategy, I will accept testimony number one at face value since it is time sensitive, and from this will follow a true belief. Since it is time insensitive, I will inquire about the sincerity and reliability of the attester when I am presented with testimony number two. This should suggest that the testimony is false and hence will lead to another true belief. The same goes with testimony number three, except here I will realize that the testimony contains a true proposition. Finally, since it is time sensitive, I will accept testimony number four without further investigation, which will result in a false belief. Accordingly, following the intuitive strategy leads me to three true beliefs and one false belief.

In adopting the undemanding strategy, however, I will end up with two true beliefs and two false beliefs. Since there is nothing that would undermine the sincerity and reliability of the attesters, I will accept all four testimonies at face value without further investigation, regardless of their time sensitivity.

Finally, in adopting the cautious strategy, I will end up with two true beliefs only. Since there is no available information that strengthens the

sincerity and reliability of the attester, I will investigate their sincerity and reliability in all four instances. Accordingly, I will form true beliefs about the time insensitive testimonies, but will fail to form beliefs about the time sensitive ones.

In examining the results of following these strategies, one should notice that the intuitive strategy is indeed epistemically superior. While we will gain at best two true beliefs following the other two strategies, we will form three true beliefs following the intuitive strategy. Between the two other strategies, however, the cautious strategy seems to be preferable since it will result in its follower forming two true beliefs and no false beliefs. This can be compared to the undemanding strategy, which adopts an overly tolerant approach that will lead to its follower forming two false beliefs and two true beliefs.

The second reason for adopting the intuitive strategy is that, in following it, we are better fulfilling our epistemic duties than if we followed either the undemanding or the cautious strategies. Fulfilling epistemic duties means that, in forming a belief, the agent exhausts the available means and sources to attain truth and avoid error. When a testimony is time sensitive and there is nothing that undermines the sincerity and reliability of the attester, the word of the attester is usually the only source available to the listener. Further, when a testimony is time insensitive, there are usually other means and sources available to the listener through which he could determine the sincerity and reliability of his interlocutor. Therefore, by depending on the words of the attester—which are the only source available to the listener in a time sensitive testimony—the listener is exhausting all of the available means, and thus he is fulfilling his epistemic duty. On the other hand, by enquiring further about the sincerity and reliability of his attester in time insensitive testimony, the listener, then (and only then), exhausts the available sources and means. Only the intuitive strategy yokes these two modes together.

In following the undemanding strategy, however, the epistemic agent would fail to fulfill his epistemic duties in certain instances. To demonstrate, let us go back to the time insensitive example of the mother and the stranger who claims that her child is sick. In that example, the character of the attester is opaque. The undemanding strategy would encourage the mother

to believe the stranger right away even if there are means available to en-
quire about his sincerity and reliability. This seems problematic; to appre-
ciate this, let us assume that the mother knows that the stranger is a friend
of her cousin Sally, whom she is just about to meet. Yet, she chooses to
believe the stranger's claims at face value without consulting her cousin at
all. Apparently, by doing so the mother would not be fulfilling her epistemic
duties efficiently as this option would make her unnecessarily vulnerable to
error. All things being equal, the mother as an epistemic agent should do
what is better for attaining truth and avoiding error. Since the situation is
not time sensitive, what is better for attaining truth and avoiding error is
clearly the one where she enquires about the sincerity and reliability of the
attester. The vulnerability to fallibility in testimonies is something one can-
not avoid altogether, but at the same time it is something that one should
not go through without a good reason. Therefore, on certain occasions, the
undemanding strategy fails to enable the listener to fulfill his epistemic du-
ties.

On the other hand, the cautious strategy seems to ask too much of the
listener and in adopting it the listener might, in some instances, also fail to
fulfill his epistemic duty to accept a testimony. To understand this point,
let us re-examine the example of the time sensitive testimony of the alleged
doctor and the choking child. According to the cautious strategy, the
mother is not justified in accepting the claims of the stranger until she
investigates his sincerity and reliability. This requirement seems problem-
atic. Since the testimony is time sensitive, by accepting the words of the
stranger, the mother seems to exhaust all of the available means and
sources to justify her acceptance. Anything beyond what is available to
the mother should not be part of her epistemic duty, and in that instance,
investigating the sincerity and reliability of the stranger is beyond her
available sources. If she investigates the sincerity and reliability of the
stranger, the child will probably die in the process and the stranger's
testimony that 'I can save the child' will become void. Indeed, not accept-
ing the testimony while exhausting all the available justifications for it is
failing to fulfill one's epistemic duty to accept a relevant proposition.
Hence, the cautious strategy also fails to provide a means of maintaining
the listener's epistemic duties.

In comparing the three competing strategies, one should notice that the intuitive strategy appears favorable as it helps the listener to fulfill his epistemic duty more efficiently than either the undemanding or the cautious strategies.

4. Potential objections to the intuitive strategy of acceptance

There are two potential objections to the intuitive strategy of acceptance. The first targets the concept of time sensitivity upon which I based the intuitive strategy. The second targets my argument for the intuitive strategy in which I argued that the strategy has an advantage over the undemanding and cautious strategies with respect to providing more true beliefs and fewer false beliefs through testimony.

First, one might object that time sensitivity is a relative concept. There is no clear distinction between a time sensitive testimony and a time insensitive testimony. On the other hand, however, the intuitive strategy seems to assume a clear distinction between the two and it seems to be based upon that distinction. Since the distinction is indeed unclear, the intuitive strategy is useless.

This objection, however, is not compelling. The distinction between time sensitive and time insensitive testimony is only unclear if one considers time sensitivity unconditionally, which is not the case with the intuitive strategy. In fact, the time sensitivity/insensitivity of testimony is considered *with respect to the relevant instance in which the testimony is presented*. With time sensitivity being conditioned in this manner, the distinction between time sensitive and time insensitive testimony becomes clear. Accordingly, the question one should ask is whether or not one can investigate the reliability and sincerity of the attester further without the testimony losing its value *with respect to the instance in which the testimony is presented*. If the answer to that question is affirmative then the testimony is time insensitive, and if the answer is negative then the testimony is time sensitive. To illustrate, let us reexamine the alleged doctor examples introduced above. In the first example,

1. A stranger identifies himself as a pediatrician to a mother and child waiting at a bus station and diagnoses the child with a non-urgent illness, prescribing a particular medicine.

In this example, we notice that the mother has the time to investigate further the reliability and sincerity of the attester. She has time to do so *with respect to the instance itself*. The truth of the testimony should not change while the mother investigates further. The testimony, therefore, is time insensitive.

Things are different with the second example:

2. As her child is choking and about to become unconscious, a stranger comes to the mother and identifies himself as a doctor. He tells the mother that her child needs an emergency tracheotomy and volunteers to perform it.

In this example, we notice that the truth of the testimony would probably change if the mother investigates the reliability and sincerity of the attester further since the child might die, for instance. The mother does not have time to investigate further *with respect to the instance itself*. Since the mother cannot investigate further without rendering the testimony valueless, this testimony, therefore, is time sensitive. Notice that apart from the instance itself, some parts of the testimony can be considered time insensitive, such as the attester's testimony that he is a doctor. This sense of time sensitivity is, however, irrelevant to my proposition. Therefore, the first objection fails. There is indeed a clear distinction between time sensitive and time insensitive testimony if the concepts were conditioned to the instance where the testimony is presented and are considered with respect to the truth of the testimony in particular.

The second objection might be offered to my argument for the intuitive strategy, in which I postulated that the intuitive strategy has an advantage over the undemanding and cautious strategies with respect to providing more true beliefs and fewer false beliefs. One might object that my argument presupposes an equal distribution of true and false testimony across time sensitive and time insensitive testimony while this might not be the case. Namely, it might be that, in reality, there are, for example, far more false time sensitive testimonies than true time sensitive testimonies. Therefore,

in practice, the strategy would not help one in obtaining more true beliefs and fewer false beliefs as it is claimed.

This objection, however, is based on a misunderstanding of the superiority of the intuitive strategy defended above. There are two ways to represent my argument for the intuitive strategy. In this objection, it seems to be represented as follows:

(1)   A testimony comes in four forms: a true time sensitive testimony; a true time insensitive testimony; a false time sensitive testimony; and a false time insensitive testimony.

(2)   All instances of testimony in the real world distribute evenly among the four forms above.

(3)   Following the intuitive strategy will lead one to obtain true beliefs through testimony three out of four times.

(4)   Following the undemanding or the cautious strategies will lead one to obtain true beliefs two out of four times.

Therefore, following the intuitive strategy leads one to have more true beliefs and fewer false beliefs through testimony.

If one understands my argument as such, then it is obvious that premise two is incorrect and, hence, the argument is unsound. This, however, is a misrepresentation of the argument. The argument should read as follows:

(1)   A testimony comes in four forms: a true time sensitive testimony; a true time insensitive testimony; a false time sensitive testimony; and a false time insensitive testimony.

(2)   External factors determine how instances of testimony in the real world distribute among the four forms.

(3)   Isolated from external factors, instances of testimony should distribute evenly among the four forms.

(4)   Isolated from external factors, the intuitive strategy will lead one to obtain true beliefs through testimony three out of four times.

(5)   Isolated from external factors, the undemanding or the cautious strategies will lead one to obtain true beliefs two out of four times.

(6)   Hence, isolated from external factors, the intuitive strategy leads one to have more true beliefs and fewer false beliefs through testimony.

Therefore, the intuitive strategy has an *inherent* advantage over the undemanding and cautious strategies with respect to providing more true beliefs and fewer false beliefs.

However, the question might arise as to what justifies isolating external factors of testimony in premises 3–5 in my argument above.

External factors (i.e., the actual contexts in which testimony is given) vary significantly and perhaps, when taking these into account, there will not be a single, unconditional, advantageous strategy. Different contexts will require different strategies of acceptance. For instance, knowing that stand-up comedians tend to lie in their stories to enhance the humor, we would perhaps be wise to adopt the demanding strategy when accepting their testimony. In addition, knowing that doctors are reliable and tend to tell the truth when discussing the results of tests with patients, the undemanding strategy would appear to be the most advantageous strategy to adopt in such instances. No doubt, this is practically helpful. However, when taking external factors into consideration, we will recognize, if we are successful, which strategy is better *relative* to a specific context. This, however, is irrelevant to my argument. The aim of my argument is to establish that the intuitive strategy is internally, or in theory, relatively advantageous in providing more true beliefs and fewer false beliefs. This advantage cannot be examined unless we isolate testimony from external factors. I am aware, though, that this *internal* advantage might be overpowered or cancelled out by external factors in actual contexts. It is an advantage nonetheless. Hence, the second objection is irrelevant.

Finally, it might be worthwhile noting that the intuitive strategy is concerned principally with achieving more true beliefs and fewer false beliefs. In some cases, however, getting things right might not be essential (as in the case of the stand-up comedian's stories mentioned previously). The intuitive strategy seems to be unhelpful in such cases. Yet, this should not be a disadvantage for the strategy for two reasons. First, the primary aim of any strategy of acceptance (indeed, the primary epistemic aim *simpliciter*) is to obtain true beliefs and avoid false beliefs. This is also the main epistemic duty required by epistemic agents. Other advantages, if they are relevant (like differentiating between significant and less significant testimony), should be secondary to the primary epistemic aim. Secondary

advantages could always be supplemented by other tools. As long as the strategy does not obstruct the secondary advantages, there seems to be nothing objectionable about it.

The second and more important reason is that while judging the truth of testimony is something absolutely objective, determining the significance of the testimony could be widely subjective and relative to a particular listener (or a group of listeners). What is important to one person might be trivial to another. It should not be a disadvantage of any strategy if it fails to account for the relative and various personal preferences of listeners.

5. Conclusion

In this paper, I have defended the intuitive strategy of acceptance which, unlike competing strategies, takes into account the time sensitivity of the testimony. I have argued that the intuitive strategy is epistemically superior to the adjacent strategies of acceptance: the undemanding strategy and the cautious strategy. One is likely to obtain more true beliefs and fewer false beliefs in adopting the intuitive strategy. Additionally, in following the intuitive strategy, one fulfills one's epistemic duties more efficiently than would be the case with either of the two other strategies.

References

Audi, Robert. 1997. "The Place of Testimony in the Fabric of Knowledge and Justification." *American Philosophical Quarterly* 34 (4): 405–22. https://doi.org/10.1093/acprof:oso/9780190221843.001.0001

Burge, Tyler. 1997. "Interlocution, Perception, Memory." *Philosophical Studies* 86 (1): 21–47. https://doi.org/10.1093/acprof:oso/9780199672028.001.0001

Coady, Cecil A.J. 1992. *Testimony: A Philosophical Study.* Oxford: Clarendon Press. https://doi.org/10.1093/0198235518.001.0001

Feldman, Richard. 1988. "Epistemic Obligations." *Philosophical Perspectives* 2: 235–56. https://doi.org/10.2307/2214076

Fricker, Elizabeth. 1995. "Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony." *Mind* 104 (414): 393–411. https://doi.org/10.1093/mind/104.414.393

Lackey, Jennifer. 2003. "A Minimal Expression of Non-Reductionism in the Episte-
mology of Testimony." *Noûs* 37 (4): 706–23. https://doi.org/10.1046/j.1468-
0068.2003.00457.x

Moran, Richard. 2006. "Getting Told and Being Believed." In *The Epistemology of
Testimony*, edited by Jennifer Lackey and Ernest Sosa, 272–306. New York:
Oxford University Press. https://doi.org/10.1093/ac-
prof:oso/9780199276011.003.0013

Weiner, Matthew. 2003. "Accepting Testimony." *Philosophical Quarterly* 53 (211):
256–65. https://doi.org/10.1111/1467-9213.00310

Laudan, Intuition and Normative Naturalism

Howard Sankey*

Abstract: The aim of this paper is to document Laudan's rejection of the appeal to intuition in the context of his development of normative naturalism. At one point in the development of his methodological thinking, Laudan appealed to pre-analytic intuitions, which might be employed to identify episodes in the history of science against which theories of scientific methodology are to be tested. However, Laudan came to reject this appeal to intuitions, and rejected this entire approach to the evaluation of a theory of method. This is an important stage in the development of his normative naturalist meta-methodology.

Keywords: Epistemic normativity; meta-methodology; method; theory-change.

1. Introduction

What is the relationship between intuition and the theory of epistemic normativity? For some, intuition enables us to explore our concept of knowledge or justification (e.g. Goldman 2007). For others, intuition serves only to identify obvious and uncontroversial items of knowledge (e.g. Kornblith 2002, 10–11). For still others, intuition is unable to play an evidential

* University of Melbourne

 🖉 School of Historical and Philosophical Studies, University of Melbourne, Parkville, Victoria, Australia 3010

 ✉ chs@unimelb.edu.au

role because it is influenced by philosophically irrelevant factors (e.g. Weinberg, Nichols and Stich 2001). In this paper, I consider an episode from the recent history of the philosophy of science in which appeal to intuition was rejected precisely in order to develop a theory of epistemic normativity.

The episode relates to the proposal in the 1970's of opposing models of scientific theory-change by Imre Lakatos and Larry Laudan. In the attempt to provide a rational account of the dynamics of theory-change, Lakatos proposed a methodology of scientific research programmes on which scientists adopt progressive while rejecting non-progressive research programmes (Lakatos 1970). To overcome perceived shortcomings with Lakatos's model, Laudan developed an account of scientific research traditions, which he combined with a problem-solving model of scientific rationality. To assess competing theories of the methodology of science, Lakatos and Laudan both saw a need to identify earlier episodes in the history of science against which the theories might be tested. For Lakatos, it was the "value judgements" of the scientific élite (e.g. Lakatos 1978, 124), while, for Laudan, it was the "pre-analytic intuitions" of "scientifically educated people," that are employed to identify the episodes.[1]

Serious questions arise in relation to the evidential role that may be played by such value-judgements and pre-analytic intuitions. Laudan came to recognize that the appeal to intuition is confronted with severe problems. Given this, he rejected the appeal to intuitions, and developed instead his normative naturalist meta-methodology.[2] Thus, Laudan's development of a naturalistic approach to the normative appraisal of methodological criteria formed part of his rejection of the appeal to intuition. In this paper, my aim is to document Laudan's rejection of intuition in the context of his proposal of a naturalistic theory of epistemic normativity. Though largely historical in substance, the paper has the systematic intent of suggesting that appeal to intuition may be avoided in the theory of epistemic normativity.

---

[1]   For discussion of Lakatos's appeal to value judgements, see (Sankey 2018).

[2]   For detailed analysis of Laudan's normative naturalist meta-methodology, see Nola and Sankey (2007, especially section 12.2).

2. Laudan and pre-analytic intuitions

In 1977, Laudan published *Progress and its Problems*, in which he pro-
posed a model of scientific theory-change in the attempt to improve upon
Lakatos's model of theory-change.[3] Like Kuhn and Lakatos, Laudan
thought that scientific development rests on underlying theoretical struc-
tures which inform research in an area of science for extended periods of
time. Laudan called these theoretical structures "research traditions." Like
Lakatos, Laudan took there to be 'core' elements within a research tradi-
tion. But, unlike Lakatos, he allowed that the core elements of a research
tradition may be modified or replaced over time (1977, 99). Laudan con-
joined his model of theory-change with a novel problem-solving account of
rationality.[4] This involves the idea that science is fundamentally a problem-
solving activity. As such, "the aim of science is to maximize the scope of
solved empirical problems while minimizing the scope of anomalous and
conceptual problems" (1977, 66).[5] Given this characterization of the aim of
science, what it is to be rational in science is to act in a way that increases
the problem-solving effectiveness of a research tradition (1977, 124–5).

---

[3]   By starting my discussion of Laudan with *Progress and its Problems,* I pass over
his collection of historical essays on theories of method, *Science and Hypothesis.* The
reason is that I wish specifically to focus on the transition away from the intuitionist
meta-methodology that Laudan shared with Lakatos at the time of writing *Progress
and its Problems.* For parallels and contrasts between the intuitionism of Laudan
and Lakatos, see Laudan (1986, especially 124–6).

[4]   Though others (e.g., Popper and Kuhn) thought of science as a problem-solving
activity, Laudan developed this insight into an explicit theory of rationality.

[5]   In spelling out the problem-solving conception of rationality, Laudan develops
a taxonomy of problems (1977, chapters 1 and 2). Empirical problems are substan-
tive questions that arise with respect to the objects in a domain of scientific study.
Unsolved empirical problems are not solved by any research tradition. A solved
problem is solved by at least one research tradition. It thereby becomes an anomaly
for a competing research tradition so long as it is unsolved by the latter tradition.
By contrast with empirical problems there are conceptual problems, either internal
ones (e.g. inconsistency, ambiguity or circularity) that arise within a tradition, or
external ones which arise due to a conflict between a tradition and another theory
or tradition, a methodological view or even a non-scientific world-view.

Though a scientist may pursue a range of theories or traditions, it is rational for a scientist to accept the research tradition which displays the highest degree of success in solving problems (1977, 109).[6]

Lakatos spoke of value judgements rather than intuitions. By contrast, Laudan does explicitly employ the term 'intuition,' though he sometimes speaks of judgements and convictions as well. He proposes that a model of scientific rationality is to be tested against key episodes from the historical development of the sciences.[7] A number of cases may be specified from the history of science about whose rationality or irrationality we have clear intuitions. The intuitions which relate to the resulting list of cases give rise to a set of "preferred pre-analytic intuitions about scientific rationality" (1977, 160). This set of pre-analytic intuitions may serve as touchstone in the evaluation of a theory of method or rationality. It is a necessary condition of adequacy for a theory of method or rationality that it fit with the set of pre-analytic intuitions. In a particularist spirit akin to Lakatos, the intuitions relate to specific episodes in the history of science.[8] Like Lakatos, Laudan holds that we have clearer intuitive reactions to particular cases than with respect to abstract theories of method or rationality. Unlike Lakatos, Laudan takes the intuitions to relate to a small set of cases rather than, potentially, the whole history of science. In a further departure from Lakatos, Laudan makes no appeal to the judgements of élite scientists. Instead, he speaks of "scientifically educated persons" (1977, 160).

---

[6]   Laudan usefully distinguishes pursuit from acceptance (1977, 108–10). In the context of pursuit, a scientist might explore a promising theory or research tradition without being fully committed to it. Acceptance involves a stronger degree of commitment, e.g. taking a theory to be true.

[7]   At this stage, Laudan often speaks of a theory of rationality rather than a theory of method. However, methodological considerations play a role in his problem-solving conception of rationality. For example, an external conceptual problem may arise for a tradition if it comes into conflict with an accepted principle of scientific methodology (1977, 57–61). As we shall shortly see, Laudan later takes the view that the theory of rationality is to be sharply distinguished from the theory of method.

[8]   For analysis of particularist elements of Lakatos's approach to meta-methodology, see my (2018), which brings Lakatos's approach into contact with the epistemological particularism of Roderick Chisholm (1973).

Laudan's account of the role that intuition may play in the evaluation of a theory of method was subject to significant criticism.[9] As a result, Laudan renounced his intuitionism, ultimately going on to develop a naturalistic meta-methodology instead. In 1986, Laudan published a response to a critical paper by Daniel Garber (1986), in which Garber questioned the role of the history of science as opposed to that of our own intuitions in the evaluation of theories of method. In his response, Laudan replied in concessive spirit to Garber's objections while at the same time repudiating the intuitionist approach he had previously adopted. Laudan raised a number of concerns about the appeal to intuitions (1986, 123 ff.). For one thing, he points out that if a theory of method is grounded in a set of pre-analytic intuitions, the capacity for the theory of method to serve as basis for criticism of those grounding intuitions is severely limited. It would not be possible to reject the intuitions on the basis of the theory of method, since the sole basis for adoption of the theory of method in the first place is its conformity with those very intuitions. For another thing, Laudan notes that intuitions are not always universally shared with respect to methodological questions. Given lack of unanimity with respect to intuition, an appeal to specific cases can hardly be expected to resolve disagreement with respect to methodological matters. Furthermore, even if there were to be agreement in intuition, it is entirely possible that competing theories of method may fit with all the same historical cases picked out by the shared intuitions. Having presented these and other reasons for rejecting the intuitionist approach, Laudan concludes by indicating that an alternative approach lies "ready to hand" (1986, 126). He does nothing at that point to characterize this alternative, though it seems likely that he was thinking of the normative naturalist meta-methodology that he went on to develop in subsequent years.[10]

---

[9]   For example, Janet Kourany raises questions about the relevance of the intuitions of "scientifically educated persons": such intuitions may fail to reflect the rationality of actual science, differ from notions of rationality found at earlier periods in the history of science and, given their origin in a person's science education, potentially constitute evidence that is lacking in neutrality (Kourany 1982, 535–6).

[10]   In fact, Laudan refers to a "monograph-length treatment" entitled *Science and Method* on which he was working. So far as I am aware, no such monograph did materialize. But papers on normative naturalism start to appear the following year.

3. Laudan's normative naturalism

The year after the response to Garber, Laudan's major articulation of the normative naturalist position was published (1987).[11] By contrast with *Progress and its Problems*, Laudan now distinguishes sharply between a theory of rationality and a theory of method. This is primarily due to the fact that the rationality of an action depends on an agent's aims and background beliefs, and methodological rules may be employed by scientists in an attempt to attain their cognitive aims. It would be inappropriate, therefore, to judge the rationality of scientists of an earlier epoch by attempting to determine whether they employed the methods which we currently adopt to pursue our aims, given that the earlier scientists might have adopted neither our methods nor our aims. Equally, a scientist of an earlier period might have held substantively very different beliefs from ours. Even if they did share our aims, they might have had different beliefs about how to achieve those aims, whether or not they shared our methodological views.

A sharp distinction between questions of rationality and method brings out the fatal flaw in the intuitionist approach to the appraisal of theory of method. The attempt to evaluate a theory of method by determining whether it counts the actions of an earlier scientist as rational is quite wrong-headed. As Laudan notes:

> Because our aims and background beliefs differ from those of past scientists, determinations of the rationality of their actions and of the soundness of our methodological proposals cannot be collapsed into one and the same process. Rationality is one thing: methodological soundness is quite another. (1987, 23)

---

[11] I will not attempt to bring Laudan's 1984 book, *Science and Values*, into this discussion of the development of his ideas. That would detract from the focus on the development of his meta-methodological views. Suffice to say that the reticulated model that he presents in *Science and Values* is primarily designed to provide an account of the rational evaluation of variable cognitive aims. There is a closer connection between the reticulated model and normative naturalism than is immediately apparent. For the naturalistic approach to the evaluation of methodological rules may be readily integrated into the reticulated model.

The result of enforcing a sharp distinction between rationality and method is that the appraisal of a theory of method comes apart from questions of the rationality of past scientists. The problem, now, is how to determine the soundness of a theory of method. This is where Laudan's turn to naturalism comes in.

The key to Laudan's naturalistic approach is the suggestion that the rules of method may be construed as hypothetical imperatives. Specifically, Laudan proposes that a rule of method has the form, "If one's goal is y, then one ought to do x" (1987, 24), where the goal is a cognitive or scientific goal, and what one ought to do is to employ some proposed method or procedure. This construal of the logical form of a rule of scientific method has the decided advantage of making empirical considerations relevant to the appraisal of methods. For a rule of method now rests on a substantive empirical claim to the effect that the employment of a specific method will lead to the realization of a specific desired cognitive or scientific end. Such a claim may be true or false, depending on how the world in fact is. Moreover, it is in principle possible to obtain empirical evidence for the truth or falsity of the empirical claim embedded in the rule of method. The upshot is that it is possible to provide empirical evidence of the extent to which a rule of method is an effective means of attaining a desired cognitive end.

On the intuitionist approach, a theory of method is to be evaluated in terms of whether it reveals selected episodes in the history of science as rational. Laudan now rejects both the appeal to intuition and the role of rational reconstruction in the appraisal of method. Nevertheless, the history of science continues to play a crucial role in determining the soundness of a rule of method. For, rather than appeal to intuition or the rationality of past scientists, the appraisal of a methodological rule now turns on the empirical question of whether use of the rule conduces to its purported aim. This is an empirical matter which turns on historical matters of fact. Investigation of the history of past science may reveal whether or not utilization of a specific rule of scientific method has in fact led to the realization of the aim to which it was thought to lead. Thus, even while rejecting the intuitionist appeal to past science, Laudan's normative naturalist meta-methodology accords history of science a crucial role in the appraisal of rules of method.

Laudan's approach to the empirical appraisal of rules of method is strongly naturalistic precisely in virtue of the way it treats the appraisal of the rules of method as an empirical matter. But such a naturalistic approach does not render epistemological questions a matter of descriptive psychology in the manner at one point seemingly suggested by Quine.[12] Rather, Laudan's naturalism is a *normative* naturalism on which the rules of method have normative force. They convey normative force because their employment does in fact conduce to desired cognitive ends. A scientist whose belief or theory-choice conforms with such rules thereby possesses epistemic warrant with respect to the belief or theory-choice precisely in virtue of their conforming with rules that lead to desired cognitive ends. Thus, Laudan's normative naturalist meta-methodology constitutes a strongly naturalistic epistemological theory about the basis of the epistemically normative force of the rules of scientific method.

4. Conclusion

In this short paper I have sought to document how critical reflection on the role of intuition contributed to the turn to normative naturalism in the work of Larry Laudan. As we have seen, Laudan was critical of the role that might be played by intuition in relation to the theory of scientific method and the rationality of science. At one level, this is a historical point in relation to the development of Laudan's methodological thought. At another level, the lessons of this episode seem to me to have significant negative implications for the appeal to intuition in the context of a theory of epistemic normativity.

Acknowledgements

---

[12]   Quine (1969, 82–3) is sometimes read as a purely descriptive, anti-normative position, though other passages suggest an alternative more strongly normative interpretation (e.g. 1992, 19–20).

References

Chisholm, Roderick. 1973. *The Problem of the Criterion: The Aquinas Lecture 1973*. Milwaukee: Marquette University Press.

Garber, Daniel. 1986. "Learning from the Past." *Synthese* 67 (1): 91–114. https://doi.org/10.1007/BF00485512

Goldman, Alvin. 2007. "Philosophical Intuitions: Their Target, Their Source and Their Epistemic Status." *Grazer Philosophische Studien* 74 (1): 1–26. https://doi.org/10.1163/9789401204651_002

Kornblith, Hilary. 2002. *Knowledge and its Place in Nature*. Oxford: Oxford University Press. https://doi.org/10.1093/0199246319.001.0001

Kourany, Janet. 1982. "Toward an Empirically Adequate Theory of Science." *Philosophy of Science*. 49 (4): 526–48. https://doi.org/10.1086/289076

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos, and Alan E. Musgrave, 91–196. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139171434.009

Lakatos, Imre. 1978. "History of Science and its Rational Reconstructions." In *The Methodology of Scientific Research Programmes: Philosophical Papers Volume I*, edited by John Worrall, and Greg Currie, 102-37. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621123.004

Laudan, Larry. 1977. *Progress and its Problems*. London: Routledge & Kegan Paul.

Laudan, Larry. 1981. *Science and Hypothesis*. Dordrecht: Reidel. https://doi.org/10.1007/978-94-015-7288-0

Laudan, Larry. 1984. *Science and Values*. Berkeley: University of California Press.

Laudan, Larry. 1986. "Some Problems Facing Intuitionist Meta-Methodologies." *Synthese* 67 (1): 115–29. https://doi.org/10.1007/BF00485513

Laudan, Larry. 1987. "Progress or Rationality? Prospects for Normative Naturalism." *American Philosophical Quarterly* 24: 19–31. https://www.jstor.org/stable/20014171

Nola, Robert, and Howard Sankey. 2007. *Theories of Scientific Method: An Introduction*. Stocksfield: Acumen. https://doi.org/10.4324/9781315711959

Quine, W.V.O. 1969. "Epistemology Naturalized." In *Ontological Relativity and Other Essays*, 69-90. New York: Columbia University Press.

Quine, W.V.O. 1992. *Pursuit of Truth*. Cambridge, Mass.: Harvard University Press.

Sankey, Howard. 2018. "Lakatosian Particularism." *Logos & Episteme* IX (1): 48–59. https://doi.org/10.5840/logos-episteme2018913

Weinberg, Jonathan M., Shaun Nichols and Stephen Stich. 2001. "Normativity and Epistemic Intuitions." *Philosophical Topics* 29 (1 & 2): 429–60. https://doi.org/10.5840/philtopics2001291/217

RESEARCH ARTICLE

If Sounds Were Dispositions:
A Framework Proposal for an Undeveloped Theory

Jorge Luis Méndez-Martínez*

Abstract: In the realm of the philosophy of sounds and auditory experience there is an ongoing discussion concerned with the nature of sounds. One of the contestant views within this ontology of sound is that of the Property View, which holds that sounds are properties of the sounding objects. A way of developing this view is through the idea of dispositionalism, namely, by sustaining the theory according to which sounds are dispositional properties (Pasnau 1999; Kulvicki 2008; Roberts 2017). That portrayal, however, is not sufficient, as it has not inquired the metaphysical debates about dispositions beyond the conditional analysis. In this paper, I try to advance this view by including recent developments (for instance Bird 2007; Vetter 2015) in the field of dispositionalism and I analyse whether this new version can sort out known and new objections to Property View.

Keywords: Audibility; dispositions; ontology of sound; potentiality; property view; sounds.

* National Research University – "Higher School of Economics"
  ✎ International Laboratory of Logics and Linguistics, School of Philosophy, National Research University – "Higher School of Economics", Staraya Basmannaya st. 21/4, 105066 Moscow, Russian Federation
  ✉ jmendez@hse.ru; slavinskii@gmail.com

0. Introduction

Imagine that you are sitting next to a river, having a picnic with friends. It is a sunny but slightly chilly day, so you are wearing gloves. You take your beer bottles to make a toast and, as the bottles collide, you are surprised to realise that your bottle seems to be made of plastic instead of glass. The gloves had prevented you from realising that it was a plastic bottle and, because in shape and colour it imitates glass beer bottles, only the lack of the characteristic sound of clinking glasses gives away the material it is made of. When you finally take a sip, the feeling on your lips confirms what the sound had hinted: it is indeed a plastic bottle.

This little scene relates to different issues regarding the philosophy of sounds: the relation of sounds and sources (Nudds 2010; Casati, Di Bona, Dokic 2013; O'Callaghan 2007b; Fowler 2013), the problem of perceptual justification (Handel 2006), the issues of sense multimodality (O'Callaghan 2011) and, last but not least, the problem of the nature of sounds (O'Callaghan 2007; O'Shaughnessy 1957; Pasnau 1999; Roberts 2017). The latter is the one I am concerned with in this paper. Is it possible to say that plastic *has* a different sound than that of glass? I think it is intuitive to answer 'yes.' In order to justify this intuition, I will appeal to a view in the ontology of sounds that should explain this in a satisfactory sense, that is the dispositional view or sound dispositionalism.

Typically, by dispositions we mean things such as fragility, solubility, irascibility and the like. Dispositions are properties that, under certain circumstances, could manifest themselves. The suffix 'ity' is quite indicative of those cases. This paper is about those sorts of properties and it examines the possibility of claiming that "sound" is a sort of disposition. It can be said that an object has "the disposition to sound" or, seemingly, "sonority" if so and so occur. It will be a matter of dispute what the nature of this "so and so"—namely, the conditions—is about.

The debate is circumscribed in the frame of Sound Ontology (SO), which, in this sense, has split into three views: the wave view (WV), where sounds are considered as acoustic waves (WV) as physics and, more precisely, acoustics tend to say; the property view (PV), where sounds are either properties of the perceiving mind (PV1), as psychology presumably

argues, or properties of the object (PV2); and, finally, the event view (EV), where sounds are considered as events (O'Callaghan 2007a, 2009, among others, favours this stance). Seemingly, the dispositional view would belong to PV2, or so it has been interpreted by those who have taken this route.

I must clarify that my purpose is not to decisively advocate for a dispositional view on sounds. I am not trying to convince you to accept the thesis that sounds are dispositional properties of objects. My commitment in this paper is related to the consequences and implications of such dispositional account.

This is not the most popular view in SO. Actually, the so-called *property view*, mostly labelled as such when it is criticised, rarely unfolds in a way that takes the global philosophical discussion on properties into account: are they universals? Are they tropes? There is only a handful of allusions in this sense: P. F. Strawson (1959), on the one hand, and Edmund Husserl (1984), on the other, have made some type-token considerations of sound as universal (for instance the *C* note) and as a particular (the playing of the note C for instance), but that does not take sound *as if* it were a property *of* an individual.[1] This, I argue, is due to a problem of *under-specification,* common to all the views that figure in the SO.[2] The field is thus in need of further development.

This applies for the dispositional account of sounds, which is a species within PV, and more precisely PV2. Three authors in particular, Pasnau (1999), Kulvicki (2008), and Roberts (2017), have advanced this view. Yet, sound's characterisation as a disposition *has not* been elaborated close to the spirit of the long metaphysical debate on dispositions. Only Roberts (2017, 347) mentions in passing the problems pertaining the conditional analysis, let alone newer considerations on the problems of modality and

---

[1]   As known, the fact that sounds do not coincide with the idea of basic particular in Strawson's metaphysics, namely that of the material body, is what motivates the examination of such problematic 'individuals.' The revival of sounds as a matter of philosophical discussion most likely comes from Strawson.

[2]   A notorious exception in that of Jonathan Cohen (2010, 205) who considers universal and/or trope form of abstract properties, while tackling linguistic objections against PV.

potencies (Vetter 2011a, 2011b, 2013a, 2015). Thus, a first goal of the paper is to achieve a more complete picture of the dispositional view.

The first section will provide a picture of Sound Ontology, hopefully short but compelling. The second section will present the current state of affairs of the dispositional view on sounds. A third section introduces the classical considerations from the dispositions' debate—its features under the conditional analysis and the newer potentiality argument—to the discussion on sound. In a fourth section, I want to learn if a new version of sound dispositionalism can withstand the criticisms that some authors (O'Callaghan 2011; Casati and Dokic 2014) have raised against the dispositional account and to evaluate its prospects for the future within the realm of SO.

In this article my contention is to bring two debates together: that of the philosophy of sounds and auditory experience and that of dispositions. A brief literature review, especially in the first and third sections, is necessary to make intelligible the conceptual exchange between both discussions. This characterisation, however, is not free of difficulties and I do not take for granted that sounds are dispositions. Even more, since I am not sure sounds or sonority are dispositions, I have opted for the title "If sounds were dispositions…" and by doing so I also want to make more explicit the character of this inquiry as hypothetical.

1. The ontology of sound

The expression "Ontology of sounds" refers to the effort to define what sounds are. There are, to my knowledge, two general proceedings in that quest: the purely ontological taxonomy of theories, which is elaborated by O'Callaghan (2007a, and with Nudds 2009), and the topological approach (the label is mine), which inquiries *where* sounds are (Casati and Dokic 1994, 2005). Both taxonomies overlap in some regards, but not in others.

The first ontological choice to make is that of deciding whether sounds are properties or individuals. For some (Roberts 2017), this is what divides the ontologies of sound. There are, however, more diversified approaches. In O'Callaghan's taxonomy, for instance, this amounts to three possibilities: the wave view (WV), the property view (PV), and the event view (EV).

The Wave View (WV) claims, from Aristotle[3] to modern acoustics, that sounds are acoustic waves. Despite its popularity, there are a number of issues that WV cannot account for. The most important one is that it pre-supposes an Error Theory of perception:[4] we, for one, do not perceive sounds to be at any acoustic wave as a vibration in an elastic medium. We perceive them as *coming from* a source (a sounding object, an event, etcetera).

This gives rise to what we can consider as *the phenomenological desideratum*, which any prospect of SO should comply with. At least, O'Callaghan (2007a, 14) notices, it is better to choose a theory that explains the phenomenological aspect of sound, to one that does not. Authors differ on the degree of importance they assign to it, some take it as a necessary constrain, some as an inescapable requirement, and some consider it a discussion they could bypass.[5]

Another option is that of PV, which is twofold: it either describes sounds as properties of the perceiving mind (PV1) or it describes sounds as properties of the sounding objects (PV2). The most common one is that of PV2, ever since John Locke described them in his *Essay* as "secondary qualities."[6] Typically, secondary qualities are thought of as being qualities of the objects somehow enabled, detected and identified by our perception. However, it is far from decisive that secondary qualities in general (let alone sounds) are

---

[3]   Aristotle is usually considered in such fashion in many historical accounts (Pasnau 1999, 310; Casati and Dokic 2014).

[4]   In neuroscience we can find a similar problem while dealing with vision and the so-called "inverse problem" in a Berkeleyan fashion. The information in the retina does not correspond directly to the real structure of the world, so how is it possible that we respond successfully on the basis of vision? (Purves 2010)

[5]   Such assessment is visible while pointing out to the 'Error Theory' as a major inconsistency to be resolved or that undermines the whole effort for searching a co-herent view on the nature of sound, that is the case with Pasnau (1999), Casati and Dokic (2014).

[6]   In the seventh chapter of the second book ('On Ideas') from the *Essay Concerning Human Understanding*, while addressing the simple ideas, Locke formulates the notion of "primary and secondary qualities." Primary qualities are those that are inseparable of a body, for instance, solidity or extension; whereas, secondary qualities are *"nothing in the objects themselves but power to produce various sensations in us"* (Locke 1999, 117). Such is the case of sounds.

exclusively objects' properties as there is an ongoing discussion on whether they could also, or instead, being regarded as properties of the perceiving mind (see, for instance, Egan 2006). In such scenario, we would face PV1.

PV1 would be focused on sound as a purely an auditory phenomenon and, therewith, it would permit auditory hallucinations to be considered as sounds, since sounds' privacy would be plausible (this is an anathema for O'Callaghan 2007a). Not only that, but there would be as many sounds as hearers (Casati and Dokic 2014). Sound would lie, thus, in the ear of the beholder. No philosopher seems to be explicitly endorsing this stance, but there are interpretations that may implicitly argue in favour of it: O'Callaghan (2007), Casati and Dokic (2014) point to D. L. C. Robert Maclachlan's *Philosophy of Perception* (1989, New York, Prentice Hall); Casati and Dokic (2009, 103), additionally, mention O'Shaughnessy (1957). There is even mention of Edmund Husserl and Franz Brentano, usually overlooked in the analytical debates on sounds, as potentially approaching this view (Méndez-Martínez 2020).

In the same guise, O'Callaghan (2009, 34) appeals here to the argument from the vacuum, for which there are two ways of arguing this: either one says that a medium is a necessary condition for sound (a bell struck in the vacuum does not make any sound); or you say that the audible properties of sound (namely, timbre, pitch and loudness) cannot be afforded in the absence of a medium, and do not produce a veridical perception. PV2, as I elaborate below, is at odds with these arguments.

Finally, we have the event view (EV), defended by O'Callaghan (2007a, 2009)[7] and somehow by Casati and Dokic (2014). O'Callaghan defends what the latters label as "Located Event Theory," whereas the option defended by Casati and Dokic is that of the "Relational Event Theory."

To say that "sounds = events" is to claim that they are spatiotemporal localisable occurrences that are to be identified neither with properties nor with waves. EV tries to deliver in the case of the phenomenological desideratum. Another key aspect that will mark a difference with any potential

---

[7]   Despite being considered as the most representative figure in EV, O'Callaghan has employed other labels for more current elaborations: the mereological position, for instance, which considers the causal sources of sound as a "part" of sound (O'Callaghan 2011).

PV2, dispositional view included, is that EV relies in the *manifestation* of sound: whenever there is a sound, there is a sounding (O'Callaghan 2009, 36). EV pays a great deal of attention to the problem of relating to sources and, in that sense, it is probably the view that encompasses the most in its effort to explain sound. In reporting so, things complicate to the extent that some have preferred to cast aside causality, sources and just addressing "pure events" (Scruton 2009). A problem this theory has, among many that are beyond the scope of this paper, is that it must provide several explanations to specify the location of sounds; echoes and Doppler effect are recurring puzzling cases (Casati and Dokic 2014; O'Callaghan 2007b; Fowler 2013).

In addition, as with the other views, EV is not free from the problem of under-specification. Although maybe not as properties, events represent another large topic in ontology and metaphysics to which EV relates to at a lesser extent.

This is the purely ontological classification. Precisely because of the problem of location and spatiality, Casati and Dokic (1994, 2014) consider that deciding *where* sounds are warrants as much explaining as the matter of *what* they are. For them, there are three broad conceptions of sound: *proximal*, which locates sound *at* or *in* the hearer; *medial*, which locates sound *between* the hearer and the sounding event/object; and *distal*, which locates sound at the sounding object or event.[8]

In short, these are the views that compose SO, which, in my opinion, offer a general description of the discussion. There are other particular classifications depending on what is being inquired. Scruton (2009) thinks of physicalist and non-physicalist conceptions, where common rival views such as that of Pasnau and O'Callaghan fit, for him, in the physicalist row. Dokic (2007), who reviews the tensions between the "unique event" and the "repeatable object" ontologies, is another example.

Let us now evaluate the proposal in question: the dispositional view.

---

[8]   There are some correspondences between both taxonomies. WV, for instance, is clearly distal; whereas EV is presumably distal, although it sometimes seems to make some concessions that put it close to WV (Casati and Dokic 2014). PV1 is proximal, whereas PV2 is distal.

2. Sound dispositionalism, a state of art

A preliminary outline is useful. On these views, sounds are properties and, curiously enough, most proponents of a PV point out to dispositional properties. However, in the large debate concerning properties there are other available options on how to draw on properties, yet philosophers of sound have chosen this one.[9] I will analyse in the conclusions whether this compromises, or not, this view.

This dispositional view is a sort of PV and, seemingly, it is PV2, because the argument points out the *disposition-to-sound* in the object, rather than in or at the hearer. If the dispositional locates what is going on at the object, it is certainly not medial nor proximal, but rather distal. If it locates sounds at the hearer, then it is proximal.

These implications will be clearer further on. For now, let us elaborate on the state of affairs of the dispositional view that, though unpopular in the overall philosophical discussion on sound, is not without representatives (Pasnau 1999; Kulvicki 2008; Roberts 2017; and, at a certain extent, Cohen 2010).

Pasnau's paper is probably one of the most influential sources in the field of the philosophy of sounds, which is probably due to its contribution to the revival of the debate. In it, Pasnau proposes:

> ...identifying sound with the vibrations of the object that has the sound. More cautiously, I would say that sounds either are the vibrations of such objects, or supervene on those vibrations. The former would imply a physicalist account of sound, whereas the latter would have room for a dispositional account. (Pasnau 1999, 317)

However, the part concerning the supervenience on vibrations does not do the job in showing strong adherence to a dispositional model. An issue

---

[9]   A case of an adherent of this view who is not championing a dispositionalist alternative is Jonathan Cohen (2010). He elaborates on the idea that the characteristic temporal feature of sound is usually taken to be at odds with its characterisation as a property. Although his defence does not mention dispositions, his treatment could serve dispositional arguments.

that here seems innocuous is that of sounds "supervening." Usually, the expression used is that of "being instantiated by" while describing exactly the same (for instance O'Callaghan 2007a, 66). It is yet to be inquired, though, on whether they are equivalent and exchangeable sentences,[10] for the appeal to supervenience has been contested for some of those who are disposition realists, and a dispositional account of Anti-Humean inspiration would spare us of modelling through the idea of supervenience, as I will elaborate in the next section.

Although Pasnau does not go deeper in this, he formulates one of the key ideas for the dispositional account while saying that objects *have* sound (Pasnau 1999, 316). This goes against the grain concerning the ordinary language use of sound as something that is "made" and not "had" (Pasnau 1999, 310). Not only in this having/making sound distinction has Pasnau paved the way down for sound dispositionalism but also regarding the appeal to colour. In so doing, he has a very different attitude towards colour analogy than that of O'Callaghan's (2007a), which goes against the 'tyranny' of the visual.[11]

Besides the distinction and the appeal to colour analogy, Pasnau does not add more on how this dispositional account could be. In the end, the

---

[10]   I fear they are not, because although at a certain point we can use them without reserves, while being specific 'supervenience' commits us to a metaphysical picture (the Lewis-Humean one) which could have, or not, unwarranted features on how the world is structured.

[11]   If considered as an endeavour within the philosophy of senses or the philosophy of perception, the philosophy of sounds and auditory experiences shares a relatively common ground with that of colour and vision: there are ontological discussions on the nature of both colour and sounds; there are several positions concerning the phenomenological content of our auditory and visual experiences, and so forth. However, when compared to the forays in the realm of colour and vision, the philosophy of sounds and auditory experience is considerably less developed. Not only that, but it has browsed, at large, many of the discussions that have taken place in the philosophy of colour and vision. Trying to explore, independently, new paths for the auditory phenomena is a task that O'Callaghan undertakes, rebelling against the 'tyranny of the visual.' Of course, not all the philosophers in this new field would agree with such assessment and particularly those defending dispositional views (Pasnau, Kulvicki, Roberts), for they appeal to an analogy with colour.

goal of his paper is rather the criticism towards the "standard view." By using such label, Pasnau refers to both WV and PV—arguably PV1—, whose conflation is rather incoherent because it allows and even endows an 'Error Theory' of perception.

John Kulvicki has a more detailed argument to offer. In his view, sounds are "stable properties" of the objects. Concerning the other philosophers' taxonomies and classifications, in O'Callaghan's taxonomy, this would clearly classify as PV. The same can be said of Casati's and Dokic's classification. When Kulvicki states that: "*sounds are perceived to have locations, and those locations seem to correspond to the objects that make the sounds*" (Kulvicki 2008, 1), he chooses a "distal" view. But, beyond these brandings, his account is also 'dispositional,' as he adds that: "*Perhaps sounds are not vibrations per se but dispositions to vibrate in response to certain kinds of stimulation*" (Kulvicki 2008, 4).

The core of the argument, for Pasnau and Kulvicki, lies on colour analogy. Sounds are usually thought of as "transient" as opposed to colours that are "stable properties" of the objects that possess them (Dokic [2007] thinks precisely this while theorising on the opposition unique-repeatable). The key claim is the following:

> As objects still have their colours in the dark, they also "have" sounds even in the vacuum or without vibrating, those sounds cannot be heard.

Kulvicki (2008, 5) considers that objects have "resonant modes," dependent on their material structure—remember the plastic beer bottles—that cause them to vibrate. Objects are *disposed to* vibrate when "thwacked." Here Kulvicki is getting closer to the standard conception of dispositions through the lens of conditional analysis. The 'thwack,' a way of imparting energy to the object (Kulvicki 2008, 9), is the stimulus condition; the sound made—seemingly—, its manifestation. But having a sound is not the same as making it. It is easy to think of objects that make sounds that they don't have: a speaker, for instance. Appealing to colour analogy, one might say that a projection in the movie theatre is showing ("making") coloured images, when the colour, indeed, *isn't there*. Notice how, by switching to colour analogy, we employ different verbs.

This possession of sound, so to speak, represents, for Kulvicki, a stable property that licences the possibility of dispositional talk, although he does not draw beyond on the literature. All this would lead us to identify the expression "had sound" with dispositions, and that of "made sound" with the manifestation, which, formulated as such, is a rather circular argument. However, there are alternatives specifications.

A more elaborated and/or explicit depiction on the dispositional account for sounds is that of Roberts's (2017), who relies at large in the property view for colour. Having divided the views between property and non-property, Roberts groups eleven possible views within PV, according to their potential compliance with three features: being dispositional or not dispositional (thus, categorical); being reductive or non-reductive; being relational or non-relational. Something not clearly explicit here is whether not having some of the features mentioned by Roberts means that they *necessarily* lack them.

A first interesting aspect is that, unlike Pasnau and Kulvicki, he does consider a non-dispositional possibility for PV. Here 'non-dispositional' means categorical, for which one of the options is vibrationism (that is, appealing to vibrational structure), which is also a reductive view. As for the dispositional ones, they have revolved around the possibility of wave dispositionalism (that is, the overt disposition to produce an acoustic wave) and vibration dispositionalism (like Kulvicki's thwack).

When we consider perceived sound (or sound appearance) things get, as usual, thornier: a conditional analysis like-spirit is still present, for here "*sounds are dispositions to auditorily appear in certain ways if certain conditions are met*" (Roberts 2017, 346). Here we have an internal sound dispositionalism and an external one. The difference is that the external one does distinguish between sound and apparent sound, thus enabling us to explain differences of change and constancy. However, apparent sounds seem difficult to conceive if not taken as experiences of sounds. Finally, on the possibility of connecting the dispositional and relational features, Roberts notices that they do not have to come necessarily together. This becomes an important issue to discuss the differences between sound, in a phenomenal sense, and the objectual sound.[12]

---

[12]   Talking about 'internal' and 'external' could lead us to hasty associations. For instance, we could say that the former matches, at a certain extent, with PV1, while

So far, in Roberts' treatment the dispositional account is more complete. He concocts these varieties in a hypothetical fashion, as they are not being argued for currently. Kulvicki would happen to be a vibration disposition-alist, viewed from this lens.[13] Another issue to bear in mind is that he de-picts sound dispositionalism in the fashion of conditional analysis and, even more, he is aware that this way of dealing with dispositions might face theoretical obsolesce. Yet he does not appeal to alternative views on dispo-sitions. In the following sections, I highlight the so far implicit elements of sound dispositionalism *à la façon* de Lewis and I also undertake its shaping into a more current discussion.

3. A dispositional view, from conditional analysis to potentiality

A word at the broader level of the discussion on dispositions is required. Dispositions, as I have said, are a genre of properties. The other genre is that of categorical properties. The main difference lies in the fact that dis-positional properties are instantiated *under certain conditions*, while the categorical ones are instantiated in *all conditions*. It is common to think, as well, that the latter are constant, while the formers are not; that the latter are observable, while the others are not (if not being manifested at the

---

the latter with PV2. But that is not exactly the case, because the focus of the features sorted by Roberts is not aimed at bestowing the property either to objects or subjects, but on how they both relate to sound appearance. The same caveat goes for trying to associate this to the externalist/internalist discussion within the phi-losophy of mind and language. I do not mean that these aspects should remain separated from the features outlined here, it is just that so far the discussion is rather narrow.

[13]   The whole spectrum proposed by Roberts encompasses the following options: non-relational primitivism, vibrationism, disjunctive vibrationism, disjunctive wave/vibration dispositionalism, primitive appearance relationism, reductive appearance relationism, vibration relationism, appreareance sound dispositionalism, wave dispo-sitionalism, vibration dispositionalism, as well as an unavailable view which would happen to be dispositional, non-relational.

moment).[14] This division, as fundamental as it is, roughly mentioned by Roberts' taxonomy, is usually cast aside in the debate on sounds. This, as seen in the conclusions, could be of major importance for anyone who favours PV's varieties.

In the big picture of metaphysics, Humeans and Anti-Humeans wage a battle concerning their understanding in the structure of the world. There are many arenas for this (for instance causality, modality, and observability/non-observability), being the nature of properties one of them. More concretely, the label "Anti-Humean" is employed while alluding to David Lewis's metaphysics, which strongly appeals to Humean supervenience, whose formulation goes as follows: "All there is in the world is a vast mosaic of local matters of particular fact, just one little thing over then another" and this has an impact in how laws are conceived: "The laws of the world supervene on the totality of local matters of particular fact" (Lewis 1986b: ix). If there is an accepted understanding on both sides it is that of Humean supervenience as the core of Humeanism and Lewis as its endorser.

Typically, Humeans defend categorical properties and Anti-Humeans, dispositional ones. But this is an oversimplification as the matter is, indeed, which of the two are the most fundamental. In this sense, two monist positions are on sight: either every sparse property (namely, natural properties) is categorical or it is dispositional. A middle ground could recognize that there are properties on each side. One can also accept dispositions (or 'disposition talk' *á la* Ryle), without committing ultimately to this. The opposite view would be dispositionalist realism.

This corresponds to a deeper review of our view on metaphysics and the philosophy of sounds and auditory experience could perhaps dispend with

---

[14]   We should not understand this in a visualist form. Observables entities are those that we can perceive with any of our senses, in normal conditions; unobservables don't. Yet some unobservables are detectable by using certain instruments (subatomic particles, for instance) and some are not detectable but have rather an explanatory role (Chakravaty 2007, 14–15). There is, of course, a critical aspect in this, as being "non-observable" used to be an anathema for the empiricist *Weltanschauung*, although being non-observable does not constitute a reason to rule out dispositions anymore. As, for the strictly scientific point of view, there are many entities, like *quarks*, whose status in this sense is non-observable.

this. However, the difference could have some relevance while understanding the available dispositionalist projects and some falsifiability criteria.

Back to business, both Kulvicki and Roberts have drawn on a profile of the long debate on dispositions that depends on the conditional analysis. However, the standard view on dispositions has been discussed (Armstrong 1996; Martin 1996), criticised (Bird 1998), reviewed (Lewis 1997) and dismissed by some authors (Bird 2007; Vetter 2015). This warrants a concise revisit to the guidelines of the conditional analysis. Let us take a quick view on it. Following no other than Lewis (1997b), the *conditional analysis* states that:

> CA: Something x is disposed at time t to give response r to stimulus s iff, if x were to undergo stimulus s at time t, x would give response r.

In short, this is usually capsuled by the formula D $_{(S, M)}$, that is, the ordered pair of stimulus and manifestation (or response).[15] We can see how it fits the problem of sound.

Now, the disposition to sound would be given by the pair formed by a stimulus (eg. Kulvicki's 'thwack') and, on the other hand, a manifestation (either if it is the vibration or the wave, that is, vibration dispositionalism or wave dispositionalism in Roberts's terms). This looks simple and compelling. The conditional analysis picture can be, however, more specific. For instance, it can distinguish between covert dispositional property names or nouns (such as 'fragility', 'solubility', etcetera); covert dispositional predicates in adjective form (such as fragile, soluble); overt (canonical) dispositions descriptions 'the disposition to M when S'; and overt dispositional predicates 'x *is disposed* to M, when S' (Bird 2007, 18). Given that sound

---

[15]   Although this line of argument is immediately associated to Lewis, we had to go back a bit earlier to authors like Carnap and Ryle. Carnap (1936, 448), who introduces essentially the same definition but with a more elaborated formula, is concerned that disposition-terms don't enable semantic reduction—a 'dogma' later criticised by Quine (1961). Ryle (1949, 31), on the other hand, identifies dispositions' ascriptions as those that allude to a particular change when an object is under certain conditions. These semantics are known as the "simple conditional analysis" (Choi and Fara 2016). All this is to say that conditional analysis is not entirely Lewis's doing. His attention was dedicated to its reformation, rather than the simplistic formula.

does not commonly figure as a disposition, it is perhaps artificial to say that we can easily appeal to "sonority" or "sounding,"[16] whereas the former is, in the ordinary language, describing a conduction quality of an object (something you could say of a musical instrument or a performer of a musical instrument for example), the latter has been used to describe the manifestation. It seems that sound ought to be overtly formulated as "the disposition the sound" or "x is disposed to sound iff... certain conditions are met." The conditions could be, in a first moment, the object having resonant modes or a vibrational structure and a medium of propagation.

A specification that is worth bearing in mind is that pertaining single and multi-track dispositions. A disposition D can have multiple manifestations: a fragile glass can break, but it also can just get cracked and so forth; or, as Ryle's (1949, 107) example goes, the disposition of knowing French can be manifested in being able to speak it, to listen to it, to write in French and the like.[17] As it turns out, we have myriads of multi-track dispositions and one could say that thinking of dispositions otherwise would be mistaken.[18] Let us portrait this feature for sound. Sounds' stimuli can be very diverse, which is already noticed by Kulvicki (2008, 9). In this sense, to give a dispositional analysis account for sounds while regarding them as dispositions had to be exhaustive. This can make us doubt on whether the definition of the stimuli of "when thwacked" is correct, for if you smash, make explode and so on, you will also have a sounding object.

Manifestation can diversify as well. Let us think of a musical instrument like the violin. A violin player knows that she can obtain different 'colours' depending on how she pulls the bow. If she bows near the bridge then we

---

[16]   Audibility is also used in the debate, but I will stress the difference below.

[17]   While Ryle is the first one to brig up this idea, we owe its systematic treatment to Bird (2007, 23), who elaborated on all the combinations of stimuli-manifestation. His work on how such characterisations can, or not, account for fundamental and pure dispositions, leads him to proposes and a typology where we have simple stimulus, conjunctive manifestation; disjunctive stimulus, simple manifestation; simple stimulus, but conjunctive manifestations; conjunctive stimulus, and a simple manifestation.

[18]   For a comprehensive argument on how and why most if not all dispositions are multi-track, see (Vetter 2013b).

obtain a very specific and a bit harsh effect, a technique known as *sul ponticello*; if she bows near the fingerboard, we have a rather sweet effect, known as *sul tasto*. If she plucks the strings instead of using the bow, we have *pizzicato*, and so on. Then the manifestations can be diverse and although in these cases it seems that they connect one-to-one with the diversity of stimuli, that is not necessarily the case. We can point to different examples. If you 'thwack' someone on the nose, maybe you would hear a sneeze, maybe not. Musical instruments are useful examples for all this, but they can be misleading in one particular aspect: unlike other objects, they are *supposed to sound*, sounding is what they are meant to do. When we turn to 'ecological sounds,' that disposition (taking for granted, for the sake of the argument, that they are such) can go along other with other dispositions, let alone multiple manifestations. As I will elaborate in the conclusions this leads to the problem of parasitical dispositions, which might represent a possible objection to sound dispositionalism.

This is not where the story ends. As it is known, there are plenty of counterexamples for the conditional analysis. Such is the case of masks—i.e. entities that prevent the manifestation even if there is a disposition D and a stimulus S—and mimics—that is, when we have manifestations and stimulus but without the disposition D. The typical counterexamples for CA in the case of sound can come in two guises depending on how the manifestation is understood: in a vibrational sense or in the phenomenological sense, that is, as the possibility of having, or not, the experience of sound.

A clear-cut case of a mask is that preventing the vibration in the medium of the disposition-to-sound when actually having the stimulus. A silencer or a pillow in the case of a firearm is a mask in a vibrational sense. Masks, as seen, may vary in their masking success. The other direction for masks is the phenomenological one. Maybe there is vibration, but there is some physical obstruction in one's ear tragus that prevents you from having the experience-of-sound. Or if this is still too objectual, we could refer to some reactive response within the auditory system. For PV1 this is the mask and not the former; whereas for PV2 the emphasis would be in the former.

With mimics it happens in a similar way. The manifestation is produced either in spite of the absence of the disposition in the object to give away

a characteristic sound; or in absence of the subject's disposition to have an experience of sound. This would happen via the intervention of a mimicker, like the famous case of 'The Hater of Styrofoam' (Lewis 1997): a Styrofoam dish would break when struck by intervention of the mimicker (i.e. The Hater of Styrofoam), even though the Styrofoam dish lacks that disposition.

We may envision the example of a guitar, which is not supposed to sound like a cello with pizzicato when plucking a string, yet via the intervention of some mimicker, let us call it 'the Cello-*pizzicato* lover,' it would sound like a cello with *pizzicato*. Can we apply the difference between making and having sounds like in the case of the speakers? I think so. However, 'orphan manifestations' eventually raise some concerns (see §4). Some other counterexamples, like that of finks, are considerably more difficult to create for the case of sound than the already mentioned, and so far this is enough for contesting the conditional analysis. Conditional analysis, in order to respond to each of these cases, can reconfigure over and over with *ceteris paribus* clauses and with endless specifications that, in the long run, show that going through dispositions takes a lot to produce the correct statements, if any. This line of response is identified by Manley and Wasserman (2008, 63) as the strategy of 'getting specific.'

Parsimony and simplicity are the desiderata that critics of conditional analysis have endorsed. A glass is still fragile *even if* nobody attempts to break it; an apple is edible even if no one is willing to eat it, and so on. In this fashion, Kulvicki already formulated that an object has a sound even if not sounding (or even if it is in a vacuum), and there is no need of going through a conditional analysis to state the disposition as it is.

For some reason, however, sound dispositionalism has not gone beyond standard conditional analyses to other current views. In order to advance this view, I deem necessary to explore other options. One of those options is that of *potentialities*, which is extensively explored by Barbara Vetter (2015).[19]

---

[19]   There are other works by other authors (Bird 2007; Molnar 2003) and Vetter herself that go in this direction, yet it is the one mentioned which fully develops the project.

Joining in the attack on conditional analysis seems to represent a first obliged step. Noting, however, that objects still have their dispositions regardless of the manifestation—i.e. a glass remains fragile even if no one breaks it—enables further considerations. The first one concerns us directly. In most cases we use the suffix 'ity' points to dispositions. And it concerns us for we still don't know which is the candidate for sound in such fashion. The second consideration is that some objects having a disposition seems context-sensitive. Since there can be cases of things breaking that are not or could hardly be considered as 'fragile,' as well as people who get angry without being irascible, Vetter appeals, in a first moment, to the notion of 'easy possibility' (Vetter 2015, 72). Yet it is unclear whether this commits us to possible worlds—for instance, that we have to consider worlds where something is likely to break and to distinguish it from those worlds where it is not—and what the *relata* are (for saying that something is easy is likely to say that it is easier than something else).

In this spirit, Vetter switches from dispositions to potentialities: "*I propose that we call those properties which form the metaphysical background for dispositions ascriptions potentialities*" (Vetter 2015, 84). Let us go back to the case of a rock. Intuitively, we would not say that a rock is fragile. Making that ascription entails the definition of a context, in a world where maybe there is a different force of gravity and things of the sort. The ascription of fragility, a typical disposition, is, as said before, context-sensitive. Yet a rock *can* break. Both a glass and a rock—one of which you can ascribe fragility but not the other—*can break*. Therefore, we can appeal to a *potentiality*, which, if you like, can be expressed with the noun 'breakability' (which is not commonly used, if used at all). The difference with fragility is that fragility refers to a point of the spectrum where an object can break—you can fix the context using modal-talk, or by alluding to 'easy possibility'; whereas breakability covers the whole spectrum from a piece of diamond to a thin glass. What we commonly refer as the disposition is the maximal degree in the spectrum, which attaches the suffix 'ity' enabling linguistic intuitions. Having a high degree of potentialities also explains manifestations. And, likewise, the rough picture indicates that having a sufficient or high degree of potentialities goes against having counteracting potentialities—which halt or prevent manifestations (Vetter 2015, 99).

Concerning modal aspects, a full discussion is beyond the scope of this article, but there is a token of relevance that is worth mentioning. Potentiality, as portrayed by Vetter, is still a sort of modality, but it is *localised* in the object in contrast with metaphysical possibility, which is *non-localised* (Vetter 2015, 203). Hence, rather than formulations such as "it is possible that x", we can appeal to "x can M."[20]

It is clear by now that Vetter offers an account of potentiality that tries to be independent of other *explanans* that come about when the manifestation of dispositions/potentialities is exerted. Many of the things we come to think of as definitional of potentiality are interacting factors of the exertion of the potentiality (its manifestation or production), and yet not essential. So there is no need to appeal to the whole conditional analysis conceptual package in order to interpret potentialities, as it happens with change and causation. Of course, change and causation may play a role in how the potentiality is exerted, but they are not definitional.[21] If anything, we can explain causality *via* dispositions (Chakravarty 2007; Vetter 2015, 99). The

---

[20]   Vetter (2013a) claims that it is possible to express potentiality regardless of possible worlds. The reasons on why linking them was part of a canon is yet not as clear. And that happens as well with causality, Humeanism, conditional analysis and supervenience, which seem to be somehow entangled. This entanglement appears also when discussing modality. Lewis is well known for being, among other things, a modal realist (Lewis 1986a). In his quest for arraying his picture of possible worlds and to look for alternatives to modal operators, he first proposed a counterpart theory (Lewis 1968) and then found that supervenience is a convenient tool ('the right one') for his modelling throughout possible worlds (Lewis 1986a). In contrast, Vetter (2011b) identifies Kit Fine's efforts and hers on the side of 'new actualism,' which happens to be confronted with that of Lewis. Both sides could agree, however, on their search for new tools to reflect on modality (Lewis *idem*, Vetter 2013a). At a point, she uses modal talk identifying 'possible worlds' as heuristic (Vetter 2015). For her, new actualism and 'anti-Humean' metaphysics are on the same page (Vetter 2011b, 745). I am not entirely sure these associations apply at large concerning Humeanism/modal realism and actualism/anti-Humeanism, since the basic idea of supervenience doesn't need *ab initio* modalities (either with typical modal operators or with Lewis' counterpart theory).

[21]   For a detailed account on how and why we should avoid causal construal in dispositional explanations without going through modalities the way Vetter does and by taking into considerations epistemological aspects see Gurova (2017).

first rebuke of a certain idea of foundational causality is that of rejecting conditional analysis whenever the stimulus is a cause, and the manifestation, an effect (Vetter 2015, 96). This does not mean that potentialities are dissociated from causal schemes. But not even in all cases manifestations have causes attached. Vetter mentions the possibility of spontaneous reactions.[22] So it happens with change: a potentiality may remain as it is and, even if that is rare or uninteresting, they do exist.

Going back to sound dispositionalism, we can think of sound as a potentiality. Typically, the acoustic spectrum covers human perception thresholds according to wavelength frequencies. Vibrations below of 20 Hz, like those of earthquakes, are considered infrasonic; whereas those above that threshold, like a dog whistle, are known as ultrasounds. It is interesting that the disposition to sound does not consider a maximal degree of frequency, but rather to be located at a specific range within the spectrum. This might lead us to reconsider the idea of associating a typical disposition with the 'maximal degree' in Vetter's account. Since this perspective also frees us from the chains of linguistic intuitions and it is context independent, we could cast aside for a minute the concern of not finding a suitable noun with the suffix 'ity'. We could brand it as *acoustic potentiality*.

So far, I have said very little on how this relates to the perceiving subject or how and why we can consider or rule out other 'ity'-nouns for naming this acoustic potentiality. A last feasible element that can be addressed with the new conceptual device at hand is that of *loudness* and, presumably, *loudability*. That would be certainly a gradable potentiality. One could wonder however whether we should substitute acoustic potentiality for this. I think that should not be the case, since loudability refers to an audible quality of sound. This could give place to a known objection as the 'property of properties,' which I will address in the next paragraph. For now, it suffices to say I do not take this as a lethal objection against sound dispositionalism.

Potentiality's basic elements have been introduced (i.e. gradability, context-independence, localised modality) and how we can think of sounds with

---

[22]   Vetter (2015, 98) mentions that for a dualist philosopher getting angry could be a spontaneous affair, rather than the outcome of a sufficient cause.

them. Now some auxiliary elements may also help us to have a more complete picture. I would add only two: the difference between intrinsic and extrinsic and the idea of joint potentialities. An intrinsic property is contained in the object and it does not depend on any outer conditions. Presumably shape can be such a property; while properties such as being "taller than you" are relational and, therefore, extrinsic since they depend on external objects, conditions, and so on.[23] As for joint potentiality, its formulation goes in hand with that of extrinsic and intrinsic potentialities. An orchestra, for instance, shows the plurality of potentialities of each musician for playing a particular instrument. Another example is that of a key, which has the potentiality of opening a door. This is a useful token, because it points to both joint potentialities—the potentiality of a key to open a door, and the potentiality of a door to be opened—and to the difference between *intrinsic* (the potentiality of opening a particular door) and *extrinsic* (the potentiality of opening a general type of door, regardless of that type actually existing) (Vetter 2015, 124).

In order to implement some of the novel elements mentioned here, we have to go back to one of our initial hindrances: that we do not have an intuitive candidate attaching the suffix 'ity' for depicting sound as a disposition. In an objectual sense we could point out to 'sonority'. Although 'sonority,' in certain contexts, could allude to something else than the "disposition/potentiality to sound," when, for example, we say that:

"The sonority of Tchaikovsky Hall is better than Carnegie Hall"

Which is the same—imprecise and folk—usage we have for "acoustics." In this case we are actually referring to the disposition connected with reverberation of a space. Now, in spite of already having a more compelling idea of an *acoustic potentiality*, there is another candidate that comes around, which has, though, a slightly different character. Such is the case of 'audibility,' which is repeatedly used in the literature on the philosophy of sound's literature (for instance in O'Callaghan 2011, 400).

---

[23]   Dispositionalists discuss, however, whether extrinsic dispositions are adequate, for the canon used to be that they were all intrinsic (for instance Molnar 2003). Here I will just assume we can include them. For reviewing a comprehensive plea for extrinsic dispositions, see (McKitrick 2003).

So far, the discussion has remained in a distal sense, that is, it portrays the property as objectual and that has the advantage of bracketing perception, which is a difficult issue in the philosophy of sounds. Albeit difficult, it is nonetheless decisive and that is probably the reason why sound has been traditionally considered a secondary quality. Framing audibility in this sense seems not only pertinent, but the very way to address dispositions and potentialities in the philosophy of sound. Audibility and sonority are extensionally the same, the difference lies in the fact that the former is addressed in a merely objectual sense, but that is a rather narrow picture that does not accurately describes sound's nature as a secondary quality. I can address this topic only now that I have elaborated on the adequate conceptual elements to depict its dynamics. One of the mentioned elements suits perfectly the occasion: joint potentiality (Vetter 2015, 2019b). In this case we have more than one actor in the circuit: the sounding object, the hearer, which is given also in a spectrum of frequency concerning the audibility/inaudibility of a vibration. Audibility jointly acts, thus, with at least other two potentialities: the one bestowed on the hearer, which is close to what audiologists describe as "hearing capacity" when, for example, describing the use of audiograms (Parker and Parker 2004, 76), and the vibrational potentiality of the object. Having a sound is, thus, just a fragment of the whole picture of a jointly dispositional acting scheme. From this new perspective we do not have to think of the *medium* as a condition in terms of the conditional analysis, but rather as an entity that provides for a joint potentiality.

Besides, Vetter's approach offers yet another element to understand dispositions; a remaining problem in the philosophy of sound and auditory experience is that of sound individuation, a problem 'infamously difficult to resolve' (O'Callaghan 2007, 64). From the point of view of potentiality, a disposition is individuated by its manifestation, full stop (Vetter 2014, 752). As we have dispensed from dealing with change, causation and modality, being sound a disposition/potentiality, the problem of its individuation is no longer a great concern. This will prove even more relevant while reviewing the objections raised against PV and sound dispositionalism. Taking stock, even if this picture is, as I believe, more complete than the one offered by the authors who have pursued sound dispositionalism, it still faces challenges that I deem difficult to sort out, as I elaborate below.

4. Objections against the dispositional wiew,
known and new

Objections to the dispositional view so far have been directed mainly against Pasnau and Kulvicki and, curiously enough, they do not consider the structure of the conditional analysis of dispositions nor do they focus their criticisms on the internal aspects of dispositions themselves. The goal of the critique is usually oriented towards showing that the ontological choice they make is more compelling, so they rather pick it on problematic aspects of the theory. Thus, although there are particular objections, we still lack a robust systematic critique.

Casati and Dokic (2014) notice several objections of this nature. The first one is that our ordinary language does not recognize those uses of sounds (Pasnau [1999, 310] also stresses this). However, since the first lines of this paper I show how there is an ordinary and intuitive way of expressing that something "has a sound" and that, as a matter of fact, points out the main issue: their dispositional nature (iff, of course, sounds happen to be dispositions). A common philosophical strategy is to appeal to the intuitive nature of ordinary language. However, the reasons why philosophical concepts should accommodate to ordinary language, and to what extent, are a matter of debate. A reason on why this remains at issue may have to do with the problem of individuation and the assertion according to which "sounds are particulars".

O'Callaghan (2011) has also raised some important objections. A coincidence he has with Casati and Dokic is that of pointing to the "many properties problem" (O'Callaghan 2011, 378). If sounds were properties, the audible properties we usually attach to sound (and mostly to musical sound), namely, timbre, pitch, and loudness or intensity would be dependent or second degree. In any case, O'Callaghan anticipates a complex property-like response.[24] However, in this scenario and appealing to parsimony, he concludes that EV fares better. In fact, any view that holds that sounds are particulars would fare likewise, so it is not a virtue unique to EV.

---

[24]   Theorising about hierarchies while dealing with properties and their inter-array is not uncommon to metaphysics. An interesting view on this dealing directly with dispositions and powers is that of Molnar (2003).

Another problematic issue is that of change. Typically, in metaphysics properties are not taken to be subject to change, the way individuals are. With O'Callaghan (2011, 379) it is unclear whether he concurs to this, since, after emphasising that sounds are concrete particulars—not object-like, but rather event-like—he claims that although both sounds and objects survive change, they differ in the fact that sounds need of time for unfolding and happening. Moreover, individuation of the formers does depend in the patterns of change in their audible features.

Concerning change, Pasnau (1999, 319–20) does not grant that sounds are subject to change. Much of the evidence we have of change in sound, he says, depends on our perception. Something similar happens with the perception of size: we perceive things to be bigger or smaller in relation to our proximity to them, the same happens with our perception and 'measurement' of intensity, for which, in his opinion, "*sound itself remains unchanged.*"

Not all supporters of PV think accordingly, Roberts (2017, 341), for instance, gives an example on how a change of colour could be possible pointing out to the intensity of colours in a strange colour world.[25] The same, presumably, could be exemplified for sound: the marker for intensity would be loudness. Actually, it is easier to imagine qualitative change in sound than in colour. While playing string instruments, a change in pitch

---

[25]   It goes as follows: "Imagine a world in which all objects were (mostly) transparent but when caused to vibrate suddenly became colored and would exemplify one color after another for some time before again becoming transparent. The exact colors objects would exemplify would depend on the type of object and how much it vibrated. For example, dense objects would exemplify different determinable colors on average than less dense objects, and objects would exemplify different determinate colors of the density-determined determinable color dependent on how much they were vibrating. If a very dense object were caused to slightly vibrate it would become light red, then a lighter red, then lighter, then white, and then the white would fade and the object would again be transparent, and if a less dense object were caused to vibrate lightly it would become a light blue, then a lighter blue, then lighter, and then white before fading back to being transparent, and if the object were caused to vibrate more heavily it would become a dark blue, and then a lighter blue, and lighter, then white, and again the whiteness would fade to transparency" (Roberts 2017, 341).

can occur when using the technique known as *glissando*, or while simply tuning the instrument; a flute can produce in a note a considerable increase in intensity.[26]

By linking confronted views on how sounds survive, or not, change with compliance to ordinary language, Cohen (2010) notes that both views—that is, survivalist and non-survivalist—can be both acceptable, for which we should not use this as a definitive criterion.

The last and main objection is that which actually could undermine any sort of PV: the claim according to which sounds are particulars. This comes in two fashions: the first one is appealing to countability; the second one is by invoking experiencing particulars. The bedrock of these arguments goes like this: particulars have features that arguably properties do not. So, if some of these traits can be encountered while addressing sounds, it is unlikely that they can be taken as properties. Now, there are two considerations here: either it is impossible for properties to have these features, because they are exclusives for particulars; or it is only that they are present in particulars, but not necessarily in an exclusive sense.

Countability is also connected with this particular-like sort of objection: you can count particulars, and you can count sounds, *therefore*, sounds are particulars. Arguably, you cannot do the same with properties and, *therefore*, sounds are not properties. Here is a disanalogy with colour: you cannot count 'colours,' which is taken to be a property; but you can count sounds. Therefore, sounds are not like colours and they are not properties. Taking nonetheless the challenge, one could say that what is countable in that sense is the manifestation, which shows the way in which any disposition is individuated (Vetter 2015, 35, 108). On the other hand, sound could even be used as a mass-term as well (Cohen 2010; Méndez-Martínez 2019). Finally, in Roberts' 'strange colour world' the manifestation of a colour can be also the number of times you switch the light on and off.

This gives an answer to the question raised by Casati and Dokic: "What are the *particulars* that you hear when you hear something?" So, the property theorist could simply answer: property manifestations.

---

[26]   Changes in timbre seem more difficult to use as an example, but theoretically we should not rule them out.

So far so good, these known objections and the responses to them are all very well, even if some rely on desiderata (parsimony, concordance with ordinary language) whose relevance could be queried. A general and systematic objection that could counter PV and sound dispositionalism at large is still pendant. I would like to posit here what, in my view, should be its main guidelines.

Despite philosophy's different standards and historical development patterns to those of science, a criterion that comes in handy while examining an argument is that of falsifiability. In a Popperian spirit we could ask how sound dispositionalism could be proved false. There are several goals that can be assessed and, thus, falsified. The broader one is that which states "sounds are properties" (T1). A way to falsify this out is to deny that claim:

AT1: Sounds are not properties

EV, WV and claiming that sounds are particulars seem to entail AT1. However, just stating that they are particulars does not constitute an argument in and of itself, it requires burden of proof. That is a difficulty we encounter with O'Callaghan's (2011) arguments against PV, for he heavily relies on the characterisation of sounds as particulars (not object-like, but event-like), implying that those features could not belong to properties. However, as shown above, those features can be predicated of disposition's manifestations, as already said. Continuing with the—hypothetical—global systematic objection, if sounds were properties, then we would have to decide between PV1 and PV2. Asserting one discredits, seemingly, the other, and, thus, within PV, PV1 could be the antithesis of PV2 and *vice versa*. It is important to notice that this would be so only if we have overcome the challenge posit by AT1. So far, I have argued that, if they are properties, they are objectual properties, which, naturally, are enabled by our perceptive capacities.

If the argument develops through the path of objectual properties (namely PV2), then there are also contesting views. A fundamental one is, as mentioned earlier, that of choosing between categorical and dispositional. In this sense, either one:

- Chooses categoralist monism in order to advance this view or to discredit dispositions

–   Chooses dispositionalist monism in order to advance this view or to discredit categorical properties

–   Chooses a mild pluralist or dualist position that admits both; yet focus on one of the two.

Naturally, if one were to choose a monist position it would have to be followed to its last consequences, which could undermine or even compromise the project. Categoricalism usually tries to undermine not only dispositionalist monism, but dispositions themselves and, thus, to challenge dispositional realism (that is, the stance for which dispositions do exist). Turning a blind eye to monism, a pluralist *pax metaphysica*, could engage with their preferred type of property without troubling dispositionalist talk or categoricalist talk, depending on one's focus. I believe that in the philosophy of sound and auditory experience we face such a situation and, hence the goal is far from asserting a dispositional view in detriment of a categoricalist one. And that is fine. While there is nothing intrinsically wrong with this, it would be better to keep in mind the falsifiability routes for dispositionalism.

Finally, and narrowing the scale to those views that are not only PV, but dispositionalist themselves, the local theory choice seems to boil down to two options: conditional analysis or potentiality. Here the choice is not so simple, as conditional analysis is committed to more than it states in the analysis itself within the Humean project. However, there may be theoretical scenarios where both ways of framing dispositions are admissible. For instance, one could accept potentiality, and say that only dispositions (that is the maximal degree of a potentiality) can be framed with some reformed conditional analysis, but not the whole spectrum. To my knowledge, there are no claims in this direction, but maybe there is room for pluralism of this kind.

The advanced picture I offer here is not bulletproof and we could conceive extra hindrances to its theoretical development. Let us consider new objections. Even though Kulvicki underestimates the case of sounds made and not had as a *rara avis*, there are more cases of 'orphan sounds.' Consider the case of a thunder. EV has no problem in cataloguing these both as sounding sources and as events—and not even Kulvicki denies that

hearing informs us of the surrounding events. Many specifications are needed to frame clouds as objects. On the other hand, objectless or individ-ualess properties seem to make no sense (Armstrong 1993, 433); although Galen Strawson (2011, 304) thinks differently. Now, it may be the case that thunders and all what is involved in the natural phenomenon of the water cycle is still a particular yet not an object-like one, but event-like, as O'Cal-laghan says. Granted. Still we could wonder what happens with manifesta-tions, like orphan sounds, that are neither relatable to objects nor to dispo-sitions. Therefore, orphan sounds just make a larger problem evident: that of lonely manifestations and mere happenings.

This leads us to the point where we can find most of the objections to dispositionalist theories, including those of sound dispositionalism: the way they relate to stimulus and manifestations and, thus, to its multi-track nature. Although conditional analysis has generally been abandoned in the discussion on dispositions and potentialities, the other aspects in the circuit that are linked to dispositions can still raise concerns. Here I would like to propose the idea of 'parasitical dispositions.' Something already considered in the formulation of multi-track dispositions is the possibility for a disposition to have multiple manifestations, and multiple stimuli as well. That is not new. In this sense, we can use the typical case of a glass' fragility. When breaking, however, the glass also sounds. There is a characteristic sound of a breaking glass that is familiar to most of us. However, either both dispositions go *in tandem*, or one is dependent on the other. This is what I have in mind with parasitic dispositions: the possibility that the disposition to sound *could be* the disposition of some-thing else.

The notion of parasitical disposition is in need of a larger framework not only within philosophy of sounds, but also in the overall literature on pow-ers and dispositions. Being a parasite is being a parasite *of something*. Thus, the idea implies a hierarchy that can be fixed bearing in mind an ecological array of how organisms perceive stimulus: sounds are relevant for animals that are either hunter or pray because they indicate something else than sounds *per se*. EV would refer right away to the causal relationship between sounds and its sources, but it is clear that from this side we have to look for an alternative. The combination of joint potentialities and Gibsonian

affordances (Vetter 2018),[27] for instance, may enable this approach. A parasitical disposition is one that has not affordances and yet it is instantiated with dispositions-potentialities that do have them. It is important to see that a parasitical disposition would not be a counteracting potentiality, which in the criticism of conditional analysis is presented as a 'mask.' We must conceive unperceived properties that do nothing. The fact that sounds don't happen to be in this category is an evolutionary making. There could be possible worlds, however, where sounds would not play such a role.[28]

Finally, an extra objection that follows the line of argument of 'things particulars have but properties don't' is that of parthood. Let's call this the *mereological*[29] *objection*. Although we certainly would not say that:

"Here is a sound. There you have half of that sound."

You can appeal to lengths and durations in the following sense:

"This sound lasted half the length of the previous sound."

Perhaps, this is something a sound engineer or a composer could say. Clearly, sound is not the sort of discrete entity you can halve, like an apple or a table, and there are two non-equivalent solutions to this: either you say that this is something that happens with events in general, which are particulars in the end—as O'Callaghan would probably argue; or as the property theorists (dispositionalist included) would do, one could talk about the theoretical mereology of properties' manifestations (as in Robert's strange world of colour). The topic, however, remains unexplored.

Up to this point, it is evident that the discussion on dispositions can contribute to the metaphysics and ontology of sound. It is in the best

---

[27] In Gibson's ecological psychology, an affordance is a habitat feature that *enables* an organism to do something: "a surface that is knee-high and sufficiently steady affords sitting on." Barbara Vetter (2018) has worked on bringing together the philosophy of potentiality and this approach.

[28] This resembles a Lewinsian idea: that of *idlers*, namely, properties that just manifest and 'do nothing' (Lewis 2008, 75). Perhaps our world is full of these things and it is just a matter of evolutionary serendipity that sounds happen to be relevant for so many species.

[29] Unlike O'Callaghan (2011), I do refer here to mereology in its classical sense: as the study of the relationship between parts and wholes.

interest of property theories to explore this possibility and face its consequences for, as I have shown, it has attractive theoretical features, albeit problematic. Not only it is worth reviewing for the property theorist but also for those accounting for EV or even WV. This is where my last advice goes. Although here I provide elements that can advance sound dispositionalism, there is maybe a good reason not to take this route in a reductionist way. By 'reductionist' I mean the ontological reduction of sound to the class of potentialities (and dispositions). This explanatory misplacement is a product of the aim of terminological reduction that looms in the views in SO. And this criticism is analogous to one of the main objections towards WV. For sure, acoustic waves *have to do* with the auditory phenomena but being part of the explanatory scheme does not imply the claim "sounds are waves." There is a sense in which much of the discussion can be arranged by means of specifying our usages of 'sound' in our explanatory schemes. Maybe there is room for everyone. Yet a difference here is that whereas acoustic waves' importance is undeniable in order to have the picture of sound, dispositions or potentialities require more justification.

Avoiding a reduction of this kind would allow the necessary distance from the problem noticed by Casati and Dokic (2014) concerning the collapsing or conflation of views, which are supposedly antagonistic.[30] In conclusion, dispositions or potentialities are potentially useful tools for the philosophical investigation of the realm of the auditory phenomena and they *may* be used in the discussion, but it is perhaps unwise to reduce and circumscribe the ontology of sound to the class of dispositional properties as it might be to do so with other classes (waves or events for instance).

---

[30]   In particular, Casati and Dokic are aiming a critique towards the Relational Event Theory, which intends appealing to both the surrounding medium and the source, that is, articulating distal and medial theories. However, for these authors this view (which is O'Callaghan's arguably) "collapses unto" a medial theory when considering puzzling cases where an informational barrier prevents us to access a sounding object. In such circumstances, this theory, they say, ends up being plainly medial.

Acknowledgements

Funding

References

Armstrong, David. 1993. "A World of State of Affairs." *Philosophical Perspectives* 7: 429–40. https://doi.org/10.2307/2214133

Armstrong, David. 1996. "Dispositions as Categorical States." In *Dispositions, a Debate,* edited by David Armstrong, C. B. Martin, and P U. T. Place, 15–19. New York: Routledge.

Armstrong, David. 2005. "Four Debates about Properties." *Synthese* 144 (3): 309–20. https://doi.org/10.1007/s11229-005-5852-7

Armstrong, David, Martin, C. B., and Place, U. T. Eds. 1996. *Dispositions, a Debate.* New York: Routledge.

Bird, Alexander. 1998. "Dispositions and Antidotes." *The Philosophical Quarterly* 48 (191): 227–34. https://doi.org/10.1111/1467-9213.00098

Bird, Alexander. 2007. *Nature's Metaphysics, Laws and Properties.* Oxford: Oxford University Press.

Casati, Roberto, Di Bona, Elvira, and Dokic, Jérôme. 2013. "The Ockhamization of the Event Sources of Sound." *Analysis* 73 (3): 462–66. https://doi.org/10.1093/analys/ant035

Casati, Roberto, and Dokic, Jérôme. 1994. *La philosophie du son.* Chambon: Nimes.

Casati, Roberto, and Dokic, Jérôme. 2014. "Sounds." In *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), edited by Edward N. Zalta. Retrieved from: http://plato.stanford/archives/fall2014/entries/sounds/

Chakravarty, Anjan. 2007. *A Metaphysics for Scientific Realism, Knowing the Unobservable.* Cambridge: Cambridge University Press.

Choi, Sungho, and Fara, Michael. 2016. "Dispositions." In *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), edited by Edward N. Zalta. Retrieved from: http://plato.stanford.edu/archives/spri2016/entries/dispositions/

Cohen, Jonathan. 2010. "Sounds and Temporality." In *Oxford Studies in Metaphysics*, Vol. 5, 303–20. New York: Oxford University Press.

Dokic, Jérôme. 2007. "Two Ontologies of Sound." *The Monist* 90 (3): 391–402. https://doi.org/10.5840/monist200790327

Egan, Andy. 2006. "Secondary Qualities and Self-Location." *Philosophy and Phenomenological Research* 72 (1): 97–119. https://doi.org/10.5840/monist200790327

Fowler, Gregory. 2013. "Against the Primary Sound Account of Echoes." *Analysis* 73 (3): 466–73. https://doi.org/10.1093/analys/ant059

Gurova, Lilia. 2017. "A Reason to Avoid the Casual Construal of Dispositional Explanations." *Organon F* 24 (4): 438–55.

Handel, Stephen. 2006. *Perceptual Coherence. Seeing and Hearing.* Oxford: Oxford University Press.

Husserl, Edmund. 1928. *Vorlesungen zur Phänomenologie des inneren Zeitbewustseins.* Halle: Max Niemeyer.

Husserl, Edmund. 1984. *Logische Untersuchungen.* Hamburg: Felix Meiner.

Kulvicki, John. 2008. "The Nature of Noise." *Philosopher's Imprint* 8 (11): 1–16.

Lewis, David K. 1968. "Counterpart theory and Quantified Modal Logic." *The Journal of Philosophy* 65 (5): 113–26. https://doi.org/10.2307/2024555

Lewis, David K. 1973. *Counterfactuals*, Massachusetts, Blackwell.

Lewis, David K. 1986a. *On the Plurality of Worlds*, Oxford: Blackwell.

Lewis, David K. 1986b. *Philosophical Papers,* Vol. II. Oxford: Oxford University Press.

Lewis, David K. 1997. "Finkish Dispositions." *The Philosophical Quarterly* 47 (187): 143–58. https://doi.org/10.1111/1467-9213.00052

Lewis, David K. 2008. "Ramseyan Humility." In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, 203–22. Cambridge, Massachusetts: MIT Press.

Locke, John. 1999. *An Essay Concerning Human Understanding.* Pennsylvania: University of Pennsylvania Press.

Manley, David and Wasserman, Ryan. 2008. "On Linking Dispositions and Conditionals." *Mind* 117 (465): 59–84. https://doi.org/10.1093/mind/fzn003

Martin, C.B. 1996. "Properties and Dispositions." In *Dispositions, a Debate,* edited by David Armstrong, Martin C. B., and Place, U. T., 71–88. New York: Routledge.

McKitrick, Jennifer. 2003. "A Case for Extrinsic Dispositions." *Australasian Journal of Philosophy* 81 (2): 155–74. https://doi.org/10.1080/713659629

Méndez-Martínez, Jorge Luis. 2019. "What Counts as 'a' Sound and How 'to Count' a Sound. The Problems of Individuating and Identifying Sounds." *Synthesis Philosophica* 2019 (2): 173–90. https://doi.org/10.21464/sp34112

Méndez-Martínez, Jorge Luis. 2020. "Sound Ontology and the Brentano-Husserl Analysis of the Inner Consciousness of Time." *Horizon, Studies in Phenomenology* 9 (1): 175–204. https://doi.org/10.21638/2226-5260-2020-9-1-X-X

Molnar, George. 2003. *Powers, a Study in Metaphysics*, Oxford: Oxford University Press.

Nudds, Matthew, and O'Callaghan, Casey. Eds. 2009. *Sounds and Perception. New Philosophical Essays.* New York: Oxford University Press.

Nudds, Matthew. 2009. "Sounds and Space." In *Sounds and Perception. New Philosophical Essays,* edited by Matthew Nudds and Casey O'Callaghan, 69–96. New York: Oxford University Press.

Nudds, Matthew. 2010. "What Are Auditory Objects." *Review of Philosophy and Psychology* 1 (1): 105–22. https://doi.org/10.1007/s13164-009-0003-6

O'Callaghan, Casey. 2007a. *Sounds, a Philosophical Theory.* London: Oxford University Press.

O'Callaghan, Casey. 2007b. "Echoes." *The Monist* 90 (3): 403–14. https://doi.org/10.5840/monist200790328

O'Callaghan, Casey. 2009. "Sounds as Events." In *Sounds and Perception, New Philosophical Essays*, edited by Matthew Nudds and Casey O'Callaghan, 26–49. New York: Oxford University Press.

O'Callaghan, Casey. 2011. "Hearing Properties, Effects, or Parts." *Proceedings of the Aristotelian Society*, New Series (111): 375–405. https://doi.org/10.1111/j.1467-9264.2011.00315.x

O'Shaughnessy, Brian. 1957. "The Location of Sound." *The Monist* 66 (264): 471–90. https://doi.org/10.1093/mind/LXVI.264.471

Parker, J. and Parker, P. Eds. 2004. *Audiology, A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References.* San Diego: ICON.

Pasnau, Robert. 1999. "What is Sound." *The Philosophical Quarterly* 49 (196): 309–24. https://doi.org/10.1111/1467-9213.00144

Place, U. T. 1996. "Structural Properties: Categorical, Dispositional or Both." In *Dispositions, a Debate,* edited by David Armstrong, Martin, C. B., and Place, U. T., 105–26. New York: Routledge.

Purves, Dale. 2010. *Brains, How They Seem to Work*, New Jersey: Pearson.

Quine, Willard Van Orman 1961. *From A Logical Point of View*, New York: Harvard University Press.

Roberts, Pendaran. 2017. "Turning Up the Volume on the Property View of Sound." *Inquiry* 60 (4): 337–57.
https://doi.org/10.1080/0020174X.2016.1159979

Ryle, Gilbert. 1949. *The Concept of Mind.* New York: Routledge.

Scruton, Roger. 2009. "Sounds as Secondary Objects and Pure Events." In *Sounds and Perception. New Philosophical Essays,* edited by Matthew Nudds and Casey O'Callaghan, 50–68. New York: Oxford University Press.

Strawson, Peter Frederik. 1959. *Individuals, An Essay in Descriptive Metaphysics.* London: Routledge.

Strawson, Galen. 2009. *Selves, An Essay in Revisionary Metaphysics.* New York: Oxford University Press.

Vetter, Barbara. 2011a. "On Linking Dispositions and Which Conditionals?" *Mind* 120 (480): 1173–89. https://doi.org/10.1093/mind/fzr077

Vetter, Barbara. 2011b. "Recent Work: Modality Without Possible Worlds". *Analysis* 71 (4): 742–54. https://doi.org/10.1093/analys/anr077

Vetter, Barbara. 2013a. "'Can' Without Possible Worlds: Semantics for Anti-Humeans." *Philosophers' Imprint* 13 (16): 1–27.

Vetter, Barbara. 2013b. "Multi-Track Dispositions." *The Philosophical Quarterly* 63 (251): 330–52. https://doi.org/10.1111/1467-9213.12024

Vetter, Barbara. 2015. *Potentiality, from Dispositions to Modality.* New York, Oxford University Press.

Vetter, Barbara. 2018. "Perceiving Potentialities: A Metaphysics for Affordances." *Topoi.* https://doi.org/10.1007/s11245-018-9618-5

Vetter, Barbara. 2019a. "A Plenitude of Powers." *Synthesis.*
https://doi.org/10.1007/s11229-018-1758-z

Yates, David. 2015. "Dispositionalism and the Modal Operators." *Philosophy and Phenomenological Research* 91: 411–24. https://doi.org/10.1111/phpr.12132

Could Tajtelbaum Question What Tarski Could Not?

Jan Wiślicki*

Abstract: The paper discusses Tarski's approach to quotation. It starts from showing that it is vulnerable to semantic inconsistencies connected with what is known as *Reach's puzzle*, formulated in 1938 by a Czech logician Karel Reach. This fact gives rise to serious problems concerning the relation between the metalanguage and an object language. Moreover, the paper touches upon a historic aspect, pointing out that the problem at hand is discussed in the only paper signed up as *Al. Tajtelbaum,* i.e. Alfred Tarski's original name. It argues that the puzzle reveals the importance of reopening the discussion on the understanding and limitations of deriving the metalanguage from an object language.

Keywords: Enquotation; metalanguage; quotation; Reach's puzzle; Tajtelbaum; Tarski.

1. Introduction

The present squib delivers arguments for reopening the discussion on Tarski's approach to metalanguage. There are two crucial issues underlying the present discussion. First, Tarski's view on quotation is more often than not taken as unproblematic, also for Tarski himself; I show that these

---

* University of Warsaw
  ✎ Department of Formal Linguistics, Faculty of Modern Languages, University of Warsaw, ul. Dobra 55, 00-312, Warsaw, Poland
  ✉ j.wislicki@uw.edu.pl

assumptions are not justified. Second, reconsidering this problem gives rise to serious questions concerning the present understanding of metalanguage.

The paper is organised as follows. In Section 2 I present the problem arising for the functional approach to quotation (Tarski 1933/1983). In Section 3 I show that not only the crucial problem had been formulated in the late thirties (Reach 1938) and became known as *Reach's puzzle*, but also that it was acknowledged in Tarski's seminal paper (1933/1983). In Section 4 I discuss a note signed by Tarski's original name, *Al. Tajtelbaum* (Tajtelbaum 1957), that has remained nearly unnoticed for the last 60 years. I contrast this paper with Reach's puzzle and the related discussion, showing that Tarski might have been less certain about the nature of quotation than it is standardly assumed. In Section 5 I show why these facts are problematic for Tarskian semantics and what is the general lesson following from them. Section 6 summarizes the discussion and suggests a path for future work.

2. Quotation-function: a problem

Let us first have a look at the problem underlying the whole discussion. The puzzle arises for the functional operation of enquotation yielding quotational names. Tarski (1933/1983) defines quotation-function as follows:

> The expression " 'p' " […] must be regarded as a function, the argument of which is a sentential variable and the values of which are constant quotation-mark names of sentences. We shall call such functions *quotation-functions*. The quotation marks then become independent words belonging to the domain of semantics, approximating in their meaning to the word 'name,' and from the syntactical point of view they play the part of functors. (Tarski 1933/1983, 161)

This fragment shows that Tarski, at least in his seminal work, did not simply subscribe to the atomic treatment of quotational expressions. A toy formalisation of the above definition results in a trivial formula of the form:

(1) $Q(x) = \text{'}x\text{'}$

where x ranges over expressions of the object language (here sentences), Q stands for the operation of enquotation (most standardly marked by quotes), and 'x' for its value being a quotational name of x.

Assuming, as standardly, that a function is the meaning of the corresponding functor, I take the formula in (1) to be equivalent to the following interpretation:

(2)   The meaning of 'Q' maps the meaning of 'x' onto the meaning of ''x'.'

Note that, though not unproblematic, the equivalence of (1) and (2) is coherent with Tarski's view expressed elsewhere:

> We may, admittedly, replace this formula $[3 = 2 + 1]$ by a sentence which expresses the same idea but is about symbols, namely, by a sentence which asserts that the symbols "3" and "$2 + 1$" designate the same number. (Tarski 1994, 55)

So, in order to obtain the quotational name *'dog'* picking out the noun *dog*, the computational mechanism makes the substitution $[\text{dog}/x]$ yielding (3):

(3)   The meaning of 'Q' maps the meaning of 'dog' onto the meaning of ''dog'.'

Though at first sight innocent, (2) is in fact quite problematic. To see this, let us make the following substitution [the first word of this paper/x], perfectly fine according to (1):

(4)   The meaning of 'Q' maps the meaning of 'the first word of this paper' onto the meaning of ''the first word of this paper'.'

A quick look shows that (4) is untenable; the line of reasoning, assuming the referential approach to meaning enriched by the minimum context securing the computation of indexicals, is very simple:

(5)   i.   The meaning of the expression *the first word of this paper* is the word *could*.

         ii. Thus, the expression *the meaning of 'the first word of this paper'* can be substituted *salva veritate* by the expression *the word 'could.'*

    iii. The result of the substitution is:

        The meaning of '*Q*' maps the word 'could' onto the meaning of ''the first word of this paper'.'

    iv. This means that the quotational name of the expression *the word 'could'* is the quotational expression *'the first word of this paper.'*[1]

However, the quotational name of the expression *the word 'could'* is *'the word 'could','* not *'the first word of this paper.'* Thus iv. is contrary to the expected effect. It follows, then, that the definition of enquotation as formulated in (1) is untenable.

So, the straightforward Tarskian implementation of quotation-function gives rise to serious semantic problems. In the next section I show that the problem is not new, going back at least to Tarski (1933/1983).

3. Quotation-function and its domain

The puzzle underlying the effect laid out in (5) was first tackled by a Czech logician Karel Reach (Reach 1938). Reach's line of reasoning goes as follows. Given an object and its name, it is impossible to express the name relation holding between the two. The reason is that whenever one formulates a sentence meant to express this relation, one uses not the name but the quotational name of this name. Because any name (including quotation) and its bearer are two distinct entities, the name relation cannot be informatively expressed.

---

[1]   Without going into a detailed ontological discussion, I take an abstract word to be the meaning of its quotational name, and the particular token of this word to be what satisfies the meaning at hand. Note also that the problem laid out in (5) cannot be resolved by making further assumptions connected with the ontology of language. For instance, one might, as suggested by an anonymous reviewer, assume that the meaning of the particular expression is not a word but a token. Still, this would not resolve the problem. The obtained effect would be that the quotational name of the expression *the token 'could'* is the quotational expression *'the first word of this paper,'* contrary to expectations. Although *the first word of this paper* may seem to be an appropriate way to refer to the token *could*, it is not its quotational name.

To illustrate, suppose that someone asks about another person's name. The answer is something like (6):

(6)   His name is *Alfred*.

Clearly, when uttering (6) the speaker is not using the name, contrary to (7):

(7)   Alfred is reading a book.

Rather, what he is using is the quotational name of his name. What obscures this fact is that it is the special property of quotational names that their bearers—the units of the object language—can be immediately grasped from the form of the quotational name. Thus, the word *Alfred*—the bearer of the quotational name *'Alfred'*—can be immediately grasped on the basis of the form of its quotational name *'Alfred'* (cf. Read 1997, Yourgrau 1982).

While Reach does not explicitly address the problem of quotation, his observations share one important point with the discussion in Section 2. What is crucial for Reach's analysis is that there is a problem with setting up the relation between a name and its bearer. Put more formally, no function standing for the name relation can be defined by letting it take expressions of an object language as arguments (Anscombe 1965; Geach 1980; Mendelsohn 2005; Gaskin, Hill 2013; a.o.).

Indeed, it is exactly this fact, i.e. letting the domain of the quotational name relation be a set of expressions of an object language as in (1), that lies at the heart of the problem presented in (5). The problem arises because Q is assumed to be extensional and thus to take as arguments expressions that undergo substitution *salva veritate* for expressions of a different form but with an equivalent meaning. And what is especially interesting from the historic point of view is that Tarski was aware of that; immediately after the passage quoted is Section 2 he continues:

> But then new complications arise. The sense of the quotation-function and of the quotation marks themselves is not sufficiently clear. In any case such functors are not extensional […] a deeper analysis shows it to be impossible to give any precise meaning of such functors. (Tarski 1933/1983, 161)

Moreover, in Footnote 2 on the same page he suggests that name-forming functors should be considered as something distinct from sentence-forming functors. However, no further comments on these problems are provided, despite the fact that they are by no means marginal for Tarski's concept of metalanguage. But what had just been touched upon by Tarski was discussed in a more detailed way by Tajtelbaum. In the next section I take a closer look at Tajtelbaum's paper and try to find a more far reaching explanation of the effect at hand.

4. Tajtelbaum's doubts

What is now known as *Reach's puzzle* would have probably passed unnoticed had Anscombe (1956, 1965) not raised it in the context of her discussion on Wittgenstein's *Tractatus*. Anscombe (1956) posed the problem taken from Reach's observations in the form of competition. The winning response appeared as Tajtelbaum (1957)—a paper signed by Tarski's original name.[2] The offered solution, however, is by no means coherent with the functional approach presented in Tarski (1933/1983). The core of Tajtelbaum's approach rests upon the following idea:

> we have the (tacit) convention that a name and its name are denoted by the same word, and so the name of a name "tells" us the name. (Tajtelbaum 1957, 53)

Despite the lack of unequivocal evidence, the paper has all the hallmarks of being authored by Tarski (Sundholm 1993). This fact is not historically unmotivated. It is well known that Tarski himself was rather reluctant to express his philosophical views in articles (Sundholm 1993; Murawski 2011), sometimes deliberately (Tarski 1930/1983). Viewed from that angle the fact that the paper was published under the name *Tajtelbaum* is not that

---

[2]   The paper is not mentioned in (Burdman Feferman, Feferman 2004) who explicitly state that 'after that [1924] all his papers were published under the name *Alfred Tarski* because, shortly before receiving his Ph.D. degree in 1924, he officially changed his surname and concomitantly converted to Catholicism' (Burdman Feferman, Feferman 2004, 37–38).

surprising. Notwithstanding these historical questions, a more pressing task is to determine how Tajtelbaum's account differs from Tarski's functional approach. Note that Tarski took quotational names to be the values of the function defined on names (expressions of the object language). By contrast, Tajtelbaum's idea is that a name and its quotational name are expressed by one and the same expression, which means that there is no mapping as in (1).

As it stands, Tajtelbaum's idea invites further questions. First of all, it cannot be the case that a name and its name are one and the same. And this, of course, is not Tajtelbaum's claim; he clearly states that the two are denoted by the same word, not that they are the same object. Still, the author does not make explicit how the proper meaning—the name or its quotational name—denoted by a given word can be discriminated within a formal semantic system.

The idea sketched by Tajtelbaum has been given a new life in the form of so-called identity theory of quotation (*aka* use theory of quotation) developed by Geach (1980), Washington (1992), Saka (1998), a.o. (see also Saka 2013; Maier 2014 for some critical overviews). The idea is that every expression is primarily ambiguous. Accordingly, the meaning of the expression E is the set $\{e_1, ..., e_n\}$, where e_i: $1 \leq i \leq n$ stands for the extension, intension, lexical entry or the phonic/written form. It is the role of quotes to (partially) disambiguate the expression. When flanked by quotes, E is interpreted as denoting something else than its extension, most standardly (though not exclusively) a lexical entry or a phonic/written form. Put more formally, every use of an expression E in the context C maps this pair onto an element of the power set of E. Thus *use*: $(E, C) \rightarrow \boldsymbol{P}(E)$. If the particular utterance is unambiguous, the obtained value is a singleton.

If, then, it is the role of quotes (the quotation functor Q) to disambiguate the use of expressions, the situation roughly looks as follows:

(8) $use(Q(E), C) = \{e_i, ..., e_k\}$, where:
         $\{e_i, ..., e_k\}$ is an element of $\boldsymbol{P}(E)$ & $i \leq j \leq k$ & $j \neq$ *the extension of E*

Within this approach *Alfred* in both (6) and (7), repeated below, is a primarily ambiguous single expression:

(6)   His name is *Alfred.*

(7)   Alfred is reading a book.

It is the context of its use that is responsible for disambiguation. In (7) the *use* function maps it onto its extension, perhaps as a default value chosen due to the lack of other markers (e.g. modal operators or a marker of quotation). By contrast, in (6) the operation of enquotation (here marked by the italic) results in the *use* function yielding the expression of the form *Alfred.*

The picture sketched in (8) is different from the one in (1). Here quotation-function is not defined on expressions of an object language, as was crucial for Tarski (1933/1983). In this regard it is safe from the threat exposed in (5). Instead, it takes as its argument an underspecified linguistic object E, a sort of root in the sense of Marantz (1997; 2000), as investigated within the generative inquiry. When additionally mapped by quotes, E denotes the word at hand, so that *the name of a name "tells" us the name*, as put by Tajtelbaum in the passage quoted above. Accordingly, an expression of the metalanguage is an output of operation defined on an underspecified item, rather than on an expression of object language.

So, Tajtelbaum's approach to quotation is orthogonal to the account of quotation provided within Tarski's truth theory. This fact, when viewed in the context of Reach's puzzle, deserves a more careful treatment. As concluded by Tajtelbaum (1957, 53), the problem *draws attention to a limitation of formal analysis*. What the discussion taken up by Tajtelbaum suggests is that there are serious doubts whether defining the metalanguage does not extend those limitations. Put differently, there are doubts whether the relation between an object language and the metalanguage can be expressed in an informative and formal way. In this sense they are more severe than those cited from Tarski (1933/1983) in Section 3. In the next section I will discuss some issues following from these doubts.

5. The emergence of the metalanguage: reactivation

The above concern raised by Tajtelbaum with respect to the limitation of formal analysis unearths a deeper problem than just a formal aspect of

a theory of quotation. It shows that the present understanding of how the metalanguage is derived from an object language might be much weaker than it has standardly been assumed.

It might be apt at this point to recall a different yet closely connected problem raised by Tarski and tackled by Soames (1999). Assume the following definition of truth (cf. Tarski 1933/1983, 159–60):

(9)   for every t ('t' is true iff t)

The puzzle is that the substitution of t in (9) by any sentence cannot be blind. In particular, it cannot target the letter t occurring in the word *true.* The approach proposed by Soames (1999) is based on the fact that (9) works perfectly fine if we are able to *distinguish a set of substitutional variables that we wish to use to quantify into such constructions* (Soames 1999, 88). Thanks to this distinction, *true* is somehow frozen and thus protected from the unwanted substitution.

Note, however, that that kind of solution works insofar as the relevant distinction is possible. In the case presented above, the distinction separates variables from constants, being thus hardly questionable. But the problem given in (5) looks different, mainly because it does not boil down to the logical distinction as the one above. The problem in (5) arises if what is flanked by quotes undergoes substitution *salva veritate.* Thus for (5) not to arise, the material within quotes must have been somehow frozen, on a par with t in *true* being frozen for the unwanted substitution. Still, regardless of the reasoning behind this effect, it would mean that freezing is absolute, targeting equally each expression flanked by quotes. In this sense whatever is flanked by quotes (i.e. taken by the quotation-function as an argument) would be treated as a purely material string. But if such a blindly overall freezing is assumed, a reasonable question arises whether within this approach the domain of quotation-function is not shifted from expressions of object language into material strings they are represented by.

If this reasoning is on the right track, then Tajtelbaum's doubts unearth an important problem, pointed out by Dummett:

> The presence in our language of various meaning-theoretic terms forces us, as we saw, to *impose* on it a distinction between object-language and metalanguage which is not there in reality. And we

shall want to draw the line so as to put into the metalanguage only those terms, and those uses of such terms, which really do serve the purpose of expressing some imperfectly formed ideas we have about how our language functions—or, to put it differently, which could be understood only as having a place in a meaning-theory for the rest of the language. Now, if one of these terms, considered as subject to a certain type of characterisation, would not play any useful role in such a meaning-theory, it is either useless or belongs (in so far as it is so characterised) on the other side of the line, to what we ought to take as constituting the object-language. And that is how we ought to view the term 'true,' considered as characterised either directly by the requirement that each instance of the (T) schema holds, or by a Tarskian truth-definition to which the fact that the metalanguage is an expansion of the object-language is taken as essential. (Dummett 1991, 71–72)

That is, leaving aside the exact way one formalises the metalanguage, it is crucial for Tarski's semantics that the metalanguage is defined as being derived from an object language.[3] In a similar vein Simchen (2003) points out that the a priori character of T-sentences, crucial for Tarski's theory, is secured insofar as the metalanguage extends an object language. Of course, a functional account as the one in (1) is not the only way of defining this relation, and neither Dummett nor Simchen claims so. As pointed out by Shapiro (1997, 49), the minimum condition is that of metalanguage being *a faithful representation of the object language.* Still, the way one is related to the other must be established. If, on the other hand, this extends the limitation of formal analysis, then the problem is quite severe. This limitation was suggested in Tajtelbaum's note and can be inferred from (8). According to this approach, an object language and the metalanguage are just

---

[3]   The quote from Dummett is important for yet another reason. In his seminal paper Tarski (1933/1983) proposes a number of other approaches to quotation, many of which have been developed in the literature (see Cappelen & Lepore 2007; Saka 2013; Maier 2014 for some relevant surveys). Nevertheless, as pointed out by Dummett, it is the functional approach, according to which the metalanguage is derived from the object language, which is crucial for Tarskian semantics.

two different values of the *use* function defined on underspecified linguistic objects. Still, the function does not relate expressions of an object language with their metalinguistic names. Put differently, we know how to obtain expressions of an object language from some underspecified objects. We also know how to obtain expressions of the metalanguage from such underspecified objects. Perhaps we can even stipulate criterions of faithfulness mentioned by Shapiro.[4] However, what remains beyond the present understanding is whether it is possible to set up (and, if yes, how) a relation securing a systematic, extensional connection between the two. In this sense while the truth theory itself is secured by the fact that we can derive an object language and the metalanguage, the understanding of truth theory is weaker to the extent the relation between the two extends the limitation of formal analysis, and thus a precise understanding.

6. Conclusion and future prospects

Tajtelbaum's idea, developed as the identity theory of quotation, is not the only way for circumventing[5] the unwanted effects discussed in Section 2. Perhaps Davidson's (1979) demonstrative theory is the most established one in the current literature (see Reimer 1996; Predelli 2008; Maier 2014 for some relevant comments). Still, the theory has hardly been discussed from the point of view of Tarski's semantics in the context of the abovementioned comments provided by Dummett (1991) and Simchen (2003). And the problem is relevant, bearing in mind that various authors (Goldstein 1985; Kamp 1995) pointed out that under this theory the operation of enquotation is defined not on words of object language, but on their inscriptions. This,

---

[4]   It is also not at all unproblematic whether *faithfulness* is the right notion. Various linguistic tests show that quotation is a vague, scalar (*Jones said something very similar to the Japanese 'Kanpai'*) and highly context-dependent object (De Brabanter 2017). In this regard linguistic data suggest that it is conditions of a successful demonstration, rather than faithfulness, that should be investigated in this context. Accordingly, the problem of how successful demonstration relates to the problem of the a priori character of T-sentences is pressing and challenging (Simchen 2003).

[5]   Some of them mentioned in Tarski (1933/1983, 162).

of course, provides a solution to the puzzle laid out in (5). To see this, let capital letters stand for inscriptions of expressions and consider the following line of reasoning:

(10)   i. The inscription of the expression *the first word of this paper* is THE FIRST WORD OF THIS PAPER.

   ii. The meaning of 'THE FIRST WORD OF THIS PAPER' is an object of the form of THE FIRST WORD OF THIS PAPER.

   iii. Quotation-function takes inscriptions as arguments. Thus, the following is true:

        The meaning of 'Q' maps the meaning of 'THE FIRST WORD OF THIS PAPER' onto the meaning of ' 'THE FIRST WORD OF THIS PAPER'.'

   iv. Given ii., *the meaning of THE FIRST WORD OF THIS PAPER* cannot by substituted *salva veritate* by *the word 'could.'*

Given that the quotation-function maps inscriptions of expressions, not expressions as such, onto quotational names, it is not vulnerable to the puzzle laid out in (5). However, since inscriptions and expressions are two different classes of objects, quotation-function as described in (10) no longer relates an object language with the metalanguage. This, in turn, suggests that problems arising for Tarski's semantics as those discussed above in the context of comments given by Dummett and Simchen are not just a matter of the particular approach to quotation, but rather of the nature of quotation as such.

The lesson from these observations is that building semantics upon the metalanguage as functionally derived from an object language is misguided. It is not a matter of the way linguistic expressions are mapped onto expressions of the metalanguage. Rather, it is a matter of how expressions are evaluated, so that they can be interpreted as extensional, metalinguistic or other expressions. In this sense the quotes seem to be part of modality, i.e. part of grammar securing various evaluations of particular expressions, rather than of functions mapping one object onto another. The present squib was aimed at showing that observations underlying this approach have strictly logical motivations, whose importance did not remain unnoticed by Tarski.

Acknowledgments

References

Anscombe, G.E.M. 1956. "The Tenth 'Problem'." *Analysis* 16 (6): 121.

Anscombe, G.E.M. 1965. *Introduction to Wittgenstein's Tractatus*. New York: Harper Torchbooks.

Cappelen, Herman and Lepore, Ernest. 2007. *Language Turned on Itself. The Semantics and Pragmatics of Metalinguistic Discourse*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231195.001.0001

Davidson, Donald. 1979. "Quotation." *Theory and Decision* 11 (1): 27–40.

De Brabanter, Philippe. 2017. "Why Quotation Is Not a Semantic Phenomenon, and Why It Calls for a Pragmatic Theory." In *Semantics and Pragmatics: Drawing a Line*, edited by Ilse Depraetere an Raphael Salkie, 227–54. Springer. https://doi.org/10.1007/978-3-319-32247-6_14

Dummett, Michael. 1991. *The Logical Basis of Metaphysics*. Cambridge, Massachusetts: Harvard University Press.

Burdman Feferman, Anita and Feferman, Solomon. 2004. *Alfred Tarski. Life and Logic*. New York: Cambridge University Press.

Gaskin, Richard and Hill, Daniel J. 2013. "Reach's Puzzle and Mention." *dialectica* 67 (2): 201–22. https://doi.org/10.1111/1746-8361.12021

Geach, Peter T. 1980. "Quotation and Semantics." In Peter T. Geach, *Logic Matters*. 189-211. Berkeley and Los Angeles: University of California Press.

Goldstein, Laurence. 1985. "The Title of This Paper is 'Quotation'." *Analysis* 45 (3): 137–40. https://doi.org/10.1093/analys/45.3.137

Kamp, Gary. 1995. "Salmon on Fregean Approaches to the Paradox of Analysis." *Philosophical Studies* 78 (2): 153–62. https://doi.org/10.1007/BF00989679

Maier, Emar. 2014. "Pure Quotation." *Philosophy Compass* 9 (9): 615–30. https://doi.org/10.1111/phc3.12149

Marantz, Alec. 1997. "No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon." *University of Pennsylvania Working Papers in Linguistics* 4 (2): 201–25.

Marantz, Alec. 2000. "Roots: The Universality of Root and Pattern Morphology." Paper presented at the Conference on Afro-Asiatic Languages, Université de Paris.

Mendelsohn, Richard L. 2005. *The Philosophy of Gottlob Frege*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511497964

Murawski, Ryszard. 2011. *Filozofia matematyki i logiki w Polsce miedzywojennej.* Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Predelli, Stefano. 2008. "The Demonstrative Theory of Quotation." *Linguistics and Philosophy* 31 (5): 555–72. https://doi.org/10.1007/s10988-008-9042-1

Reach, Karel. 1938. "The Name Relation and the Logical Antinomies." *The Journal of Symbolic Logic* 3 (3): 97–111. https://doi.org/10.2307/2267594

Read, Stephen. 1997. "Quotation and Reach's Puzzle." *Acta Analytica* 19: 9–20.

Reimer, Marga. 1996. "Quotation Marks: Demonstratives or Demonstrations?" *Analysis* 56 (3): 131–41. https://doi.org/10.1093/analys/56.3.131

Saka, Paul. 1998. "Quotation and the Use-Mention Distinction." *Mind* 107 (425): 113–35. https://doi.org/10.1093/mind/107.425.113

Saka, Paul. 2013. "Quotation." *Philosophy Compass* 8 (10): 935–49. https://doi.org/10.1111/phc3.12069

Shapiro, Stewart. 1997. *Philosophy of Mathematics: Structure and Ontology.* New York: Oxford University Press. https://doi.org/10.1093/0195139305.001.0001

Simchen, Ori. 2003. "Meaningfulness and Contingent Analyticity." *Noûs* 37 (2): 278–302. https://doi.org/10.1111/1468-0068.00439

Soames, Scott. 1999. *Understanding Truth.* New York: Oxford University Press. https://doi.org/10.1093/0195123352.001.0001

Sundholm, Göran. 1993. "Tractarian Expressions and Their Use in Constructive Mathematics." In *Philosophie der Mathematik, Proceedings 15th International Wittgenstein Symposium*, edited by Johannes Czermak, 105–18. Wien: Verlag Hölder-Pichler-Tempsky.

Tajtelbaum, Al. 1957. "It Is Impossible to Be Told Anyone's Name." *Analysis* 17 (3): 52–53. https://doi.org/10.1093/analys/17.3.52

Tarski, Alfred. 1930/1983. "Fundamental Concepts of the Methodology of the Deductive Sciences." In A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, edited by John Corcoran, 60–109. Indianapolis: Hackett Publishing.

Tarski, Alfred. 1933/1983. "The Concept of Truth in Formalized Languages." In A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, edited by John Corcoran 152–278. Indianapolis: Hackett Publishing.

Tarski, Alfred. 1994. *Introduction to Logic and to the Methodology of Deductive Sciences*, Fourth Edition. New York: Oxford University Press.

Washington, Corey. 1992. "The Identity Theory of Quotation." *The Journal of Philosophy* 89 (11): 582–605. https://doi.org/10.2307/2941057

Yourgrau, Palle. 1982. "Frege, Perry, and Demonstratives." *Canadian Journal of Philosophy* 12 (4): 725–52. https://doi.org/10.1080/00455091.1982.10715813

Falsificationism and the Pragmatic Problem of Induction

Danny Frederick*

Abstract: I explain how Karl Popper resolved the problem of induction but not the pragmatic problem of induction. I show that Popper's proposed solution to the pragmatic problem of induction is inconsistent with his solution to the problem of induction. I explain how Popper's falsificationist epistemology can solve the pragmatic problem of induction in the same negative way that it solves the problem of induction.

Keywords: Conjecture; falsification; induction; Karl Popper; pragmatic problem of induction; rational action.

1. Introduction

The problem of induction seems to threaten the possibility of empirical knowledge. The standard response is to try to show that the problem is not real by arguing, in some specious way or other, that we have justified claims to knowledge. Karl Popper's response is to bypass the problem by showing that we can have knowledge without justification. However, unjustified theories do not seem to supply a basis for rational action. That is the pragmatic problem of induction. Popper's solution to that problem has satisfied few.

*   unaffiliated
    ✎ 13 Sandhurst Road, Yeovil, Somerset, BA20 2LG, United Kingdom
    ✉ dannyfrederick77@gmail.com

I show that Popper's solution to the pragmatic problem of induction is inconsistent with his solution to the problem of induction. I advocate resolving the pragmatic problem of induction in line with Popper's epistemology.

2. Induction

The problem of induction was formulated by David Hume. An individual observation provides us with information about itself but it provides us with no information about other observations that we may make. Consequently, our general theories about the world cannot be derived from, or justified by, our observations. For example, even if every swan that we have observed has been white, it does not follow that swans that we have *not* observed will also be white; and it does not follow that it is *probable* that swans that we have *not* observed will be white, because we can have no idea whether our experience of things so far has been *representative* of how things are in general. We can therefore have no justification for thinking that our general theories are either true or even probably true. We may find our general theories plausible, or 'subjectively probable,' given our particular experiences or inclinations; but we can have no grounds for linking such plausibility or subjective probability to truth or to objective probability (Hume 1888, part III, section vi, 86–94 and section xii, 139; part IV, section ii, 218; 1975, section IV, part II, 32–39; section V, part I, 40–47).

In general, philosophers who have responded to Hume's problem have tried to show that Hume was mistaken. Some philosophers, including Kant, have attempted to show that some of our general theories can be justified *a priori*. Such attempts were discredited more than a century ago. Gottlob Frege's fifth axiom for arithmetic, which had been assumed to be self-evident, was shown by Bertrand Russell to be self-contradictory. As a consequence, for all we can know, any proposed *a priori* justification might fail because it contains a latent inconsistency (Russell 1959, 58). The point was underscored in 1931 by Gödel's second incompleteness theorem, which showed that any theory that assumes arithmetic, as virtually all theories of any interest do, cannot be proved to be consistent. Other philosophers have tried to show that our general theories are justified in a weak sense, that they are "supported" or "prima-facie justified." However, these philosophers

admit that such weakly justified theories may turn out to be false. They therefore provide no solution to Hume's problem. It remains the case that we are not justified in thinking that our general theories are true or even probably true in an objective sense of 'probability.' Indeed, even if it could be shown that supported or prima-facie justified theories were, objectively, probably true, we would not be justified in thinking them to be true, since even a highly probable theory may still be false.

Karl Popper (1972a) showed the way out of this quagmire. The fact that we cannot obtain general theories that are justified is not a calamity because we can proceed well enough with general theories that are *conjectures* that we *test* against statements describing particular events that we observe. All such observation statements are theory-laden; but we can usually reach agreement on such observation statements because we can usually find a way of describing our observations that is neutral between whatever theories are currently in contention (Frederick 2016, 641). If a general theory contradicts such an observation statement, we reject it, otherwise we retain it for additional testing. Further, since any of our conjectures may be false and may, indeed, end up being rejected after testing, we should strive to come up with *better* conjectures than the ones we already have, either by modifying existing conjectures or by propounding novel ones, and we should look for ways to test those conjectures. We can agree that one unrefuted conjecture is *better* than another if it explains more or explains more simply and if it implies more novel empirical predictions that survive testing. What we should not do is waste time trying to justify any existing conjectures, not only because it cannot be done, but also because it is an attempt to block progress, to stand still instead of moving forward. Contemporary epistemology, being preoccupied with justification, is thus epistemically perverse: it stands opposed to the growth of knowledge.

Philosophical objections to Popper's response to Hume's problem tend to focus on issues concerning falsification and truth. It is frequently objected that scientific theories are usually so general that they clash with no observation statements: it is only a conjunction of such theories plus statements describing particular situations that are inconsistent with observation statements. However, all that means is that, when such an inconsistency appears, we have a choice of which theory or statement to revise. Popper proposes

that any revision is acceptable if, but only if, it solves a problem in addition to the problem of removing the inconsistency; and, ideally, it should imply novel falsifiable predictions that survive testing. Thus, acceptable revisions are those that constitute progress in the growth of knowledge (Popper 1972a, chapter 4, sections 19 and 20).

It is often objected that accepted observation statements, being theory-laden, may be false; so a theory that is falsified by such a statement might be true. However, observation statements may be tested also, so a previously accepted observation statement may be rejected later. For instance, a statement of the form 'this is a glass of water' implies that the liquid in the glass will behave in the law-like way that water does; and that implication may be inconsistent with other observation statements (Popper 1972a, chapter 5 and appendix *x, (2)). Admittedly, we can never be sure that an observation statement is true or, correlatively, that a falsified theory is false; but we can, if we are lucky, obtain theories that are better than any that we have so far and we can also adopt procedures that prohibit us from accepting a new theory, or an amendment to theory, that constitutes a regress in the growth of knowledge. We do that by prohibiting "ad hoc manoeuvres" or "conventionalist stratagems" which save a theory from falsification without solving any additional problem (Popper 1972a, chapter 4, sections 19 and 20).

It is sometimes objected that such progress in the growth of knowledge may be achieved even if all our accepted theories and observation statements are false. But that sort of objection just returns to the justificationist quagmire. We cannot know whether our theories are true or false; but, as Popper showed, we can recognise when we are making epistemic progress, that is, when we have theories that explain more, or more simply, than their predecessors and that successfully predict novel facts.

3. Rational action

The fact that, although we cannot know whether any of our theories is true, we can repeatedly propose new and more promising conjectures and attempt ruthlessly to falsify them and replace them with better ones, may be inspiring, even exhilarating, from a theoretical point of view. But from

the point of view of practical action, the fact that even our best theories may be false can be frightening. Acting on a false theory can produce unwelcome, even calamitous, consequences. We may perhaps take some comfort from acting on the theory that has best survived criticism so far; but how can it be rational to do that if even the best theory available may be false and may, indeed, be falsified in the next moment, in fact in the very moment that we act on it? If rational action is to be possible, it may seem, there must be some principle of induction that vouches for the reliability of the currently best theory. That is the pragmatic problem of induction.

Popper's solution to the pragmatic problem of induction has satisfied few. It is to insist, first, that there can be no guarantee that the theory that has best survived criticism so far is reliable, and, second, that it is nevertheless rational to act on that theory because it is identified as the theory that has best survived criticism so far by means of a rational process (or the most rational process):

> From a rational point of view we should not 'rely' on any theory, for no theory has been shown to be true, or can be shown to be true [...] But we should prefer as the basis for action the best-tested theory [...] This will be 'rational' in the most obvious sense of the word known to me: the best-tested theory is the one which, in the light of our critical discussion, appears to be the best so far, and I do not know of anything more 'rational' than a well-conducted critical discussion (1972c, 21–22).

> [P]ractical preference for the theory which in the light of the rational discussion appears to be nearer the truth is risky but rational (1972b, 282).

> When faced with the *need to act*, on one theory or another, the rational choice was to act on that theory—if there was one—which so far had stood up to criticism better than its competitors had (1974, 82).

> [I]n so far as we accept, or reject, a scientific theory as a basis for practical action, this means choosing one theory rather than another. Where we are in the position to make such a choice, it will be rational to choose, of the two competing theories, that which has survived prolonged critical discussion, including tests (1983, 62).

Some philosophers find Popper's solution unsatisfactory because it does not provide the justification of the currently-best theory that they think is needed to make action in line with that theory rational. Even philosophers who have been attracted to, and have accepted, Popper's conjecture-and-refutation response to the problem of induction have stumbled back into justificationism with regard to the *pragmatic* problem of induction. For example, Elie Zahar (1997, 144–45) says:

> In theoretical science, the notion of truth as correspondence alone serves as a regulative idea. We can dispense with any inductive principle, and attribute different degrees of corroboration to our laws, without having to make a judgement on their future performance. Popper has not however been willing to admit that matters are different in technology, where the notion of reliability remains unavoidable (quoted and translated in Miller 2006, 114).

Similar misgivings are expressed by Watkins (1999, viii) and Worrall (1989). But such misgivings are misplaced. They amount to exclaiming that, when it comes to action, we want theories that we know we can rely on! That may be true; but we would simply be deluding ourselves if we thought that we could have such theories—because of the problem of induction.

A more serious problem with Popper's solution is that, in recommending that we act on the theory that has best survived criticism to date, Popper seems to have slipped into a form of inductivism. In that case, Popper's solution to the pragmatic problem of induction is inconsistent with his solution to the problem of induction. Given that Popper's solution to the problem of induction appears to be correct, his solution to the pragmatic problem of induction should be rejected. *It cannot be rationally required to act on the currently best theory.* That may seem counter-intuitive; but we can see that it is true if we again consider scientific practice.

Recall that Popper's response to the problem of induction is to commend that we invent numerous conjectures that we attempt to falsify. Even the theory that has best survived criticism so far should be subject to further attempts to refute it. However, a scientist who attempts to refute the theory that has best survived criticism so far is attempting to create a situation described by a statement which is inconsistent with the theory. That is, she is trying to do something that the theory that has best survived criticism

so far says cannot be done. She does not, therefore, act as if the theory that has best survived criticism so far is true. Yet her action is rational, because the growth of knowledge requires that we test our theories. Further, her action may be *successful*; for she might refute the theory. For example, despite the most advanced optical theory (Kepler's) ruling out the possibility of a telescope, Galileo attempted to construct one and he succeeded (Feyerabend 1975, 105, footnote 21); and Edison's production of his electric light refuted the unanimous scientific opinion that such a light was impossible (Kuhn 1977, 238). If those people had insisted on acting in accord with the theory that had (so far) best survived the most searching criticism, they would have failed to achieve their successes.

Popper's proposed solution to the pragmatic problem of induction says that we are rationally required to *act on* the theory that has best survived criticism so far. His non-inductive epistemology says that we are rationally permitted, and even encouraged, to *act against* the theory that has best survived criticism so far (by trying to refute it). The two are inconsistent. The epistemology does not say that we are rationally *required* to act against the theory that has best survived criticism so far, since one is not rationally required to spend one's whole life attempting to refute the theory that has best survived criticism so far. According to the epistemology, one is rationally permitted to choose *either* to act on *or* to act against the theory that has best survived criticism so far. Of course, there are various ways of acting against a theory, or trying to refute it: to comment on the hackneyed example (for example, Worrall 1989, 258–60), if one wants to refute Galileo's 'law' of freefall, one may choose many ways to do it other than throwing oneself out of a tenth-story window.

Popper's mistake, then, was to maintain that one is rationally *required* to act in accord with the theory that has best survived criticism so far. That makes sense only on the assumption that the theory that has best survived criticism so far will continue to be successful, that it will not be refuted. It is a version of the inductivist 'principle of continence'

> perform the action judged best on the basis of all available relevant reasons (Davidson 1980, 41).

Given his anti-inductivist epistemology Popper should instead have maintained that, while it is rationally *permitted* to act in accord with the

theory that has best survived criticism so far, it is also rationally *permitted* to act against it (as, in fact, scientists often do). This is the only solution to the pragmatic problem of induction that is consistent with Popper's epistemology and the only solution that is consistent with the logical facts. Moreover, it is a solution consistent with the behaviour of scientists, who often try to refute the theories that have best survived criticism so far, and of people generally when deciding how to act (recall Edison).

Popper sometimes comes close to stating this solution:

> it is perfectly reasonable to *act* on the assumption that [the future] will, in many respects, be like the past, and that well-tested laws will continue to hold (since we can have no better assumption to act upon); but it is also reasonable to believe that such a course of action will lead us at times into severe trouble, since some of the laws on which we now heavily rely may easily prove unreliable (1963, 56).

However, that he did not quite get there can be seen from the emphasis he places on the word 'act,' which shows him *contrasting* the pragmatic with the logical or epistemic problem of induction.

4. Conclusion

It seems that we naturally assume that some of our general theories about the world are justified by our observations. Yet our observations, being of particular events, give no legitimate grounds for general propositions. That is the problem of induction. Generally, the responses of philosophers have been attempts to solve the problem positively by, forlornly, defending the natural assumption. In contrast, Popper resolved the problem negatively, by pointing out that the falsity of the natural assumption does not prevent us from gaining ever-better general theories about the world, by reaching agreement on descriptions of particular events that we observe, formulating explanatory general theories about the world, and then testing those theories against the agreed descriptions of observed events.

It seems that we naturally assume that rationality requires that we act as if our best-tested theories are true. Yet the problem of induction shows

that our best-tested theories may be false. That is the pragmatic problem of induction. Generally, the responses of philosophers have been attempts to solve the problem positively by defending the natural assumption. That would work if there were a successful positive solution to the problem of induction; but there cannot be such. Popper's attempt to combine a positive solution to the pragmatic problem of induction with his negative solution to the problem of induction was self-contradictory.

Like the problem of induction, the pragmatic problem of induction has only a negative solution. The natural assumption is that rationality requires that we act as if our best-tested theories are true. But the natural assumption is false. Rationality *permits* us to act in accord with our best-tested theories, since they may be true; but it also *permits* us to act against them, precisely because our best-tested theories may be false and may, indeed, be refuted when we act against them. As Popper often emphasised (1994, 12), the growth of our knowledge is marked by the revolutionary overthrow of theories that had previously been successful.

References

Davidson, Donald. 1980. "How is Weakness of the Will Possible?" In Donald Davidson, *Essays on Actions and Events*, 21–42. Oxford: Clarendon Press.

Feyerabend, Paul. 1975. *Against Method*. London: New Left Books.

Frederick, Danny. 2016. "Ethical Intuitionism: A Structural Critique." *Journal of Value Inquiry* 50 (3): 631–47. https://doi.org/10.1007/s10790-016-9547-8

Hume, David. 1888. *A Treatise of Human Nature*, Book I, edited by L. A. Selby-Bigge. Oxford: Clarendon.

Hume, David. 1975. *An Enquiry Concerning Human Understanding*. In David Hume, *Enquiries*, edited by L. A. Selby-Bigge, third edition, 1–165. Oxford: Clarendon.

Kuhn, Thomas. 1977. "The Essential Tension." In Thomas Kuhn, *The Essential Tension*, 225–39. Chicago: University of Chicago Press.

Miller, David. 2006. "Induction: A Problem Solved." In David Miller, *Out of Error*, 111–32. Abingdon: Ashgate.

Popper, Karl. 1963. "Science: Conjectures and Refutations." In Karl Popper, *Conjectures and Refutations*, 33–65. London: Routledge.

Popper, Karl. 1972a. *The Logic of Scientific Discovery*. London: Hutchinson.

Popper, Karl. 1972b. "Addendum." In (Popper 1972a, 281–82).

Popper, Karl. 1972c. "Conjectural Knowledge: My Solution of the Problem of Induction." In Karl Popper, *Objective Knowledge* 1–31. Oxford: Clarendon.

Popper, Karl. 1974. "Intellectual Autobiography." In *The Philosophy of Karl Popper*, Vol. 1, edited by Paul Arthur Schilpp, 3–181. La Salle, IL: Open Court.

Popper, Karl. 1983. *Realism and the Aim of Science.* London: Routledge.

Popper, Karl. 1994. "The Rationality of Scientific Revolutions." In Karl Popper, *The Myth of the Framework*, 1–32. London: Routledge.

Russell, Bertrand. 1959. *My Philosophical Development.* London: George Allen and Unwin.

Watkins, John. 1999. *Human Freedom After Darwin.* La Salle, IL: Open Court.

Worrall, John. 1989. "Why Both Popper and Watkins Fail to Solve the Problem of Induction." In *Freedom and Rationality*, edited by Fred D'Agostino and Ian Jarvie, 257–96. Dordrecht: Kluwer.

Zahar, Elie. 1997. *Leçons d'épistémologie.* Paris: CREA/École polytechnique.

RESEARCH ARTICLE

Fregean Monism: A Solution to the Puzzle of Material Constitution

Soo Lam Wong*

Abstract: The puzzle of material constitution can be expressed in at least two ways. First, how can the constituting object and the constituted object, which are materially and spatially coincident, be regarded as different objects? Second, how can the constituting object and the constituted object, which are qualitatively distinct, be regarded as identical objects? Monists argue that the constituting and constituted objects are identical since they are materially and spatially coincident and the property differences between then are simply differences in description, perspective or context. In contrast, pluralists argue that the constituting and constituted objects are not identical even if they are materially and spatially coincident since they are qualitatively distinct. This paper proposes a solution to the puzzle of material constitution called 'Fregean Monism' (FM), and shows that it can better account for the property differences between the constituting and constituted objects without the need to regard them as two distinct objects. On the FM view, the puzzle of material constitution is partly a semantic puzzle and partly a metaphysical puzzle, and shows how a solution to the semantic part of the puzzle, based on the Fregean distinction between sense and reference, can yield a satisfactory solution to the metaphysical part of the puzzle. The key idea is that while the reference of a term picks out both the

* Singapore University of Social Sciences
  ✎ Singapore University of Social Sciences, Centre for University Core, 463 Clementi Road, Singapore 599494
  ✉ huang__shunan@yahoo.com.sg

referent object and referent properties, the sense of the term determine which referent properties are picked out.

1. Introduction

Material constitution is a one-one or whole-whole relation. It concerns the relation of one object (for example, a piece of bronze) constituting another object (for example, a bronze statue) where both the constituting and the constituted objects are materially and spatially coincident. The puzzle of material constitution can be expressed in at least two ways. First, how can the constituting object and the constituted object, which are materially and spatially coincident, be regarded as different objects? Second, how can the constituting object and the constituted object, which are qualitatively different, be regarded as identical objects? In contrast, material composition is a many-one or part-whole relation. It concerns the relationship of two or more objects (for example, two or more pieces of bronze) constituting a further object (a bronze statue) where both the constituting object and the constituted object share at least some of the same material parts. As this paper is about the puzzle of material constitution, and not material composition, I shall not discuss problems arising from material composition further.

Solutions to the puzzle of material constitution can be divided into two broad camps. The first camp claims that material constitution is simply identity (for example, the piece of bronze constituting the bronze statue is identical to the bronze statue). As the piece of bronze and the bronze statue are materially and spatially coincident, they are identical. In contrast, the second camp claims that material constitution is not identity despite their material and spatial coincidence (for example, the piece of bronze constituting the statue is not identical to the bronze statue) because they do not share all of the same properties (for example, the bronze statue, but not the piece of bronze, possesses aesthetic appeal). Following Kit Fine (2003), I shall label the first camp monism and the second camp pluralism. In holding that the constituting and constituted objects are identical, pluralist

critics claim that monism cannot account for how they have different prop-
erties. And in holding that the constituting and constituted objects are dis-
tinct, monist critics claim that pluralism cannot account for why there can
be two or more materially and spatially coincident objects.

My sympathy lies with the monist camp. However, I agree with the
pluralist critics that if the constituting and constituted objects are identical,
their qualitative differences have to be explained, and not explained away
as mere differences in description, perspective, or context. And I propose
a version of monism called 'Fregean Monism' (FM), to better account for
the property differences between the constituting and constituted objects
without the need to regard them as two distinct objects. On the FM view,
the puzzle of material constitution as partly a semantic puzzle and partly
a metaphysical puzzle, and shows how a solution to the semantic part of
the puzzle, based on the Fregean distinction between sense and reference,
can yield a satisfactory solution to the metaphysical part of the puzzle. The
key idea is that while the reference of a term picks out both the referent
object and its referent properties and thereby addresses the metaphysical
part of the puzzle, the sense of the term determines which referent proper-
ties of the referent object are picked out and thereby addresses the semantic
part of the puzzle.

In what follows, I shall outline the puzzle of material constitution, ex-
plicate the Fregean-Monist solution to the puzzle, show how this solution
can address pluralist criticisms against monism, and critically evaluate the
master argument employed by pluralists—the argument from Leibniz's
Law, before concluding that the property differences between the constitut-
ing and constituted objects can be accounted for without the need to regard
them as two distinct objects.

2. The puzzle of material constitution

The puzzle of material constitution can be illustrated through the example
of the piece of bronze and the bronze statue. Suppose that, on Monday,
a sculptor purchases an unformed piece of bronze, which he names 'Lump'.
Suppose further that, on Tuesday, he sculpts the bronze into the form of king
David and names it 'David.' Monists claim that the sculptor possesses only

one object, for David is identical to Lump because they are materially and spatially coincident. Yet, pluralists claim that Lump and David differ in various respects. First, Lump and David differ in their temporal properties: Lump existed on Monday, David did not. Second, Lump could survive being squashed, David could not. Third, they differ in kind: Lump is a piece of bronze, David is a bronze statue. They claim that if Lump and David differ in even one way, they are not identical, for Leibniz's Law tells that for any x and y, if x and y are distinct then there is at least one property that x has and y does not, or conversely, for any x and y, if x and y are identical, then x and y share all of the same properties. Thus, they conclude that the sculptor possesses two objects: a bronze statue and a piece of bronze. More generally, pluralists claim that there must be two material objects existing in the same place at the same time because they do not share all of the same properties. Moreover, pluralists claim that two material objects can exist in the same place at the same time because they are objects of a different kind.

On the contrary, monists claim that spatially coincident objects are impossible. Wiggins (1968) calls the monist claim—that two things cannot completely occupy exactly the same place or exactly the same volume for exactly the same period of time—a truism. L.A. Paul (2010) neatly summarises the monist arguments for constitution as identity. First, monists can claim that any differences in essence or other properties between David and Lump are only apparent; they are just differences in description based on different contexts. For example, we call the object 'David' when we ascribe an essence including being statue-shaped and having aesthetic appeal and we call it 'Lump' when we ascribe a different essence and properties such as its chemical structure and what it is suited to build. In other words, the property differences between the constituting and constituted objects are only skin-deep; they are context-dependent rather than observer-independent features of the world. Second, monists can claim that their view is simpler and that the pluralist view fills reality with layer upon layer of excess ontological fat (Bennett 2004). We can adopt a leaner and meaner view by identifying the constituting and constituted objects, just as a picture is just an arrangement of pixels on paper (Lewis 1994). More generally, we can view the apparent incompatibilities of properties as simply the consequence of incompatibilities in perspective or description.

3. Towards a Fregean-Monist solution to the puzzle
of material constitution

Perhaps a useful first step to navigate a way through the tortuous puzzle of material constitution is to spell out the agreements between the monists and the pluralists. First, both sides agree that the constituting and constituted objects are materially and spatially coincident but disagree on whether they are identical. Second, both sides agree that the puzzle of material constitution has a semantic aspect. While most monists claim that any differences in essence or other properties between the constituting and constituted objects are only apparent, or are differences in description based on different contexts, most pluralists claim that the choice to include objects into or exclude objects from our ontological scheme is based on whether certain sortal terms exist in ordinary language. Perhaps an adequate solution to the puzzle of material constitution is in part semantic and in part metaphysical.

The next step is to highlight the most plausible elements of both monist and pluralist positions. Even if we reject the monists' claim that the property differences between the constituting and constituted objects are only skin-deep or context-dependent, we need not reject their aim to trim excess ontological fat off the pluralists' account. That is, the monist account has the principle of parsimony on its side as it has the simpler explanation of material constitution involving the fewest entities, provided that its explanatory power is not inferior to that of the pluralist account. And even if we reject the pluralists' claim that there are two or more objects occupying the same place at the same time for the sake of parsimony, we need not reject their aim to account for property differences. That is, the pluralist account has the virtue of saving appearances on its side as it aims to explain the way things seem like to us, what Lynn Rudder Baker calls 'a metaphysics of ordinary things' (Baker 2008).

Keeping the first two steps in mind, the third step is to define the concepts of object and property since they are key to resolving this puzzle. A definition offered by C.B. Martin and John Heil, which I endorse, is as follows: 'Objects can have parts, but an object's properties are not its parts, they are the particular ways the object is' (Martin and Heil 1999, 45). When

material parts are arranged, organised, or structured in certain ways (in the static sense), or when material parts connect, interact or enter into processes with one another in certain ways (in the dynamic sense), then the object constituted by these material parts bear certain properties. Objects can be treated as concrete particulars while properties can be treated as abstract particulars, distinct from and yet grounded on objects or concrete particulars. Different terms can be used to represent a single object or concrete particular in different ways, including or excluding certain properties or abstract particulars. For example, the term 'Lump' can be used to represent a bronze object, or a chunk of bronze parts arranged in a certain way occupying a certain place at a certain time, including some of its the properties (weight and size) but excluding others (statue-shape and aesthetic appeal). And the term 'David' can be used to represent the same bronze object, or the same chunk of bronze parts arranged in the same way, including more of its properties than 'Lump' (weight, size, statue-shape, and aesthetic appeal).

I shall call this view Fregean Monism (FM) since it is largely based on Frege's famous distinction between reference (extension) and sense (intension) with a few modifications. FM is based on the following claims:

  i.   Terms refer to objects and their properties.

 ii.   Terms have reference (extension) and sense (intension). The reference (extension) of a term picks out an object and its properties; the sense (intension) of the term determine which properties of the object are picked out.

iii.   Referent objects are objects picked out by their corresponding terms; referent properties are properties picked out by their corresponding terms.

 iv.   Referent objects are concrete particulars (arrangements of material parts); referent properties are abstract particulars (ways arrangements of material parts are).

  v.   Two or more terms can have the same reference or referent object.

 vi.   Even when two or more terms have the same reference (pick out the same referent object), they can have different senses (pick out different referent properties of that referent object). For example, 'David' picks

out properties including weight, size, statue-shape and aesthetic appeal whereas 'Lump' picks out properties such as weight and size but not statue-shape and aesthetic appeal, even when both terms refer to the same object—the bronze piece.

On this view, semantic pluralism (the linguistic component of the view) explains property pluralism (the metaphysical component of the view). While 'David' and 'Lump' refer to the same object or concrete particular, the properties or abstract particulars picked out by 'David' and 'Lump' are different because of their different senses. While the senses of both 'David' and 'Lump' pick out the weight and size of the object, the sense of 'David' but not 'Lump' picks out its statue-shape and aesthetic appeal. Unlike other versions of monism, FM explains the monists' intuition that materially and spatially coincident objects are identical without explaining away property differences as mere differences in description, perspective, or context. And unlike pluralism, FM explains the pluralists' intuition about property pluralism without appealing to object pluralism.

To explain property pluralism, FM posits two kinds of reference—referent objects and referent properties. While reference (extension) picks out both the referent object and its referent properties and accounts for the metaphysics, sense (intension) determines which referent properties of the referent object are picked out and accounts for the semantics. This conceptual distinction provides a solution to the puzzle of material constitution, which could have arisen because terms do not pick out all of the referent properties of a referent object. And as terms do not pick out all of the referent properties of a referent object, different terms represent the referent object in a different and partial way by picking out some of its referent properties but not others. As such, FM is not only a semantic solution, but also a metaphysical one. It does not explain away differences in properties as mere differences in description, perspective, or context. Instead, it explains why a referent object can seem to have different referent properties by showing how different terms pick out some but not all of the referent properties possessed by the referent object.

4. Pluralist arguments against monism
and Fregean-Monist responses

The first argument pluralists employ against monists is the argument from Leibniz's Law (Fine 2003; Frances 2006). According to this argument, David and Lump cannot be one and the same object because there is at least one difference between them, and by holding that David and Lump are one and the same object despite their property differences, monism violates Leibniz's Law. Yet, Leibniz's Law is not violated when the property differences are accounted for by different terms picking out the different properties of one and the same object, as terms referring to one and the same object pick out some but not all of its properties. An object is called 'David' when its aesthetic appeal is picked out and called 'Lump' when its aesthetic appeal is not picked out. Like other monists, proponents of FM assume that there is only a single object, a piece of bronze, occupying a place at a time, being referred to as 'David' or 'Lump.' But unlike other monists, proponents of FM tell a deeper story by emphasising that 'David' and 'Lump' have the same referent object but different referent properties. And the senses of 'David' and 'Lump' determine which referent properties are picked out. FM shows how the semantic difference explains the ontological difference. On this view, the aesthetic appeal of the piece of bronze is not merely a way of describing the piece of bronze within a certain context (a predicate); it also picks out the way the piece of bronze is under certain circumstances (a property). In other words, aesthetic appeal is possessed by the piece of bronze but it is picked out by the sense of 'David' but not the sense of 'Lump.' Thus, FM does not violate Leibniz's Law by explaining property differences in terms of different terms picking out the different properties of one and the same object, as terms referring to one and the same object pick out some but not all of its properties.

The second related argument pluralists employ against the monists is the argument from different persistent conditions between the constituting and constituted objects (Baker 1997, 2000). On the pluralist view, David ceases to exist as the constituted object should it lose its essential/fundamental property—aesthetic appeal—by losing its statue-shape while Lump persists as the constituting object. Yet, everything else about David remains

intact other than its aesthetic appeal. Why, then, hold that David as an object or concrete particular, ceases to exist? On other versions of monism, 'David' is just a different description of 'Lump' from another perspective or in another context. In other words, the difference between the constituting and constituted objects is merely semantic. However, there is also an ontological difference between them, for what cease to exist are properties or abstract particulars (aesthetic appeal), even if the object or concrete particular (bronze object) persists. On the FM view, different terms designating the same referent object pick out different referent properties. When a certain referent property (aesthetic appeal) of a referent object (bronze object) no longer exists, the term ('David') that picks out the lost referent property (aesthetic appeal) no longer applies. To borrow Aristotle's examples, an eye remains numerically the same eye after losing its sight and an axe remains numerically the same axe after losing its sharpness. At most, it can be argued that the term 'axe' that picks out the lost referent property, sharpness, is no longer applicable to the now blunt axe-shaped object, or that the term 'eye' that picks out the lost referent property, sight, is no longer applicable to the now blind eye-shaped object. And perhaps other terms can be used to pick out the blunt axe-shaped object or the blind eye-shaped object. But it is another thing to argue that an axe, as an object, ceases to exist when it becomes blunt or an eye, as an object, ceases to exist when it becomes blind, when the material parts arranged axe-wise or eye-wise respectively, still occupy the same place at the same time. Hence, the appearance of two materially and spatially coincident objects with different persistence conditions can be explained as two different terms picking out the same referent object but different referent properties.

The two above pluralist arguments can be re-expressed as a reductio argument against monism. According to this argument, holding that the constituting and constituted objects are a single object implies that it both possesses and lacks a property such as aesthetic appeal, or that it both exists and not exists at the same time, and hence results in a contradiction. This argument works only on the assumption that property pluralism implies object pluralism, an assumption challenged by the proponents of FM. On the FM view, the constituting object (Lump) and the constituted object (David) are not two different referent objects but two different terms

('Lump' and 'David') picking out different referent properties of the same referent object (bronze object). The bronze object is called 'David' when aesthetic appeal is present and is called 'Lump' when aesthetic appeal is absent. And the bronze object can gain aesthetic appeal and be called 'David' or lose aesthetic appeal and be called 'Lump.' It is not the case that a new object (David) comes into existence and coincides with an existing object (Lump) after gaining a new property (aesthetic appeal) and goes out of existence and no longer coincides with the existing object (Lump) after losing the new property (aesthetic appeal). In other words, it is not the case that some properties possessed by Lump are not possessed by David and vice versa. On the contrary, it is the case that David and Lump are a single object referred to by two different terms with the same extension but different intensions that pick out some but not all the referent properties. Hence, there is no contradiction in holding that the constituting and constituted objects are a single object and the reductio argument does not succeed.

The third argument pluralists employ against the monists is the argument from primary kinds. Originating from Aristotle, an object's primary kind defines what it is essentially or fundamentally (Baker 2002, 2008). For example, a statue (David) is a different kind of object from a piece of bronze (Lump) as the former is essentially an art object whereas the latter is essentially a piece of raw material. So, how many objects there are in the same place at the same time depends on how many kinds of objects there are. Monists object that the notion of primary kinds is not without problems. First, postulating how many kinds of materially and spatially coinciding objects there are remains arbitrary as it is contingent upon our epistemic interests and linguistic conventions. Second, the correspondence between the sortal terms in our language and the kinds of objects in the world requires further explanation and the decision to include some kinds and exclude others requires further justification (Wilson 2007). For these reasons, FM treats primary kinds as conceptual categories rather than ontological categories, like other forms of monism. Unlike other forms of monism, however, FM shows how the semantic pluralism of primary kind terms account for their property pluralism without resorting to object pluralism. On the FM view, different primary kind terms are employed to look at

a single object in different ways, including or excluding certain properties, rather than to pick out distinct materially and spatially coincident objects because of differing properties. That is, different primary kind terms, with different senses, pick out different properties of a single object instead of different objects. Categorising a piece of bronze as a statue kind (David) is seeing it inclusive of its aesthetic appeal and categorising the same piece of bronze as a raw material kind (Lump) is seeing it exclusive of its aesthetic appeal. Thus, primary kinds need not be ontological categories that pick out different objects because they do not share all of the same properties; they can be conceptual categories that pick out different properties of a single object.

On the FM view, objects are concrete particulars, consisting of only their material parts and arrangement, implying that only one object can occupy one place at a time. This definition rules out the pluralist claim that more than one object of different primary kinds can be in the same place at the same time, with one object belonging to a primary kind constituting a second object belonging to another primary kind, as materially and spatially coinciding objects of different kinds cannot count as more than one concrete particular. To claim that materially and spatially coincident objects are distinct objects, pluralists need to deny that all materially and spatially coincident objects are concrete particulars. There are at least two ways pluralists can do so. The first way is to claim that some of these objects are abstract and that concrete and abstract objects can be materially and spatially coincident. Yet, the distinction between abstract and concrete objects is usually taken to mean the type-token or universal-particular distinction, and pluralists would not want to claim that a piece of bronze is a token-particular of a statue type-universal. The second way is to claim that some objects are immaterial and that material and immaterial objects can be materially and spatially coincident. Yet, pluralists would not want to claim that statues are immaterial either. So, if objects are concrete particulars that consist only of their material parts and arrangements, then not more than one object can occupy a place at a time. And if not more than one object can occupy a place at a time, then it is more plausible to treat property differences between constituting and constituted objects as different properties of a single object picked out by different terms with different senses.

5. A closer look at the argument from Leibniz's Law

In the earlier section, I have argued that FM does not violate Leibniz's Law. In this section, I want to show that Leibniz's Law does not necessarily support object pluralism. Let us look at a version of Leibniz's Law (adapted from King 2006 and Smid 2017) more thoroughly:

Premise (1)   F(t)   ⟶   (Property Pluralism)
Premise (2)   ¬F(s)

Conclusion (1)   ∴ t ≠ s   ⟶   (Semantic Pluralism)
Premise (3)   t → y & s → x

Conclusion (2)   ∴ y ≠ x   ⟶   (Object Pluralism)

where 'F' is a predicate expressing a property, 's' and 't' are terms (names and definite descriptions) expressing objects, 'x' and 'y' denote objects, and '→' denotes 'picks out/refers to.'

Applying this to the case of David and Lump, where F = aesthetic appeal, t = 'David' and s = 'Lump,' y = David the bronze statue, x = Lump the piece of bronze, we have:

Premise (1a)   'David' possesses the intension/sense of aesthetic appeal.
Premise (2a)   'Lump' lacks the intension/sense of aesthetic appeal.

Conclusion (1a) Therefore, 'David' is not identical to 'Lump'.
Premise (3a)   'David' has David (the bronze statue) as its extension/referent and 'Lump' has Lump (the piece of bronze) as its extension/referent.

Conclusion (2a) Therefore, David (the bronze statue) is not identical to Lump (the piece of bronze).

Proponents of FM affirm premises (1) and (2) and conclusion (1). They affirm that term t has the intension/sense of property F whereas term s does

not, and hence terms t and s are not semantically identical terms. At this stage, whether t and s refer to one or two objects remains undetermined. In other words, only semantic pluralism, but not object pluralism, is ascertained. Proponents of FM also deny premise (3) and conclusion (2). They deny object pluralism, held by the pluralists, that the terms t and s pick out distinct objects y and x respectively. Does the denial of object pluralism violate Leibniz's Law? I contend that it does not, because Leibniz's Law says nothing about whether premise (3), the claim that the terms t and s pick out distinct objects y and x respectively, holds. And I further contend that since Leibniz's Law remains silent on whether premise (3) is true, pluralists cannot rely on premise (3) to derive conclusion (2). Therefore, Leibniz's Law does not necessarily support object pluralism.

Since pluralists cannot appeal to Leibniz's Law to justify premise (3) without begging the question, other independent arguments are required. That is, pluralists have to show that objects y and x are the referents picked out by the terms t and s respectively before they can conclude that y is not identical to x. Otherwise, premise (3) remains a contested assumption or mere stipulation. Perhaps the strongest pluralist argument for premise (3) is the argument from primary kinds, the idea that while objects of the same kind cannot occupy the same place at the same time, objects of different kinds can. Yet, as mentioned above, the argument from primary kinds remains problematic. First, postulating how many kinds of materially and spatially coinciding objects there are still seems arbitrary as it is contingent upon our epistemic interests and linguistic conventions, Second, the correspondence between the sortal terms in our language and the kinds of objects in the world requires further explanation and the decision to include some kinds and exclude others requires further justification. Hence, since the pluralists' appeal to the argument from primary kinds to support premise (3) remains contested, and premise (3) is required for conclusion (2), then the pluralists' appeal to Leibniz's Law to support conclusion (2) remains contested.

On the contrary, FM offers a plausible argument for denying the contested premise (3) and conclusion (2). First, objects, such as x and y, are concrete particulars and since concrete particulars are defined only by their material parts, arrangement, and location and x and y share exactly the

same material parts, arrangement, and location, then x and y are numeri-cally identical; they are a single object. Despite affirming the material and spatial coincidence of x and y, pluralists insist that they are two objects instead of one because of the property differences between them. In con-trast, FM explains the property differences x and y without resorting to an object difference. Second, terms s and t need not pick out two distinct ref-erent objects y and x. Instead, they may pick out the same referent object, say z, but pick out different referent properties of z. For example, t may pick out referent property F but s may not, even when s and t pick out the same referent object z. That is, t and s have the same extension/referent but different intension/sense: the term F(t) represents object z with prop-erty F and the term ¬F(s) represents object z without property F. While the pluralist view holds that object pluralism is required to account for property pluralism, FM claims that semantic pluralism can account for property pluralism—an ontological difference—without object pluralism. Third, FM is not claiming that z both possesses and lacks F, but that z can be represented in two ways, with or without F, by using the terms t and s respectively. It treats primary kinds as conceptual instead of ontological categories. Classified under one kind, z can be represented as F(t) and clas-sified under another kind, z can be represented as ¬F(s). On the FM view, only one concrete particular, or a particular arrangement of material parts, occupies a place at a time, represented in different ways using different terms, or classified under different kinds, including or excluding certain ab-stract particulars.

Applying this analysis to the David and Lump case, 'David' and 'Lump' are two names for a single object—a particular arrangement of bronze parts. The term 'David' is used to denote it when it is represented as an art object (falls under the statue kind/category), with its aesthetic appeal; the term 'Lump' is used to denote it when it is represented as a piece of raw material (falls under the raw material kind/category), without its aesthetic appeal. When this particular arrangement of bronze parts retains its aesthetic ap-peal, it can be denoted by both 'David' and 'Lump' as it falls under both the statue and raw material kinds/categories. But when it loses its aesthetic appeal for some reason, then it can only be denoted by 'Lump' but not 'David' as it no longer falls under the statue kind/category even if it still

falls under the raw material kind/category. There is only one object all along, denoted by different terms picking out different properties. So, on the FM view, David and Lump are numerically identical; they are a single bronze object capable of bearing all the properties ascribed to both David and Lump by the pluralists at certain times but not others, and is called 'David' or 'Lump' accordingly, depending on whether it possesses/gains or lacks/loses aesthetic appeal.

Pluralists may object that FM, like other versions of monism, changes the puzzle of material constitution from an ontological problem to a mere semantic problem, or reducing 'a metaphysics of everyday things' to merely 'a description of everyday things.' In the above analysis of Leibniz's Law, for example, proponents of FM seem to change the discussion from one about objects (x and y) and properties (F) to one about subjects and predicates (F(t)) and ($\neg$F(s)). This is not the case, however. Proponents of FM apply a semantic tool, the distinction between the sense and reference of terms, to clarify, not dismiss, the ontological part of the puzzle. While the reference of a term (t or s) picks out an object (z) and its properties (F), the sense of a term determines which of the object's (z's) properties are picked out and which are not. Both terms, t and s, pick out object z, but whereas F(t) pick out property F, $\neg$F(s) does not. Again, assuming that terms 'David' and 'Lump' refer to a single bronze object, 'David' represents the bronze object (inclusive of its aesthetic appeal) as an art object whereas 'Lump' represents the same bronze object (exclusive of its aesthetic appeal) as a piece of raw material. Proponents of FM treat referent properties or abstract particulars as ways referent objects or concrete particulars are, and not merely ways referent objects or concrete particulars are described or represented. In other words, claims about referent properties are claims about ways referent objects are. They are ontological claims, not merely semantic ones. As proponents of FM employ semantic tools to clarify rather than dismiss ontological claims, the objection that FM changes the puzzle of material constitution from an ontological problem to a mere semantic problem is unwarranted.

Moreover, pluralists may object that proponents of FM have not clearly shown that the terms t and s refer to a single object, z and not two objects, x and y. Yet, the various arguments presented above, in tandem, offer good

reasons in favour of the claim that t and s refer to a single object. First, the claim that—if x and y are materially and spatially coincident then they are identical objects—remains a truism despite being challenged. As argued above, the challenges presented by both the argument from Leibniz's Law and the argument from different primary kinds remain inconclusive. Second, the claim that—if x and y do not share property F, then they cannot be identical objects—remains questionable. As argued above, whether object z possesses/gains or lacks/loses property F does not imply that it must therefore be two distinct objects x and y, it may just be two different ways of presenting object z, with or without property F, using terms t and s respectively. Third, the monist claim that t and s refer to a single object has the principle of parsimony on its side so long as it explains appearances as well as the pluralist claim that t and s refer to two distinct objects. As argued above, FM does not explain away the qualitative difference between x and y as simply differences in description, perspective, or context. Rather, it tells a deeper story about how the qualitative difference is dependent on whether referent property F is picked out by terms t or s. Therefore, the reasons in favour of the monist claim seem stronger than those in favour of the pluralist one.

6. Conclusion

The puzzle of material constitution is in part ontological and in part semantic and an ideal solution to the puzzle of material constitution should address both ontological and semantic issues and possess the twin virtues of parsimony and saving appearances. FM is one such attempt. First, by holding that objects are concrete particulars or arrangements of material parts and that only one concrete particular or arrangement of material parts can occupy a place at a time, FM preserves the virtue of parsimony, unlike pluralism. Second, by distinguishing between properties as ways objects are (ontological category) and terms as ways objects are described or represented (semantic category), FM does not explain away property differences as mere differences in description, perspectives, and context, but explain the property differences by applying a semantic tool, the distinction be-

tween the sense and reference of terms, to clarify, not dismiss, the ontological part of the puzzle, unlike other versions of monism. Third, by maintaining that the reference (extension) of a term picks out an object and its properties, and that the sense (intension) of the term determines which properties of the object are picked out, FM explains why a referent object can seem to have different referent properties by showing how different terms pick out some but not all of the referent properties possessed by the referent object, thus preserving the virtue of saving appearances, unlike other versions of monism. Fourth, in accepting property pluralism and rejecting object pluralism, FM does not violate Leibniz's Law. And as Leibniz's Law remains silent on whether two semantically distinct terms pick out two objects or a single object (partially) represented in different ways, it does not necessarily support object pluralism. Fifth, to support a metaphysics of ordinary things, it seems sufficient to have different terms picking out different referent properties of a single referent object occupying a place at a time rather than to have more than one materially and spatially coinciding referent objects. These points make FM worth-considering as a solution to the puzzle of material constitution.

References

Baker, Lynne Rudder. 1997. "Why Constitution is Not Identity." *The Journal of Philosophy* 94 (12): 599–621.

Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View.* Cambridge: Cambridge University Press.

Baker, Lynne Rudder. 2002. "On Making Things Up: Constitution and its Critics." *Philosophical Topics* 30 (1): 31–51. https://doi.org/10.5840/philtopics200230112

Baker, Lynne Rudder. 2008. "A Metaphysics of Ordinary Things and Why We Need It." *Philosophy* 81 (1): 5–24. https://doi.org/10.1017/S0031819108000284

Bennett, Karen. 2004. "Spatio-Temporal Coincidence and the Grounding Problem." *Philosophical Studies* 118 (3): 339–71. https://doi.org/10.1023/B:PHIL.0000026471.20355.54

Fine, Kit. 2003. "The Non-Identity of a Material Thing and its Matter." *Mind* 112 (446): 195–234. https://doi.org/10.1093/mind/112.446.195

Frances, Bryan. 2006. "The New Leibniz's Law Arguments for Pluralism." *Mind* 115 (460): 1007–1021. https://doi.org/10.1093/mind/fzl1007

Frege, Gottlob. 1892. "On Sense and Reference." ["Über Sinn und Bedeutung".] *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50.

Gibbard, Allan. 1975. "Contingent Identity." *Journal of Philosophical Logic* 4 (2): 187–221. https://doi.org/10.1007/BF00693273

King, Jeffrey. 2006. "Semantics for Monists." *Mind* 115 (460): 1023–58. https://doi.org/10.1093/mind/fzl1023

Lewis, David. 1999. "Reduction of Mind." In *Papers in Metaphysics and Epistemology*, edited by David Lewis, 291–324. Cambridge: Cambridge University Press.

Martin, C.B. and Heil, John. 1999. "The Ontological Turn." *Midwest Studies in Philosophy* 23 (1): 34–60. https://doi.org/10.1111/1475-4975.00003

Paul, L.A. 2010. "The Puzzles of Material Constitution." *Philosophy Compass* 5 (7): 579–90. https://doi.org/10.1111/j.1747-9991.2010.00302.x

Smid, Jeroen. 2017. "Material Constitution is Ad Hoc." *Erkenntnis* 82 (2): 305–25. https://doi.org/10.1007/s10670-016-9818-6

Wasserman, Ryan. 2017. "Material Constitution." *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), edited by Edward N. Zalta. Last updated July 5. 2017. https://plato.stanford.edu/archives/fall2018/entries/material-constitution

Wiggins, David. 1968. "On Being at the Same Place at the Same Time." *The Philosophical Review* 77 (1): 90–95. http://doi.org/10.2307/2183184

Wilson, Robert. 2007. "A Puzzle about Material Constitution & How to Solve it: Enriching Constitution Views in Metaphysics." *Philosophers' Imprint* 7 (5): 1–20.

Zalta, Edward N. 2019. "Gottlob Frege." *The Stanford Encyclopedia of Philosophy*, (Fall 2020 Edition), edited by Edward N. Zalta. Last updated Dec 5. 2019. https://plato.stanford.edu/archives/fall2020/entries/frege

RESEARCH ARTICLE

Subject and Predicate in Existential Propositions: A Survey of Frege's Problem and Its Solutions

Siavash Asadi*

Abstract: Frege argues that considering Socrates as an object in the proposition "Socrates exists" raises two problems. First, this proposition would be uninformative. Second, its negation entails a contradiction. Attempting to solve these problems, Frege claims that Socrates is representing the *concept* of a man whose name is Socrates. Therefore, existence is a second-order concept. This paper surveys the main modern theories about the types of existence, in order to find another response to Frege's problems. For, if Socrates' existence differs from the type that "exists" implies, "Socrates exists" is informative and its negation is not a contradiction. At last, this paper argues for an idea, in which "existence" is not a concept or property. Existence is the principle of the objects. So, "Socrates exists" is in fact "*the* existence is Socrates," and "Socrates does not exist" is "there is no existence that be Socrates." This idea could be an alternative for responding to Frege's problems.

Keywords: Existential proposition; existence; Frege; object; concept.

* Duquesne University
  ✎ 601 Fisher Hall, Duquesne University, 600 Forbes Ave, Pittsburgh, Pennsylvania, USA
  ✉ asadis@duq.edu

1. Introduction

During the history of philosophy, the meaning of "existence" has always been the subject of most ambiguous philosophical debates. From the era of ancient Greek philosophers to the present days, accomplished philosophers have clearly acknowledged this obscurity. As Aristotle emphasizes in *Metaphysics*: "the question which was raised long ago, is still and always will be, and which always baffles us—'What is Being?'..." (7, 1028b, 1). And, Williamson (1988) believes, "Both actualism and anti-actualism are obscure doctrines, for the crucial term 'exist' is ambiguous" (Williamson 1988, 259).

From the late nineteenth century, however, Gottlob Frege made the ontological debates, at least in analytic tradition, more complicated when he published his idea about "existence" as a second-order predicate. His opinion drew the term "existence" into a new phase of difficulty, and added unprecedented problematic challenges to all previous quandaries. The epitome of Frege's idea is that in an existential proposition,[1] the subject's referent is not a specific object, but it is a concept. Indeed, such a proposition is expressing that the concept—which is the de facto subject's referent— has, at least, one factual extension. Frege surprisingly demonstrates "Socrates (a specific object) exists" as neither true nor false, but a meaningless proposition. Instead, he presents "the man whose name is Socrates (a concept) exists (has an extension)" as a meaningful proposition, which could be true or false. This deconstructing idea about existential propositions was the opposite of the philosophers' opinions up to that day and the common understanding of this kind of proposition. Accordingly, some philosophers were stimulated to look Frege's problematic idea through, finding detailed interpretations about the entity of the subject's (a proper name's) referent.

---

[1]   In this paper, an "existential proposition" is a proposition in the "x exists" format, in which "x" is a proper name of a specific object, such as Socrates, Eiffel Tower, etc. Frege typically applies the term "thought" instead of "proposition." Anyway, of both "proposition" and "thought," we mean the content of a sentence. So, two sentences might be different, while expressing a single proposition or thought, such as "John is Anna's brother" and "Anna is John's sister." See (Frege 1960b, 49).

2. Frege's claim

Consider "Socrates was an accomplished philosopher." In this proposition, "Socrates," that is grammatically the subject, has a sense,[2] and a referent which is a person called Socrates in the realm of spatiotemporal objects. More importantly, this term could be saturated or unsaturated, or in a mathematical terminology, could be an independent variable (argument) or a function. "Socrates," in the above example, is saturated or is an argument, but "was an accomplished philosopher" is unsaturated or is a function, because it must follow a subject to become complete (Frege 1960c, 31). Frege calls saturated referents "objects," and unsaturated referents "concepts." The objects are normally the referents of proper names, while concepts are typically the referents of concept-words (Frege 1960b). For instance, human is a concept and the unsaturated referent of the concept-word "human."

Concentrating on predicative propositions, Frege indicates that the division on subject and predicate is logically unimportant. Therefore, he replaced these notions with argument/function distinction in which the referent of an argument could be an object or a concept while a function's referent should be a concept. In "Socrates was an accomplished philosopher," the subject refers to an object, while in "human is an animal," the subject is a concept-word and its referent is a concept. According to Frege, if a concept describes an object, such as "being an accomplished philosopher" in the former example, it is called a first-order concept. But, if a concept belongs to another concept, such as "being an animal" in the latter example, it is called a second-order or second-level concept. More precisely, Socrates falls under the concept of "being an accomplished philosopher." But human is not under the concept of "being an animal," it has a relation to "being an animal," or belongs to it. Frege demonstrates this meaning with this quote: "To do Justice at once to the distinction and to the similarity, we might perhaps say: An object falls *under* a first-level concept; a concept falls *within* a second-level concept" (Frege 1960b, 50–51).

---

[2]   Frege believes in Indirect Reference Theory. According to this theory, a proper name has a sense in addition to its referent.

Back to the matter at hand, Frege argues that, in existential proposi-tions, the predicate "exists" could not describe an object. It must predicate to a concept, and so it should be a second-order concept; or, as Frege himself says: "I have called existence a property of a concept" (Frege 1960 b, 48). It means that in "Socrates exists," "Socrates" is a concept-word and its referent is the concept of Socrates, not Socrates as an object. There is an analogy between existential and numerical predicates in Frege's works to illustrate this meaning (Frege 1960b, 53; 1953, 65). In Frege's view, "the solar system has 9 planets" does not mean the solar system (an object) is under the concept of "having 9 planets." But, this proposition connects the number 9 to the concept of "planets of the solar system." So, the number 9 is a second-order predicate. Similarly, "Socrates exists" connects the exist-ence to the concept of Socrates, and says that this concept is not empty.

As far as we can see, Frege has not given a clear argument for this claim. But, as Mendelsohn has pointed out (Mendelsohn 2005, 102), Frege's argu-ment could be concluded from his posthumous dialogue with his colleague and friend, *Punjer*. According to this document, Frege argues that if a sub-ject's referent of an existential proposition was an object, this proposition would not be informative. For, "Socrates exists" could be interpreted in two similar ways. First, the concept of Socrates has an extension, and second, there is at least one x that is identical with Socrates, or Socrates is identical with himself. And, "Neither in 'A is identical with itself,' nor in 'A exists' does one learn anything new about A" (Frege 1979, 62). In symbolic lan-guage, both interpretations could be shown as $(\exists x)(x=Socrates)$. So, we could find no more information in this proposition than this linguistic iden-tification. In other words, considering "Socrates exists," we know we are talking about an existent Socrates, otherwise we can say nothing about it. Because, as Williamson demonstrates (Williamson 1999, 260), we cannot inform of absolute nonexistence. Thus, "Socrates exists" is equal to "the existent Socrates exists," which is an uninformative proposition.

More importantly, the negation of an existential proposition is a self-contradictory proposition. "Hank (a Keebler elf)[3] does not exist" or "Hank exists not" is equal to "the existent Hank, does not exist," and it is an

---

[3]   This example has been used by Bennett in her 2006 paper (p. 270).

obvious contradiction (Frege 1979, 59). On this basis, Frege concludes that an existential proposition would be neither true nor false, but senseless, if its subject's referent is considered as an object (Frege 1960b, 50). Rejecting the objectivity of the subject's referent, Frege's only option is a conceptual subject's referent. This means that "exists" is a second-order predicate.

I emphasize the logical contradiction in "Hank does not exist," because only being uninformative—as with positive propositions—does not necessarily lead Frege to call "Socrates exists" senseless.[4] As David Londy says, considering "not to exist" a property of objects, results in ridiculous conclusions: suppose a shepherd, who is looking for his nonexistent sheep, as well as those which exist (Miller 1975). Also, Ayer confirms the contradiction in "Hank does not exist," and, following Frege, introduces Hank as a concept.[5] He indicates the grammatical similarity between existential and descriptive sentences as the source of a logical ambiguity. "Hank does not exist" and "Hank does not lie" have the same grammatical structure, but it does not follow that they also have logical similarity (Ayer 1949, 24–26).

3. Various approaches to existence

It seems that considering the subject as a concept is an acceptable solution for Frege's problem among many philosophers after him. With this innovation, an existential proposition is considered informative because in "Socrates exists" the audience is addressed that the concept of "a man whose name is Socrates" is not empty. Also, in "Hank does not exist" there is no contradiction, because it is expressing that the concept of "a Keebler elf whose name is Hank" is empty. Apart from Frege's solution, we attempt to study the possible another solution concentrated on various approaches to existence.

Here, the main question is "is there only one kind of existence?" For, a straight way to respond to Frege's problem is that the subject's type of

---

[4]   As far as we have searched, there is no attributing senseless to the other analytic propositions, such as "each husband is married" in Frege's works.

[5]   In addition to Ayer, most of other positivists have accepted Frege's idea about 'existence,' which is not an experienced object.

existence would differ from the type of the predicate "exists." In a part of his dialogue with *Punjer*, Frege says: "if you are using the word 'exists' in the same sense as the expression 'there is,' then you have at the same time both asserted and denied the same predicate of the same subject" (Frege 1979, 59). Thus, if the type of Socrates' existence differs from the type of "exists," "Socrates exists" would be informative, because we predicate a concept to Socrates that he did not have before. Also, "Hank does not exist" has no contradiction if Hank's existence differs from "exist."[6]

Plantinga, for instance, distinguishes two kinds of existence: the existence of individual essences in possible worlds,[7] and the existence of individual essences in the actual world (Plantinga 2003, 48–49). An individual essence, Plantinga says, is a property that can be exemplified by only one particular individual, such as Socrates. It exists in possible worlds, but it would be actual or exemplified only in the actual world (Plantinga 2003, 199–203; 1974, 45–63).

---

[6]   Today, the various kinds of existence could be discussed under the actualism category. After Kripke's paper (1959) titled, "A Completeness Theorem in Modal Logic," which was based on possibilism, some philosophers attempted to harmonize Simplest Quantified Modal Logic with the principal proposition of actualism: "Everything that exists, exists in the actual world." Some actualists might share with Frege this idea that existence is a second-order concept, and some might not. In this paper, however, sharing this idea with Frege, comparing actualism and possibilism, and studying their strengths and weaknesses are not our task. What we want to do is employ their achievements to respond to this question: does "Hank does not exist" include a necessary contradiction? So, we will use the main important theories to reach our purpose, without discussing under which of these doctrines a theory could be placed. Concentrating on our central problem, we avoid engaging in unrelated debates, and simplifying our paper, we avoid using symbolic language as well as possible.

[7]   The idea of possible worlds emerged in Leibniz's works for the first time. He believes that among uncountable worlds that could be created, God has created our actual world as the best. See (Leibniz 1988, 416). As Kripke says, we could apply other phrases, such as counterfactual situations or possible states, instead of possible worlds. For example, a possible world could be a world, in which Einstein was not the founder of Relativity Theory. Nevertheless, some like Lewis emphasize that possible worlds exist as well as the actual world. See (Lewis 2001, 84). For criticism to Lewis and helpful information about possible worlds, see (Stalnaker 1976).

On this basis, Socrates is an individual essence—we do not know whether he is actual. Expressing "Socrates exists," we inform that Socrates is an actual existent. Therefore, it could be an informative proposition. In the same way, "Hank does not exist" means that the individual essence of Hank—which belongs to a possible world—is not exemplified and does not belong to the actual world. Therefore, the contradiction expressed by Frege is not a true one. Based on Plantinga's view, we are not saying "the existent Hank does not exist," but we are saying "the individual essence of Hank is unexemplified in the actual world." So, in his work, Plantinga (1974) has said many times that there is a big difference between claiming that "Hank has the property of nonexistence" and claiming that "Hank does not exist."

Plantinga's view, however, has been criticized by the other philosophers,[8] especially Zalta and Linsky. The central challenge in Plantinga's view is: how can individual essences exist while they are not actual? What does existent mean without actuality? Indeed, this idea could fall in a chain of circular explanations. Zalta and Linsky, in their common paper, note that Plantinga's attempt faces difficulties, and say:

> The Problem is this: an essence such as *being Reagan* could exist at a world where Reagan doesn't exist only if it is purely qualitative (i.e., doesn't involve Reagan as a constituent). But a purely qualitative property could be exemplified by different objects at different worlds, violating clause of the definition of an essence. So, essences seem to require a non-qualitative component. But if so, then if the non-qualitative component is all that there is to an essence, the essence can no longer be seen as a property, for such non-qualitative, non-repeatable entities are not distinguishable from possible objects. If the non-qualitative component is just a part of the essence, then what else could such a component be but the contingent object itself? But then essences would ontologically depend on contingent objects. Thus, they could not exist unexemplified. (Zalta and Linsky 1994, 448)

In other words, if an essence such as "being Reagan" ontologically depends on Reagan himself, then if Reagan had not existed, the essence "being

---

8   For a brief review of criticism of Plantinga's view, see (Forbes 1987).

Reagan" would not have existed, and so, essences could not exist unexemplified. Therefore, individual essences should be considered as nonexistent. In this situation, Plantinga must demonstrate how we can talk about nonexistence; this is impossible based on Plantinga's philosophical view.

Zalta and Linsky say this argument could be applied against any other theory, such as Fine's theory,[9] that entails contingent constituents of the worlds, in which the constituents do not exist.

Instead, Zalta and Linsky propose a new classification of existents, asserting that all existents are actual. In their view, the domain of existents contains abstract and concrete objects. Abstract existents, such as numbers and propositions, and concrete existents, such as Zalta, are the two extremes of the realm of existence in all possible worlds. But, there are many things that are not under these two categories. Zalta and Linsky call these things "possibly concrete objects." These objects come in two sorts: "contingently concrete" and "contingently nonconcrete" objects (Zalta and Linsky 1994, 432). According to Tomberlin's explanation, the former are just the ordinary concrete objects that exist in the actual world; here, they are concrete, although they fail to be so in other worlds. With the latter sort, however, these are individuals obeying a pregnant condition: they are nonconcrete in the actual world, concrete in other worlds, and yet they actually exist in our world (Tomberlin 1996, 274). For example, Hank is a nonconcrete existent in our world. But, he could be a concrete existent in a possible world. Based on this theory, the proposition "it is possible that Hank is a Keebler elf" entails "there is Hank that is possibly a Keebler elf" (Zalta and Linsky 1996, 283–86).[10]

Zalta and Linsky emphasize that the domain of objects (existents) among all possible worlds is the same. If x_1, x_2, … are all things in the world W_1, they are also all things in the world W_2. But, it could be that x_1 is concrete in W_1, and nonconcrete in W_2, while x_2 is nonconcrete in W_1, and

---

[9]   According to Fine, there are two senses of "true" for propositions: an inner sense in which a proposition's being true at world W requires its existence at W, and an outer sense in which it does not. For details of Fine's theory see (Fine 1978) and for Zalta and Linsky's criticism to Fine, see (Zalta and Linsky 1994, 450).

[10]   In other words, we can say the Barcan Formula, $\lozenge(\exists x)\,F_x \supset (\exists x)\,\lozenge F_x$, is valid in our actual world.

concrete in W_2, and x_3 is concrete in both. Thus, considering Hank a concrete object in a possible world, and considering the same domain for our world and that possible world, we conclude Hank actually exists in our world too, but his actuality in our world is contingently nonconcrete.

Accordingly, "Socrates exists" could be informative if it is analyzed based on the variety of existents in Zalta and Linsky's theory. When one utters this proposition, he absolutely considers Socrates as an actual existent. But, his audience does not know what kind of existent Socrates is. So, "Socrates," as the subject of this proposition, at first might be supposed either a contingently nonconcrete or concrete object in the actual world. But, the rest of this proposition informs the audience that Socrates is a concrete existent. Similarly, "Hank does not exist" informs the audience that Hank is not a concrete object; he is a contingently nonconcrete object, which is concrete, at least, in a possible world. The mentioned contradiction in Frege's view arises when the existence of Hank and the predicate "exists" are assumed the same. But, "the contingently nonconcrete existent of Hank is not concrete in the actual world," does not entail any contradiction.

Apparently, Zalta and Linsky's idea could solve both being uninformative and the contradiction problems. It, however, faces difficulties, which especially come from Bennett and Tomberlin's criticisms.[11] Though some of these criticisms have been responded to,[12] it seems that some of them are still significant challenges for the idea of contingently nonconcrete objects.

Bennett, for example, says that considering these objects existents in the actual world violates our intuitions about the modal profiles of everyday objects (Bennett 2005, 301–302). When we talk about existents, we intuitively understand concrete objects. Menzel also asserts that only ordinary concrete objects and abstract objects fall under the same intuitive, historically well-grounded concept of (general) existence (Menzel 1993, 199).

Contingently nonconcrete objects are, in fact, alien for our linguistic intuition when we use the term "existent," even if in some formulas, such as Barcan, the quantifiers' domain includes these objects. Indeed, the root

---

[11]   Although, there are many objections to Zalta and Linsky's theory, we discuss only those that are related to our topic. Once again, we remind that this paper focuses on the existential propositions with a proper name as the subject.

[12]   For example, see (Bennett 2009).

of Menzel and Bennett's objection is placed in supposing the same domain of objects for all possible worlds. We said before that Zalta and Linsky's argument to know Hank as an actual existent in the actual world is: what exists in a world is everything in the stock, concrete or not; the entire stock is the domain of a world, like the domain of all other worlds (Bennett 2006, 269). But, we have the right to ask why the domain of all worlds is the same? Since she could not find a convincing response, Bennett could conclude that no actually existing thing has the modal property "possibly being a Keebler elf" (Bennett 2006, 270). Accepting Bennett's linguistic intuition challenge, "Socrates exists" means "concrete Socrates is a concrete existent," and "Hank does not exist" means "concrete Hank is not a concrete existent." Thus, both Frege's problems (being uninformative and the contradiction) are left unsolved. She also argues that Zalta and Linsky's view has a lot in common with Plantinga's. Though, these two parties have three main differences, Bennett says, their structures are the same. Thus, the difficulties in Plantinga's view could also apply to Zalta and Linsky's (2006, 267–72).

Tomberlin has similar objections to Zalta and Linsky's view, but his criticisms are a bit more destructive. He shows that a single and fixed domain in every possible world not only is not justified, but also could result in some incompatibilities (Tomberlin 1996, 273–76). In simple language, Tomberlin says that in Zalta and Linsky's theory, possible objects have been confused with actual existents, even if we call them "contingently nonconcrete objects." Instead, Tomberlin and his like-minded colleagues believe in possible objects, which are also known as possibilia. In this view, all objects are possibilia,[13] but some of them are actual and the rest are mere possibilia. Thus, mere possibilia do not exist in the actual world, but might have existed.[14] Menzel argues for this idea as follows: "*A possibile* is an

---

[13]   There is another version of this view that might be called moderate possibilism. According to this version, "all things are not possible," is false, but there are some ways to show that such an ontological realm is narrower than one might have supposed. See (Voltolini 2000).

[14]   For most of them, the quantifier "∃" is used as "there is" and the quantifier "E!" is used as "exists" to assert the difference between the realms of possibilia and existents. But, for an actualist there is no difference between these two quantifiers.

object that, while not actually concrete, is nonetheless *possibly* concrete. Since nothing that is possibly concrete is abstract, it follows that *possibilia* do not exist" (Menzel 1993, 199).

So, based on this idea, "Socrates exists" could be informative in the way that: we suppose the audience does not know whether Socrates (a possibile) is actual or a mere possibile. By uttering this proposition, we say to the audience: he is actual. Also, "Hank does not exist" does not contain a contradiction if we consider this proposition as "the mere possibile Hank is not actual."

This view could be analogized with Meinong's idea about the types of objects, in spite of obvious differences between these two theories (Zalta and Linsky 1994, 440). In Meinong's view, objects are divided into the objects that have being and the objects that do not have being. Even, Meinong says, impossible objects, such as "round-square," have a kind of objectivity and are placed under the latter category (Marek 2013). In other words, for Meinong the realm of objects is more extended than existents, and being an object is sufficient to be considered in mind and be the subject of propositions.[15] Thus, in "Hank does not exist" there is no contradiction; though Hank is an object (and so we can talk about him), he does not have being and is not in the existents' category.

However, all philosophers who think the domain of objects is more extensive than existents, first must cogently respond to this question: what does "being an object" mean without existence? As far as we can see, even the great possibilist philosophers, such as Kripke, only say Hank, for example, might have existed in the actual world. But, they do not illustrate what the ontological type of Hank is and how we can put it in a meaningful proposition as the subject. In other words, they do not clearly show the existential referent of Hank in "Hank does not exist." It is not satisfying to say Hank refers to a possible, but nonexistent thing. For, the phrase "nonexistent thing" could fall in a chain of circular explanations.

---

[15]   We remind the reader that what makes Frege's problem a serious one is such a presumption that, "We should not admit merely possible objects when everything we want to say (and everything that we can say) can be said with what there actually is." See (Fitch 1996, 68).

Bennett has attempted to respond to this question by introducing "proxy objects." In simple words, she believes that each mere possibile has an existent proxy in the actual word in the way that there is a specific relation (proxy function) between possible objects and their proxies (Bennett 2006, 272). Bennett does not believe that anything exists without actually existing, but believes some things are actual without existing (Bennett 2006, 282). She calls something like Hank "quasi-alien," clarifying that "although it is not in fact possible for there to be anything that does not actually exist, it is possible that some lesser-status thing have the higher ontological status—and that is all our intuition about the possibility of aliens requires" (Bennett 2005, 302). As I understand, Bennett has not done anything except alter some words. "Being something actual without existing" is as far from our linguistic intuition as "mere possibilia" or "contingently nonconcrete object." In Bennett's theory, the problem is not transferring from a lesser to higher ontological status, or substituting a possible object with an existent. But, the problem is: what are we exactly talking about when we say "a possible actual nonexistent object?"

The other alternative theory has been presented by Timothy Williamson. Williamson shows that a phrase like "possible Hank" could be interpreted in two ways: "x is Hank and x is possible," or, "it is possible that x is Hank." In his view, the former interpretation is wrong, because it attributes "possible" to x, and the latter is true since it ascribes "possible" to "being Hank" (Williamson 2000, 201). Indeed, he says "possible" is the mode of properties, not objects. Thus, each object has a necessary existence. His argumentation is as follows (Williamson 2002, 233–34):

1. Necessarily, if I do not exist then the proposition that I do not exist is true.

2. Necessarily, if the proposition that I do not exist is true then the proposition that I do not exist exists.

3. Necessarily, if the proposition that I do not exist exists then I exist.

4. Necessarily, if I do not exist then I exist.

5. Necessarily, I exist.

He asserts, however, this necessary existence is not a concrete, physical existence. A physical existent, Williamson believes, is a spatiotemporal one,

but this is only a narrow domain for extensions of the concept "existent"[16] (Williamson 1999, 259). Unlike physical existents, necessary existents are in the realm of logical objects, and so, Williamson calls them logical existents. This realm is not limited by time and space, and its domain covers a plethora of objects, such as propositions, rules, and of course, necessary existents, and thus we can talk about them. Therefore, based on Williamson's view, Hank is a logical existent, but he is not a physical one, and so "Hank does not exist" is not a contradiction. Also, each physical existent could have a corresponding logical existent. But, logical existents do not have the properties of physicals. For example, the physical existence of fire has the property of burning, but the logical existence of fire does not. In the same way, "Socrates exists" means that the logical existence of Socrates, has a physical corresponding existent.

It seems that Williamson's idea, on the one hand, has removed many of the previous challenges, and on the other hand, faces lesser challenges than the other views. However, as I understand this theory, there are some ambiguities in it. For instance, the realm of logical objects requires more clarifying. Could we call logical existents abstract objects? Is the logical existence of Hank independent from our mind? Or, without mind, is there no logical existence of Hank? And, basically, what is the relationship between the logical and physical realms?

4. Is "existence" a concept?

All above ideas about the variety of existents believe that "exists" in "Socrates exists" is a concept. The only problem is if it is a first or a second order concept. But, we have the right here to review existence as a concept. Undoubtedly, "being existent" is a concept; we can abstract this concept from the existents x, y, z, as well as children can abstract "whiteness" from "white wall," "white chair," "white dog," etc. We can say, as we say about the other mental concepts, "these existents are under the concept 'being

---

[16]   As far as I have seen, Williamson has not believed in concrete but non-physical existents. Some thinkers believe in these kinds of being and exemplify God as an instance.

existent,'" and we can apply this concept to the existents other than x, y, z. It seems that when we say "existence is considered as Socrates' property," we mention this kind of existence. Let us call this kind of existence "conceptual existence."

This is also possible for us to change our perspective we look at "existent Socrates" from. In the new perspective, we consider "Socrates" as an existence which has many properties, such as being a philosopher, keen mind, etc. In other words, "Socrates"—as a single unit of reality—can be intellectually analyzed into two aspects: an existence and a set of properties which are related to, or precisely the manifestations of, that existence.[17] Of course, this "existence" is not a conceptual but a concrete fact, and let us call it "factual existence." In this intellectual analyzing, factual existence should be prior to Socrates's properties, because the properties need something to attach to. We must note that this priority does not implicate a physical distinction between existence[18] and the Socrates's properties; they are identical in reality.

The intellectual priority of existence to everything about Socrates, leads us to invert the proposition "Socrates exists" to "the existence$_S$ is Socrates." Indeed, existence$_S$ is the principle reality of what is known as Socrates, but our natural language expresses this reality as the inverse of predication.[19] In other words, "existence$_S$" is neither a first nor a second order concept, it is an object, and its properties reveal and describe it. Therefore, Frege's problem on "Socrates exists" being uninformative is basically removed. Since, the de facto proposition is "the existence$_S$ is Socrates," which is informative and meaningful, just as "the existence$_P$ is Plato," etc.

---

[17]   The author thinks that this assumed hypothesis is not less reasonable than the others, though there is no perfect argument to prove, or at least to justify, it.

[18]   Avoiding confusion, we use "existence$_S$" for the factual existence of Socrates.

[19]   There is a theory called "Principality of Existence" in the context of Islamic Philosophy, which is founded by *Mulla Sadra*. The main claim in this theory is that existence is the principal reality of things, and quiddity is a subordinate reality. So, it is not an innovational theory of the author. Although, this theory is overlooked by modern analytic philosophers, it has many strengths that are valuable for philosophical debates. For more details about this theory, see (Asadi 2017).

Similarly, "Hank does not exist" is, in fact, "there is no existence that is Hank." It means that among all existences, none of them is the factual existence of Hank. It could be seen that in the latter proposition, the subject belongs to "existences," and this point makes it possible to utter and be a meaningful proposition. On the other hand, it is not a contradiction, because we do not say "an existent is not existent." What we say is that "existence$_1$ is not Hank," "existence$_2$ is not Hank," and so forth. In this view, we do not want to reject the other types of existences completely. Hank might have another type of existence, such as logical, though it would not be concrete.

5. Conclusion

Disregarding philosophers' conflicts about actualism and possibilism, it seems that employing their ideas cannot perfectly solve the problems brought up by Frege. As we have briefly shown in this paper, each of these views faces some difficulties that prevent it from being a generally acceptable theory. This does not mean all of these ideas are wrong, but it is a fact that they could not give us a satisfactory view of the challenges in existential propositions. If what we suggested here, which could be called "Principality of Existence Theory," does not have less challenges than the other mentioned views, it does not have more than them. We do not say there is no difficulty in it, or it is the absolutely perfect solution. But, we claim that it could be an alternative way for thinking about and discussing existential propositions among all traditional ways.

References

Aristotle. *Metaphysics.*

Asadi, Siavash. 2017. "An Interpretation of Proper Name References Based on Principality of Existence Theory." *The International Journal of Islamic Thought* 11: 43–51. https://doi.org/10.24035/ijit.11.2017.005

Ayer, Alfred. 1949. *Language, Truth, and Logic.* London and Southampton: Camelot Press Ltd.

Bennet, Karen. 2005. "Two Axes of Actualism." *The Philosophical Review* 114 (3): 297–326. https://doi.org/10.1215/00318108-114-3-297

Bennet, Karen. 2006. "Proxy Actualism." *Philosophical Studies* 129 (2): 263–94. https://doi.org/10.1007/s11098-004-1641-2

Fine, Kit. 1978. "Model Theory for Modal Logics." *Journal of Philosophical Logic* 7 (1): 125–56. https://doi.org/10.1007/BF00245925

Fitch, G.W. 1996. "In Defense of Aristotlian Actualism." *Philosophical Perspectives* (Metaphysics), 10: 53–71. https://doi.org/10.2307/2216236

Forbes, Graeme. 1987. [Rev. on] Alvin Plantinga, edited by James E. Tomberlin and Peter van Inwagen. *Noûs* 21 (1): 60–66. https://doi.org/10.2307/2215066

Frege, Gottlob. 1953. *Foundations of Arithmetic.* Translated by J.L. Austin. Oxford: Blackwell.

Frege, Gottlob. 1960a. "On Sense and Reference." In *Translations from the Philosophical Writings of Gottlob Frege*, edited by Peter Geach and Max Black, 56–78. Oxford: Basil Blackwell.

Frege, Gottlob. 1960b. "On Concept and Object." In *Translations from the Philosophical Writings of Gottlob Frege*, edited by Peter Geach and Max Black, 45–55. Oxford: Basil Blackwell.

Frege, Gottlob. 1960c. "Function and Concept." In *Translations from the Philosophical Writings of Gottlob Frege*, edited by Peter Geach and Max Black, 21–41. Oxford: Basil Blackwell.

Frege, Gottlob. 1979. "Dialogue with Punjer on Existence." In *Posthumous Writings*, edited by Hans Hermes, Friedrich Kambartel, and Friedrich Kaulbach, 53–67. Chicago: The University of Chicago Press.

Leibniz, Gottfried Wilhelm. 1988. *Theodicy.* Chicago: Open Court Publishing.

Lewis, David. 2001. *Counterfactuals.* Oxford: Blackwell.

Marek, Johann. 2019. "Alexius Meinong." In *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2019/entries/meinong/

Mendelsohn, Richard L. 2005. *The Philosophy of Gottlob Frege.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511497964

Miller, Barry. 1975. "In Defense of the Predicate 'Exist.'" *Mind* 84 (1): 338–54. https://doi.org/10.1093/mind/LXXXIV.1.338

Nelson, Michael and Zalta, Edward N. 2009. "Bennet and 'Proxy Actualism.'" *Philosophical Studies* 142 (2): 277–92. https://doi.org/10.1007/s11098-007-9186-9

Plantinga, Alvin. 1974. *Nature of Necessity.* Oxford: Clrarendon Press.

Plantinga, Alvin. 2003. "Two Concepts of Modality: Modal Realism and Modal Reductionism." In *Essays in the Metaphysics of Modality*, edited by Matthew Davidson, 192–228. New York: Oxford University Press. https://doi.org/10.1093/0195103769.003.0011

Plantinga, Alvin. 2003. "World and Essence." In *Essays in the Metaphysics of Modality*, edited by Matthew Davidson, 46–71. New York: Oxford University Press. https://doi.org/10.1093/0195103769.003.0003

Salmon, Nathan. 1987. "Existence." *Philosophical Perspectives* 1 (Metaphysics): 49–108. https://doi.org/10.2307/2214143

Stalnaker, Robert C. 1976. "Possible Worlds." *Noûs* 10 (1): 65–75. https://doi.org/10.2307/2214477

Tomberlin, James E. and McGuinness, Frank. 1994. "Troubles with Actualism." *Philosophical Perspective* 8 (Logic and Language): 459–66. https://doi.org/10.2307/2214182

Tomberlin, James E. 1996. "Actualism or Possibilism?" *Philosophical Studies* 84 (2/3): 263–81. https://doi.org/10.1007/BF00354490

Voltolini, Alberto. 2000. "Are All Alleged Possible Objects There?" *Topoi* 19 (2): 209–19. https://doi.org/10.1023/A:1006437301069

Williamson, Timothy. 1998. "Bare Possibilia." *Erkenntnis* 48 (2): 257–73. https://doi.org/10.1023/A:1005331819843

Williamson, Timothy. 1999. "Truthmakers and the Converse Barcan Formula." *Dialectica* 35 (3/4): 253–70. https://doi.org/10.1111/j.1746-8361.1999.tb00186.x

Williamson, Timothy. 2000. "The Necessary Framework of Objects." *Topoi* 19 (2): 201–08. https://doi.org/10.1023/A:1006405915896

Williamson, Timothy. 2002. "Necessary Existence." In *Logic, Thought, and Language*, edited by A. O'Hear, 233–51. Cambridge: Cambridge University Press. https://doi.org/10.1017/S1358246100008158

Zalta, Edward and Linsky, Bernard. 1994. "In Defense of The Simplest Quantified Modal logic." *Philosophical Perspectives* 8 (Logic and Language): 431–58. https://doi.org/10.2307/2214181

Zalta, Edward and Linsky, Bernard. 1996. "In Defense of the Contingently Non-Concrete." *Philosophical Studies* 84 (3/4): 283–94. https://doi.org/10.1007/BF00354491

Defending the Good Dog Picture of Virtues

Andrei Ionuţ Mărăşoiu*

Abstract: I consider and reject a specific criticism advanced by Korsgaard against virtue ethics and epistemology when these are conceived with the help of what she calls the image of the "Good Dog." I consider what virtue ethics and epistemology would look like if the Good Dog picture of virtues were largely correct. I argue that attention to the features that make Korsgaard undermine the usefulness of virtues when conceived along the lines of the Good Dog picture reveals the opposite of what she claims. On the Good Dog picture, virtue ethics and epistemology are seen as more promising approaches to rationality than Korsgaard's own advocacy of reflection.

Keywords: experience; expertise; fluency; intellectual virtue; moral virtue; necessitation; normativity; rationality; reflection.

1. Introduction

Christine Korsgaard writes:

> Some virtue theorists have offered us the (to my mind) equally rebarbative picture of the virtuous human being as a sort of Good Dog, whose desires and inclinations have been so perfectly trained that he always does what he ought to do spontaneously and with tail-wagging cheerfulness and enthusiasm. [On this picture,] the

* Institutul de Cercetare al Universității din București (ICUB)
 ✎ Str. Dr. Dimitrie Brândză nr. 1, Bucureşti, România 060102
 ✉ andrei.marasoiu@icub.unibuc.ro

experience of necessitation is a sign that there is something wrong
with the person who undergoes it.

The trouble with those two images of virtue—the Reformed
Miserable Sinner and the Good Dog—and with the philosophical
theories behind them, is not merely that they denigrate the ex-
perience of necessitation. It is also (and relatedly) that they do
not give an adequate explanation of how we are necessitated.
(Korsgaard 2011, 3–4)

Oddly enough, she doesn't return to the Good Dog picture. It mainly seems
to serve as a brief foil for introducing her own views of the role and nature
of rational necessitation to act. By "necessitation," Korsgaard means the
rational connections constitutive of our agency to the extent that values we
guide ourselves by serve as norms for us in action, norms we not only com-
ply with, but which we *follow*—and which *guide* us so we may follow.

In this text, I only incidentally explore Korsgaard's own positive views
of rational necessitation and self-constitution. Rather, I focus on the step
that seems to go by too quickly: the criticism of virtue theories implicit in
the passages just quoted.[1] I argue that this criticism, once fully articulated,
can be successfully rejected, revealing how virtues may be grasped using
precisely the Good Dog picture that Korsgaard decries.

2. The target

What is the target of Korsgaard's criticism? Surely it cannot be virtue
theories generally, for these conceive of virtues—intellectual or moral
alike—in starkly different ways. In this section, I argue that Korsgaard's
target is best construed along the lines of Railton (2011).

First note that two prominent contemporary approaches to virtues can-
not be Korsgaard's target. For Foot (2001), just as growing is the plant's

---

[1]   Korsgaard doesn't mention anyone explicitly by name. But talk of cheerfulness
and of the naturalness of the dog, as well as of its well-tailored behavior seems to
suggest she is primarily targeting views like Hursthouse (1999), Foot (2001) and
Railton (2011). It is with them in mind that I proceed.

good, it is the rational exercise of free will that is a person's good. Foot tries to accommodate this within virtue ethics, writing:

> I believe that evaluations of human will and action share a con-
> ceptual structure with evaluations of characteristics and opera-
> tions of other living things, and can only be understood in these
> terms. (Foot 2001, 8)

It follows that the cultivation of virtues should aim at enhancing one's rational exercises of free will. The view fits well with Korsgaard's talk of agency and self-constitution. For, on this view, it is by our rational and free acts that we become ourselves more by becoming better people. Foot's distinction between what is good in the natural world and what is good for a person makes it clear that Korsgaard's complaint that a Good Dog picture of virtues misconstrues them cannot target Foot's virtue theory.

Nor is Hursthouse's (1999) virtue theory a target of Korsgaard's objection. For Hursthouse starts her discussion of virtues by considering a bridge principle between rule-based and virtue-based approaches to morality. On that principle: "An action is right iff it is in accordance with a correct moral rule or principle" (Hursthouse 1999, 26). Hursthouse argues that one's virtue theory should be devised such that this principle *could*—in principle— be satisfied. And this constraint on what gets to count as a (moral or intellectual) virtue doesn't fit a Good Dog picture of virtues. This is because people and animals alike may be subject to habitudes, whereas only persons can *follow* rules. The possibility of re-describing virtue-based behavior as rule-based behavior seems to be at odds with a Good Dog picture of virtues.

I believe Railton's (2011) view best fits the Good Dog picture of virtues. For Railton, virtues are situationally flexible habitudes, open-textured enough to afford changes on the fly as we meet with slightly atypical situations, in a way that cannot—except partly and misleadingly—be captured by the formulation of a rule to govern our behavior. Consider a few examples:

> Like a skilled athlete, artisan, or artist, the virtuous agent pos-
> sesses an acquired mastery and self-control that explain her re-
> markable ability to rise to the occasion. As with any mastery, its
> intuitive exercise can be intuitive and 'non-deliberative,' yet at
> the same time mindful, self-governed—even, perhaps, the highest

> form of creativity and self-expression. The skilled individual can
> know what she is doing, and why, 'in the moment,' without need
> for self-conscious reflection. (Railton 2011, 298)

The examples illustrate the continuity between virtues—moral and intellectual alike—and everyday habitudes manifested by the athlete, skilled orator, or seasoned worker. This is one reason to endorse Railton's view, and a reason that fits well with the Good Dog picture, for it proposes a natural account of virtues, one on which these are continuous with flexibly adaptive yet rational animal behavior.

3. A reason to consider the Good Dog picture

There is a second reason to endorse Railton's view, and, more generally, views which seem to ply to the Good Dog picture of virtues. Namely, that such views offer *prima facie* credible replies to situationist objections to virtue ethics and epistemology.

According to those objections, in a nutshell, virtues are dispositions, so they span situation-types which include several distinct situations, each with their own peculiarities. A fully rational behavior should consider these peculiarities and flexibly adapt to them. Whereas virtues as dispositions cannot fully do so while still retaining their identity across distinct situations. In any given situation, to the extent that our response is rationally sensible to the circumstances, it will always be an open question whether our behavior is due to a virtue or to our awareness of the specifics of that situation. It will always be an open question exactly how virtues should be individuated so that their specification does not include situation-relative factors.[2]

To the situationist objections, Railton's reply is to distinguish run-of-the-mill dispositions from habitudes, and argue that virtues are habitudes:

> As Aristotle emphasized, acquiring a skill or mastery is not
> simply internalizing a set of rules or procedures. No set of rules
> or procedures could be sufficient; nor could it apply itself. One

---

[2]   For careful expositions of these objections, see Harman (1999) and Alfano (2012).

must instead have proper habitudes, which include dispositions to notice certain features of situations, to feel their force or urgency, to appreciate the values at stake, and to be moved appropriately in thought and action. Put in terms introduced earlier, such habitudes thus involve both evaluative perception and practical attunement. As a result of a fortunate nature, exposure to good examples and proper training in youth, and from her own growing experience, the virtuous person acquires such habitudes, enabling to her to act in the right way and for the right reasons—to respond spontaneously and aptly to relevant reasons for acting as such. (Railton 2011, 316)

Habitudes can be manifested and changed contemporaneously in response to new situations, mirroring both a history of learning and the sage's rational sensibility to her current circumstances. The notion of virtues as habitudes seems to be a good fit for what Korsgaard calls the Good Dog picture of virtues.

The reason I mention this is not to offer additional insight into how situationist challenges could be met by a virtue theory. Rather, I only note that, to the extent that Korsgaard decries the Good Dog picture of virtues that Railton seems to endorse, she incurs the burden of showing how one could conceive of virtues in a way that is situation-flexible and yet would be mischaracterized by the Good Dog picture. As far as I can tell, Korsgaard doesn't offer a positive virtue theory along those lines. This is at least an initial motivation to take the Good Dog picture of virtues, and Railton's attendant explicit theory, seriously and to inquire into the rationality of virtuous behavior conceived that way.

In what follows, I will argue that a Good Dog picture fares better than Korsgaard's own reflection-based view of virtues and rationality. As the Good Dog picture of virtues is more open-textured than Railton's own specific virtue theory, I will consider that picture in full generality, and still argue that Korsgaard's specific appeal to the nature and role of reflection fails to undermine the role which Good-Dog virtues ought to play in rational behavior.

4. Virtues

Traditionally,[3] virtues have been deemed *virtuous* for one of two reasons. Either because they are reliable, or because we can properly be said to exercise them as their due authors. The distinction is controversial, however.

First, everyone agrees virtues are *reliable.* Though, perhaps, assent is hasty. Presumably, part of the naturalness of a genuine virtue is its occasional failure to be triggered in circumstances that naturally trigger it. To the extent that we are considering theories of rationality—and virtues—for human beings like us, this fact should be given prominence instead of being ignored. It can, of course, be accommodated, given a flexible enough notion of reliability. But the traditional notions of reliability on offer—ratio of good performance (truth) over good plus bad performance (truth or falsity), or counterfactual conditions like safety and sensitivity—fail to fully codify this fact.

Second, the criteria advanced for what makes us the *authors* of our acts of virtue often fail to be criterial. To illustrate, I'll briefly mention two criteria of authorship, or responsibility for virtues, proposed by Zagzebski (1996).

First, to think something counts as an act of virtue to the extent it helps us flourish—in our daily lives or our lives of knowledge irrespective—is to put the cart before the horses. What if our lives simply turn out well by sheer luck? We get all the goods—moral and intellectual—for no effort of ours. And we don't get any of the bads—moral or intellectual—again, without merit. The possibility cannot be ignored. For we *might* flourish given all these benefits. But we wouldn't, in an important sense, author or be responsible for how our lives proceed.

Second, to think something counts as an act of virtue to the extent we're motivated by the right things in doing it is to ignore much of our current moral phenomenology. Consider the possibility of fulgurations, brute or unexplainable motivations, whims that lead to transformations in our practical or intellectual lives, moments of grace or love, deference to practices

---

[3]   E.g., cf. Zagzebski (1996, ch. 2).

whose motivations stay opaque to us, fluent participation in mystical traditions, etc.

So it's unclear both what the reliability of, or authorship for, acts of virtues comes down to. Clearly, the distinction about what makes virtues virtuous needs revisiting. Still, I believe we can make headway in beginning to address the question.

The reliability that matters for virtues is that exhibited in producing rational goods: acts or inferences. And this amounts to satisfying the norms of rationality, whatever those norms may be: logical or probabilistic consistency, explanatory power, simplicity relative to our cognitive make-up, etc. The satisfaction of the norms in question is, presumably, also up for grabs: whether we merely pass some test or examination, whether we faultlessly excel, or whether occasional failings may be passed over in an otherwise good work, far above mediocre satisfactoriness. Notice that, here, considerations of authorship drop out: our minds merely host our reliable dispositions.

What, now, makes us the authors of our acts of virtue? Here we begin to scratch the surface of Korsgaard's complaint and what might be wrong with it. Two options seem, again, available for authorship. One is to say we satisfy some formal criterion—say possible endorsement of our thought or deed were we to reflect on the matter. This is Korsgaard's earlier view (1996) of rationality as acting by the lights of norms whose authority is bestowed upon them by our own reflective endorsement.[4] Another view is to say we are good-natured: "Good dog!" We do what becomes us, what is in our character, what makes most sense given our circumstances and setting. To see what might be wrong with Korsgaard's complaint against this Good Dog picture of the virtues, let's briefly detour to consider the chances of reflection.

---

[4]   Which seems to also be how Korsgaard construes Kant's criterion that we should be able to legislate for ourselves, and are free when we abide (only) by this body of self-legislation.

5. Reflection

Reflection[5] concerns thoughts, feelings and perceptions we have. It's not a mere feeling or presentment, but is conceptually articulate. This doesn't mean that it needs to be put into words; we can reflect without talking to ourselves. Typically, we reflect for one (or both) of two reasons. Either to *search* for a solution to a problem, or to *evaluate* something that's already been done by the light of some canons.

What role may reflection play in reason? Of course: we may consciously, deliberately, articulately, evaluate performances relative to norms in reflection. But it is far from clear that reflection, in such a procedure, does anything but to take note of a *fait accompli*, viz. whether the performance is good or rational enough or not.[6]

In contrast, suppose, with Korsgaard, that reflection does play a constitutive role, endowing rational norms with their authority. Who invested reflection with that authority? To wit: *why* should our good-natured, tail-wagging character—moral and intellectual alike—not enjoy the same or better position in the rational authority-bestowing game? An argument to that effect seems to be missing in Korsgaard's criticism.

Bracket now the specific rational role reflection plays; how does reflection play any rational role at all? We could construe reflection along reliabilist lines.[7] If so, however, reflection would be scarcely recognizable from old-trodden first-order thought. It would make nonsense of reflection's supposed rational advantage to think it's the same kind of push-you-pull-me of everyday thought, only one floor up.

---

[5]   On my use of the word, which is, I believe, shared by many, Korsgaard included.

[6]   It matters little how "realistic" a stance one takes with respect to norms or values here. Even if they were nothing but the products of our minds, they need not be reflective—they could be emotive. Or even if they were the products of intellect (or reflection writ large), it still doesn't follow that we should be able to have an *occurrent* reflective thought involving the normative content in question in order for evaluation to be feasible.

[7]   Kornblith's (2010) "What Reflective Endorsement Cannot Do" is telling here in showing reflective processes are no better than other thought processes in their exposure to biases and all kinds of errors, mishaps and misconstruals.

The reliability of reflection would only have a rational afterglow if it objectified one (or both) of two things. One possibility is that reflection should be more than brute by realizing some formal procedure of rational choice operating over one's thoughts. Another possibility is that reflection should be more than brute by being ours—by being, somehow, a closer intimation of our agency than first-order thought. Why should reflection (construed either way), get the better of virtues (construed either way)? To this I now turn.

6. Virtues and reflection taken jointly

One line of thought might be this. We pick sides between competing views of rationality, and the superiority of reflection flows from that.

If we thought rational choice principles are *best* realized in conscious deliberate reflective thought, then reflection should trump virtues on that score. In point of fact, however, we need a separate argument for why rational choice principles should be better codified under the supervision of reflection, rather than letting our epistemic character do its job and seamlessly—*virtuously*—produce the right conclusions and recommend the right deeds.

If an act or thought couldn't be rational unless it were ours—unless we were its authors—perhaps that very claim involves reflection. Whatever benefit we might draw from virtues, reflection would necessarily partake in authorship. However, consider the *fait accompli*. Reflection may, perhaps even always, rubber stamp our thoughts and deeds as ours. But this editorial task isn't tantamount to the creative process authors go through prior to publication. Analogously, reflection may brand a deed or thought as ours (editorially dubbing them as ours) even though what is so branded is the toil of our epistemic character and its formation (the authors in the analogy). The connection between reflection, on the one hand, and rationality and authorship, on the other hand, is questionable.

Instead of cherry-picking views of rationality so that reflection may one-up virtues and character, we should, perhaps, readily acknowledge these aren't opposed at all. Reflectiveness is a virtue, and, if the word doesn't stand for a unique phenomenon, there may well be a constellation of reflective

abilities which we may exercise in equilibrium with other cognitive abilities. Along these lines, Peter Goldie writes:

> [T]here is a normative requirement to be motivated to have, and to have, the right habits and dispositions of thought, such that doubts will arise when and only when they should. On particular occasions, much of our thinking will be unreflective, and not part of conscious deliberation, so we will need to rely on our habits and dispositions, at work in the background of our minds, so to speak. (Goldie 2004, 251)

Goldie references what he calls a virtue of reflectiveness, exercised all and only when called for.

Suppose we recognize the fact that reflection and virtue often partner in reflectiveness. The question left outstanding, however, is why this particular virtue should be insulated from one's overall epistemic character and given place of pride in what endows our epistemic or practical norms with their authority (if one finds that thought congenial in the first place). Here is an analogy: It is as though one thought that open-mindedness alone among all intellectual virtues is responsible for all the notable progress in science. Such a view would be both initially implausible and hard to defend. Unless, that is, one waters down the meaning of "open-mindedness"—or, analogously, of "reflectiveness"—so that is simply becomes another label for most of one's relevant virtues or overall character. It is unclear whether reflectiveness as a virtue holds any special status over and above other virtues.

7. Virtues and humanity

It is worth considering another reason why one may resist the Good Dog picture of the virtues, even if this reason doesn't provide direct support for Korsgaard's own view. Thus, Gary Watson writes:

> [T]he notions of self-disclosure or deep appraisal are richer than the notion of a skill. For beings without self-reflective capacities can be more or less skillful, as dogs can be good at catching Frisbees. The appraisal of skills or talents is importantly different from aretaic evaluation in a way identified by Aristotle.

> Knowledge of an agent's ends, intentions, and efforts has a different [e]ffect on aretaic appraisals than on the others. Indifference in a performance doesn't count against one's skill, whereas a less than wholehearted effort to save someone's life does impugn my moral character. Talent and skill are fully displayed only in wholehearted performances, whereas the aretaic perspective is also concerned with the "will," that is, with one's purposes, ends, choices, concerns, cares, attachments, and commitments. Not trying can be a failure of virtue but not of skill. (Watson 1996, 244)

Watson is here arguing against the generalization of the notion of aretaic responsibility from virtues to skills. In this text, I remain neutral about that generalization, though, to reply, it is worth pointing out that although virtues do differ from skills in important respects, that falls short of showing that the notion of responsibility applicable to acts of virtue *doesn't* apply to acts of skill as well.

As a part of explaining why aretaic appraisals fit virtues and not skills, Watson remarks in passing that "beings without self-reflective capacities can be more or less skillful, as dogs can be good at catching Frisbees. The appraisal of skills or talents is importantly different from aretaic evaluation in a way identified by Aristotle." Now whether Aristotle might have agreed that the Good Dog picture of virtues is mistaken is also beside the point here. The question I'm pursuing is why think that a dog catching a Frisbee is such an ill-suited model for excellent production of virtuous acts. (Surely it may sometimes be inappropriate, depending on context; but the question here is why that suggestion is dismissed offhand, as *in-principle* inadequate.)

We should distinguish two suggestions in interpreting Watson here. The first is that we may only exercise virtues if we possess reflective—and, in particular, self-reflective—capacities. Or, in other words, that acts of virtue are necessarily such that we may, on reflection, endorse them. The second suggestion Watson may make is that only people—not dogs or other beings unable to self-reflect—may be virtuous. It's important to see Watson's first suggestion isn't supported by his second suggestion.

It seems to me that loyalty, generosity, mercy and many other virtues may properly be illustrated by, say, dogs. Indeed, this seems to me to be such a commonplace as to barely be worth pointing out, were it not met

with reluctance by authors like Korsgaard or Watson.[8] However, just to stay close to the dialectic, let us grant to the two that only humans may be virtuous, and do acts of virtue. Let us also grant, for the sake of argument, that only humans are able to reflect, and, in particular, to reflect upon themselves.

Still, *nothing* follows about how, if at all, virtues and self-reflection relate to each other. In particular, simply because only humans may perform acts of virtue (an assumption granted for dialectic purposes alone) and only humans may self-reflect (also granted for dialectic purposes alone), it doesn't follow that self-reflection—the activity itself, or our ability to undergo it—has anything special to do with carrying out acts of virtue, or with virtuous character dispositions (whether intellectual or moral).

As for wider implications concerning humanity—how it may necessarily involve either the ability to perform acts of virtue, or the ability to reflect upon oneself, or both, nothing has been settled. *A priori*, we have no good reason for presupposing that only in virtue of humanity may be we virtuous or reflective, nor that only in virtue of virtue or reflectiveness may we be genuinely human. Indeed, what more precisely we mean by "humanity"—humaneness and conscience, on the one hand, or just a form of speciesism, on the other hand—hasn't been determined here, and need not be determined in order to ascertain how virtues and reflection relate to each other. In all, it is hard to see how Watson's occasional reference to Frisbees helps Korsgaard's criticism of the Good Dog picture of virtues.

---

[8]   One might think that humans enjoy rationality, virtues, and reflection—and that these things go together such that dogs can't enjoy any, however cuddly or well-behaved. Put bluntly, this begs the question against a Good Dog picture of virtues and rationality. But more sophisticated forms of the view exist. For instance, one might think that thoughts produced well need some kind of reflective validation (dispositional, and in-principle) in order to count as genuinely rational. This is a more roundabout way of begging the question, but the diagnosis stays the same still. The question remains why we should want the reflective add-on in the first place. This dialectic is on display in Grimm's (2001) objections to Sosa (1991).

8. Moral virtues, intellectual virtues, and the experience of necessitation

So far, I left one part of Korsgaard's criticism of the Good Dog picture of virtues unaddressed. It concerns the experience of necessitation. It was left unaddressed because I first needed to argue that there is no contradiction between being able to reflect upon the task at hand—if and when the task calls for it—and, respectively, manifesting virtues and skills, be they intellectual or moral, or both. Korsgaard writes:

> The trouble with those two images of virtue—the Reformed Miserable Sinner and the Good Dog—and with the philosophical theories behind them, is not merely that they *denigrate* the experience of necessitation. It is also (and relatedly) that they do not give an adequate *explanation* of how we are necessitated. (Korsgaard 2011, 3–4, my italics)

I have, throughout, bracketed Korsgaard's talk of the "Reformed Miserable Sinner," and my defense of the Good Dog picture of virtues should be construed as neither defense nor criticism of what she says about reformation or sin. However, several points need to be made about how virtues connect with what Korsgaard calls "the experience of necessitation."

We should concede that it is at least sometimes true that a logician, expert at proving theorems, will both produce impeccable proofs, and do so without incredible strain on her mental powers. Her expertise might manifest itself, in part, in seemingly effortless proof-building. This fluency might suggest, at first blush, that she experiences no *push* or *felt need* to abide by logical rules—it simply comes naturally. On a naïve and literal interpretation of what "experience of necessitation" might mean, it might be urged that such a logician experiences no necessitation to abide by the rules of logic in thinking about, and building, the proofs in question.

This way of looking at things is mistaken without a doubt. First of all, it's quite natural for experts to experience fluency in competently solving tasks in their area of expertise. It's precisely because our logician is so apt at proofs that she took on the problem of proving even the more difficult theorems. This fluency doesn't in any way preclude the possibility of experiencing necessitation. For, were our logician to discover a blatant violation

of logical rules she overlooked, the strain felt and the need to proceed with care and apply relevant rules with *extra* caution will perhaps be *more* significant than what a logical *ignoramus* might experience.

Fluency in problem-solving isn't in conflict with intellectual virtues like conscientiousness. Although she may come up with a proof in a heartbeat, the logician might—precisely because of her expertise—then go back and carefully double-check each step in the proof. This is evidence of conscientiousness, of intellectual modesty in implicitly admitting she might have made a mistake, and (perhaps) of intellectual courage in exploring all the ways the proof might have gone wrong. Notice, here, that when and how she pays extra attention to what she is doing is *shaped* by her intellectual virtues and epistemic character. Far from possible experiences of necessitation and expert fluency excluding each other, precisely when the expert feels the former might be *called for and made possible*—as experienced—only by the latter.

In order to become an expert and for rules to become second nature, our logician first had to undergo training. Feeling the pull of laws, abiding by the rulebook, retracing our steps—all these are customary during training, and enable us to later perform more fluently, and often without recourse to any explicit recap of known rules in reflection.

So far in this section, I discussed intellectual virtues and epistemic character. But the points stand if we switch to an example concerning moral virtues. Consider courage on the battlefield, a prime moral virtue for Aristotle if there ever was one. The brave are attuned to their environment, and can readily spot an opportunity when they see one. Sometimes there may be no time to reflect at all. And yet it's important that the brave, while they don't overthink things, aren't rash or hasty either. Both these extremes would be vicious rather than virtuous.

Now ask yourself: in the heat of the moment, as they make the virtuously right decision (to defend their positions, or counter-attack, or retreat, or do a side-maneuver, etc.) with accuracy, skill, and getting the expected result out of it—*must* brave soldiers "experience necessitation"? Must they feel they do it from "the motive of duty"? Or that moral laws wouldn't "determine" them otherwise?[9] I submit, as a point of view that might, of

---

[9]   I am grateful to Nora Grigore for bringing instances of heroism to my attention. For the thorny issue of whether such cases are best construed as cases of acting from

course, be mistaken, that—in so far as these case descriptions are correct and not overly under-specified—it is hard to imagine them *having to* undergo conscious experiences where they feel "necessitated" to act as they do. This is so even if and when they are so necessitated. It is so even when soldiers abide by rules of engagement with the utmost precision *and* fluency.

Brave soldiers help another way too. It's important to see that the points I make against Korsgaard aren't disjoint. Thought and deed go hand in hand.[10] Courage requires moderation and is virtuous in that the brave soldier doesn't rush in the line of fire, but jumps in only when the odds are best to secure victory, or an honorable retreat. In other words, moral virtues are best exemplified when co-instantiated with intellectual virtues. For instance, the soldier is bravest when she knows when the right time is to act. She knows this both by being *intellectually* brave (taking the objectively best chance she can when she gets it) and by being prudent and keeping an open mind (she might err in judgement, or a different tactic might work better).

In this section, I appealed to two examples—the crafty logician and the brave soldier—to suggest that *experiences* of necessitation are neither "denigrated" nor "in trouble" if we adopt a virtue-theoretic standpoint with respect to rationality, be it theoretical rationality, practical rationality, or overall agent-rationality. Moreover, the joys and pains of training presupposed in acquiring expertise go a long way toward explaining how "necessitation" actually gets realized in natural beings like us. All this is to support—fallibly, but ostensibly—the Good Dog picture of virtues, both intellectual and moral, that Korsgaard seems to target.

---

duty or of supererogation, and what might be at stake in the debate, see Grigore (2019).

[10]   For a principled reason why thought and deed must harmonize if *full* rationality is presumed, see Mărăşoiu (2018). In response to Korsgaard, however, it matters that degrees of coherence between thought and deed—and, correspondingly, between intellectual and moral virtues—seem to tally with our overall assessments of rationality. I only submit this as a possibility here, for I have proposed no decisive positive argument in its favor. But the bare possibility is enough, *prima facie*, to raise questions about Korsgaard's preferred appeal to reflection.

9. Conclusion

Korsgaard criticizes theorists who sketch a Good Dog picture of virtues. In fact, I have argued, virtue theorists who harbor that picture are not to be criticized but praised for holding the view they do—in epistemology and ethics alike. When properly construed, the work of reflection isn't inimical to a Good Dog picture of virtue.

Instead, exploring how virtues and reflection interact may offer us better insight into the rationality of agents who expertly perform tasks requiring sufficiently high degrees of both theoretical and practical rationality of them, and who achieve these ends virtuously: correctly, often seamlessly, reflectively when needed, and achieving enough harmony between thought and deed in their endeavors.

References

Alfano, Mark. 2012. "Expanding the Situationist Challenge to Responsibilist Virtue Epistemology." *Philosophical Quarterly* 62 (247): 223–49.
https://doi.org/10.1111/j.1467-9213.2011.00016.x

Foot, Philippa. 2001. *Natural Goodness.* Oxford: Oxford University Press.
https://doi.org/10.1093/0198235089.001.0001

Goldie, Peter. 2004. "Emotion, Reason, and Virtue." In: *Emotion, Evolution, and Rationality*, edited by Dylan Evans and Pierre Cruse, 249–68. Oxford:

Oxford University Press. https://doi.org/10.1093/ac-prof:oso/9780198528975.003.0013

Grigore, Nora. 2019. "On Why There is a Problem of Supererogation." *Philosophia* 47 (4): 1141–63. https://doi.org/10.1007%2Fs11406-018-0045-z

Grimm, St. R. 2001. "Ernest Sosa, Knowledge, and Understanding." *Philosophical Studies* 106 (3): 171–91. https://doi.org/10.1023/A:1013354326246

Harman, Gilbert. 1999. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99 (1): 315–31. https://doi.org/10.1111/1467-9264.00062

Hursthouse, Rosalind. 1999. "Right Action." In *On Virtue Ethics,* 25–42. Oxford: Oxford University Press. https://doi.org/10.1093/0199247994.003.0002

Kornblith, Hilary. 2010. "What Reflective Endorsement Cannot Do." *Philosophy and Phenomenological Research* 80 (1): 1–19. https://doi.org/10.1111/j.1933-1592.2009.00307.x

Korsgaard, Christine M. 2009. "Agency and Identity." In *Self-Constitution: Agency, Identity, and Integrity*, 1–26. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199552795.003.0001

Korsgaard, Christine M. 1996. "Reflective Endorsement." In *The Sources of Normativity*, 49–89. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511554476.004

Mărăşoiu, Andrei Ionuţ. 2018. "Wisdom and Reason." *Croatian Journal of Philosophy* 18 (2): 367–74.

Railton, Peter. 2011. "Two Cheers for Virtue: or, Might Virtue Be Habit Forming?" In *Oxford Studies in Normative Ethics* 1, edited by Mark Timmons, 295–330. https://doi.org/10.1093/acprof:oso/9780199693269.003.0012

Sosa, Ernest. 1991. "Intellectual Virtue in Perspective." In *Knowledge in Perspective*, 270–94. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511625299

Watson, Gary. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24 (2): 227–48. https://doi.org/10.5840/philtopics199624222

Zagzebski, Linda. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174763

RESEARCH ARTICLE

The Judge-Dependence of Aesthetic and Moral Judgement

Max Kölbel*

Abstract: This paper develops an account of judge-dependence, conceived of as a generalization of the better known notion of response-dependence. It then solves a number of problems for the view that aesthetic judgements are judge-dependent in this sense. Finally, a parallel case for the judge-dependence of moral judgement is assessed.

Keywords: Anti-realism; aesthetic judgement; judge-dependence; moral judgement; Kant; Hume.

In aesthetics and metaethics, as elsewhere, the terms "realism" and "anti-realism" are associated with a number of distinct worries. They may include worries about

(i)   whether entities of a certain sort exist;

(ii)   whether these entities, if they exist, are somehow basic (explanatorily, reductively, ontologically, or whatever);

(iii)   whether judgements involving or concerning these entities are truth-apt; or

(iv)   whether such judgements are subject-independent or objective (in some relevant sense).

*   University of Vienna
    ✎ Department of Philosophy, University of Vienna, Universitätsstraße 7, 1010 Wien
    ✉ max.koelbel@univie.ac.at

In this paper, I shall follow Crispin Wright's (1992) advice to distinguish these worries and then to worry separately. My focus will be on (iv).

There are some reasons why I want to focus on (iv) separately. With the increasing popularity of deflationism about truth the question about truth-aptness, (iii), has become increasingly obsolete—at least as a worry by itself. Or perhaps the worry will now usually not be whether a given topic is truth-apt, but, say, whether truth can be construed as "robust" truth, as provability, knowability or whether it can transcend possible proof and knowledge. The situation is similar with respect to (i): nowadays, before we can decide whether entities of a certain sort exist, we have to specify the sense of existence we have in mind. Or perhaps we have to decide whether we want to be universalists or nihilists about the composition of objects. Question (ii) is also affected by some recent debates: under the heading of "metaphysical grounding" theorists discuss which forms of dependence are ontologically significant.

Thus questions (i) through (iii) are under a kind of foundational dispute: it is not completely clear what is at stake in them. The same goes for question (iv), however, question (iv) seems especially relevant in the case of aesthetic and moral thought, so clarifying and attempting to answer this question seems a good starting point when worrying about the metaphysical status of moral and aesthetic value. This paper will contribute primarily to explaining one conception of subject dependence—"judge-dependence"—and explore its prospects in the realm of aesthetic and moral judgements.

In §1, I shall explain the general idea of what I want to call the "judge-dependence" of aesthetic concepts, a generalization of the more familiar phenomenon of response-dependence. In §2, I explain three phenomena that might be thought, and have been thought, to present a problem for the view that aesthetic judgements are judge-dependent. In §§ 3 and 4, I explain how in the case of aesthetic judgements these phenomena can be saved, or explained away, on the judge dependence view. In §5, I shall consider whether a parallel case can be made for moral judgements. Here I shall suggest, that while some of the same motivations are available in the moral case, there are additional features that make the argument much less clear cut.

1. Judge-dependence

It is sometimes said that aesthetic judgements or judgements of taste are distinctively subjective or subject-dependent in the sense of being based on certain types of experience of the person making the judgement. It is debatable whether this characterization by itself really sets aesthetic judgements apart from other, empirical judgements. Why should one think that it is distinctive of aesthetic judgements that they can be made, or are perhaps canonically made, on the basis of certain types of experience? This would seem to be equally true of many empirical judgements that would not normally be regarded as "aesthetic judgements". To say something distinctive about aesthetic judgement, one needs to add specificity. I want to add two more specific points: (i) that the experience is a kind of pleasure or displeasure, and (ii) that the disposition to have these experiences, and on which the correctness of an aesthetic judgement is grounded, often varies from judge to judge.

(i) Kant thought that the experiences in question were experiences of *pleasure* or *displeasure*. And he went on to distinguish those aesthetic judgements where the pleasure in question is linked to the satisfaction of some desire (judgements of the agreeable), and those where it is not (judgements of beauty). The important aspect of these Kantian classifications for our purposes will be the observation that the experiences on which certain types of judgement are based are experiences that are intrinsically valuable or desirable (or in the case of displeasure: intrinsically worth avoiding). I will come back to this first point in the next section.

(ii) The second specific point that needs to be added in order to say something distinctive that sets aesthetic judgements apart from other judgements is a certain point about how different judges may be differently disposed to have the experiences in question, and what these differences mean for the correctness of such judgements—a feature I shall call "judge-dependence". To explain judge-dependence, I first have to take a step back and say something more generally about judgements and concepts.

The characterization of aesthetic judgements that I am after is a characterization in terms of the distinctive representational content of aesthetic judgements, and this distinctive representational content in turn will

depend on certain distinctive features of the *concepts* that are employed in making aesthetic judgements. In general, the representational content of any judgement captures the representational aspects of the judgement. Judgements represent the world as being a certain way and thus help us to conduct ourselves in ways beneficial to us—as long as they represent *correctly*. (For reasons mentioned in the preamble, I am here using the neutral term "correctly", rather than "truly".) Thus, judgements are not arbitrary or unconstrained, but rather aim at correct representation. For example, a judgement that the shop sells lightbulbs is incorrect if the shop does not sell lightbulbs. There is, thus, a reason to avoid making that judgement unless the shop does sell lightbulbs. There may be judgements that are correct even if the shop does not sell lightbulbs. But they will not be judgements *that the shop sells lightbulbs*—they will not be judgements with that particular representational content.

In my view, the content of a judgement, together with the way things are, does not always determine whether the judgement is correct. Sometimes, it depends on who is making the judgement, and when.[1] For example, I want to allow the content that *it is Tuesday* to be a content that it is correct to believe only on Tuesdays; and the content that *one is a philosopher* to be correctly believed only when the believer him or herself is a philosopher. I even want to allow a content that it is only correct for David Hume to believe: the content that *one is David Hume*—a content famously believed in error by Heimson (cf. Lewis 1979a). To be sure, some contents do not exhibit this variability, namely those that have been called "boring" or "portable" (respectively by Andy Egan [2007] and myself [2013, 2014]). If anyone at any time believes such a portable content and does so correctly (or incorrectly), then anyone else who believes the same content at any other time will also do so correctly (incorrectly). For example, the content that a particular shop sells light bulbs in July 2020 is such a portable or boring content.

In the current context it is not important whether we allow such centered or *de se* contents of representation, or whether all contents must be portable. I want what I say about aesthetic and moral judgements to be

---

[1] In other words, I want to construe representational contents in general as "centered" or "perspectival" contents, see e.g. (Egan 2007, 2012; and Kölbel 2015).

neutral with respect to the question of whether we allow centered contents. Whenever I say that there is a common content that it is correct to believe by some at some time, and not correct to believe by others or at other times, those who stipulate that all contents are portable (e.g. Frege) will say that there are as many distinct contents of judgement as there are different thinker-time pairs with which correctness can vary. Thus, while I could articulate what I want to say by saying that the *contents* of aesthetic judgement are typically judge-dependent or non-portable, I shall attempt to find a more ecumenical formulation. The ecumenical formulation focusses on distinctively aesthetic and moral *concepts*, i.e. on the constituents of such contents, and it thus bypasses the mentioned issues about contents in general.

The content of a particular judgement (and the corresponding conditions under which it is correct to make a judgement with that content) depends on the *concepts* involved, i.e. the concepts that constitute the judgement's content, which are the concepts employed by the judge in making that judgement. There may be various reasons why it is interesting to say that the contents of judgements are in some sense constituted or determined by concepts.[2] In the current context, the main reason is that I want to say that aesthetic judgements are judge-dependent because they involve judge-dependent concepts.

Just as judgements are correct depending on their content (and perhaps on who judged the content and when), concepts can be *correctly* or *incorrectly* applied to things. Thus, the concept *lightbulb* is correctly applied to all, and only to lightbulbs. Thinkers who are competent with the concept of a lightbulb have the disposition to apply the concept all and only to lightbulbs, at least when conditions are good. (The exact correlation between competence and correctness of judgement, including the specification

---

[2]   Amongst them a desire to explain the large repertoire of potential contents of judgement that are available to individual thinkers, and a desire to explain how different thinkers manage to make judgements with similar or identical content in such a large range of cases, thus being well positioned to *communicate* their judgements to one another. In other words, the compositional structure of contents is hoped to mirror the structure of individuals' competences as well as the similarity of the competences of individuals in a group.

of "good conditions," is a complicated issue, which I leave aside here.) Now, some concepts, like that of a lightbulb, are such that their range of correct application is the same for everyone at every time. Other concepts, like that of *here*, *yesterday* or *one's own uncle* are not: the range of things to which these concepts are correctly applicable depends on where, when or by whom they are applied. Those who, under good conditions, do not apply the concept *here* only to the place where they are, are not competent with the concept. Those who, under good conditions, apply the concept *one's own uncle* to things other than their own uncle thereby show their incompetence with the concept.[3]

I am stressing the connection between the range of correct application of a concept (for a judge at a time) and the tendency of a competent speaker to apply the concept all and only to things in that range when conditions for judgement are good. I am doing this because we can use this connection—within dialectical limits—when trying to show what the range of correct application of a given concept is. We can do so by observing what kind of behaviour, i.e. what kind of application of the concept, would count as a sign of incompetence with the concept.

In particular, I want to argue that aesthetic concepts impose conditions on their range of correct application that concern the response the judge must be disposed to have to things in that range. Thus, the concept of beauty (or ugliness) can be correctly applied by a thinker to an object if and only if that thinker is disposed, under good conditions, to respond to that object with a certain kind of pleasure (or displeasure). Kant would say: *disinterested* (dis)pleasure, i.e. (dis)pleasure that does not come from the (lack of) satisfaction of some bodily desire, such as hunger, thirst or a sexual urge. I am not sure whether Kant's point about disinterested pleasure correctly marks the narrowly aesthetic (the beautiful) from the agreeable or merely pleasant or nice. But I do think it is a correct observation that those who apply the concept of beauty to objects that fail to cause a certain

---

[3]   It may be worth clarifying that when I am speaking about concept applications, I am speaking about simple judgments in which the concept is applied to an object (i.e. not, for example, applications that occur within a disjunctive judgement. Since my topic is judgement and not linguistic performances, there is no issue of insincere or non-literal applications of concepts.

pleasurable experience in them, even when the conditions for concept application are good, thereby manifest incompetence with the concept of beauty or ugliness.[4]

There are many concepts that require for their correct application that the judge be disposed to respond in a certain way. I have already mentioned concepts like *agreeable*, *pleasant*, *nice*, which are linked to pleasant experiences of different kinds. But there are many others that suggest themselves: *attractive*, *funny*, *disgusting*, *seems F* (for variable *F*), *soporific*, *thrilling*, *horrific*, *great*, *fantastic*, etc. I believe that all (or most) of these concepts have conditions of correct application that follow roughly the following schema:

(JD)   For all subjects *s* and all objects *o*: it is correct for *s* to apply C to *o* if and only if *s* is disposed to have experiential response R to *o* under favourable conditions.

---

[4]   There are some niceties here in specifying the disposition. Thus, one may ask whether a blind person cannot correctly apply the concept of beauty to a (visual) painting, because the painting does not cause the relevant experience in the blind person, not even under favourable conditions. There are various ways of dealing with such cases. One possibility is to say that the blind person never is in favourable conditions, thus can still have the disposition. But if we said that he or she has the disposition vacuously, then we'd also be forced to say that they have the disposition to respond with a certain displeasure under favourable conditions. Thus, it would be correct for that person to judge that the object is beautiful, and also that it is ugly. Another possibility is to say that the blind person is indeed partially incompetent with the concept, at least when visual beauty is concerned. Such a view would seem to have to decide whether there is such a thing as a blind person judging correctly that the painting is beautiful, and if so, on whose response to the painting this depends—or are these uses some kind of "inverted comma uses," which do not manifest employment of the concept, but rather a meta-reflection on how it is correct to apply it?

Another possible difficulty is that it may be claimed that some forms of beauty do not cause pleasure or a pleasurable experience. I am inclined to respond that in so far as this is so, we are talking about an extended concept of beauty that may or may not be culturally or historically related to beauty in the core sense I have tried to characterize.

If we instantiate the schema for C = the concept of beauty, we should put something like "aesthetic appreciation" as the response R. If C = the concept of funniness, then the response should be amusement. If C = the concept of seeming large, then the response should be an inclination (in the absence of evidence to the contrary) to believe that o is large. If C = the concept of being great (or fantastic), the response should be enthusiasm.

Evidence for these suggested conditions of correct application is in each case the observation that if someone applies concept C to objects that do not cause response R in him or her, even under favourable conditions, then this is an indication that this person is not fully competent with concept C. It may be difficult or controversial to decide in a specific case what the conditions of correct application of a given concept are. Furthermore, it may be controversial whether concepts of this sort correspond to properties, i.e. properties that have corresponding application conditions. However, it ought not to be controversial, that there could be concepts of this sort— that there is nothing incoherent about such concepts per se. Let us call such concepts "judge-dependent".

The range of correct application of a concept need not vary from judge to judge, or from time to time just because the concept is judge-dependent (i.e. just because it conforms to schema (JD)). For example, if the potential judges all happen to be similar in their relevant dispositions to respond under favourable conditions, then that would prevent any divergences in ranges of correct application. Arguably, colour concepts, such as the concept of being red, instantiate schema (JD). Thus failure to apply the concept of redness all and only to those things that cause a certain characteristic experiential response in one,[5] when conditions are favourable, will be taken to indicate incompetence.[6] So colour concepts would appear to be a special

---

[5] NB: if we do not want to assume the possibility of making interpersonal comparisons of subjective experiential states, we may not be able to specify the characteristic experiential response in intrinsic terms but may need a relational specification.

[6] As in the case of the concept of beauty, there is a good question of what one should say about those who are constitutionally unable to have the relevant response, but who nevertheless arguably use the concept—a blind person may apply the concept of redness. Many are tempted to say that visual detection of colour is somehow

case of judge-dependence. However, I would argue that the relevant response R is here specifiable only in ways that makes sure that the ranges of correct application of different judges at different times do not only *happen* to coincide, but that they *must* coincide as a matter of conceptual necessity.[7] For example, on this view it would be a *conceptual* truth that if different competent judges are in favourable conditions, the objects to which they will apply the concept of redness will be (roughly) the same. Colour concepts are judge-dependent, but at the same time they are objective: correct application will not depend on who is applying the concept to a given case or when.[8]

Aesthetic concepts, however, do not seem to exhibit judge-dependence in the special way in which colour concepts do: the ranges of objects to

---

essential to the possession of colour concepts, so that blind people do not have full competence with colour concepts. Another way out would be to allow that under "favourable conditions" blind people would after all have the experiences in question. They are simply never in favourable conditions.

Another problematic aspect is that it may be difficult to identify the response in question. If we say that it is the type of experience brought about by red objects under favourable conditions, then the explanation is blatantly circular. Moreover, it has been argued that there could be systematic qualia-inversion, so that the way one person experiences red objects is the way another person experiences green ones. If colour-qualia inversion is indeed possible, then I tend to think that the type of response mentioned in the relevant instance of (JD) cannot be specified in an intrinsic way, but rather has to be specified, for example, with reference to a list of paradigm red things or with reference to the concept-using practice. In the latter case, R = the type of visual experience brought about under good conditions by those objects that are classified as "red" in the community in question. Systematic inversion of experiences specified in this way within the community in question is not a coherent possibility. However, inversion of experiences characterized in another, perhaps intrinsic or phenomenological way, would be.

[7]   For example, R could be specified as the colour experience caused by … [here follows a list of paradigm red things] under favourable conditions. This leaves open the possibility that there are phenomenal differences between individuals, see last footnote. Human physiology will ensure that the phenomenology is relatively constant in the same individual over time and under favourable conditions.

[8]   I am simplifying somewhat by ignoring certain subtle variations, with for example sex or age, in the application conditions of colour terms in certain contexts.

which judges are disposed to respond with aesthetic appreciation under favourable conditions will vary from judge to judge. The variation is quite regular, and it is well-known to exist. In my view, such variation is not taken to show that some judges are less competent, or that the conditions under which some are operating, are not favourable. Thus, aesthetic concepts, such as the concept of beauty, do not prescribe at the conceptual level that ranges of correct application must coincide between judges.

I realize that this last claim may be controversial, especially for the paradigmatic aesthetic concepts such as beauty and ugliness. It will be one of the tasks of §4 to explain away some of the phenomena that make it appear controversial. For the moment, however, let us settle for the claim that at least many of the judge-dependent concepts mentioned on the last page (e.g. *agreeable*, *pleasant*, *attractive*, *funny*, *disgusting*) are judge-dependent in such a way that the ranges of correct application of these concepts can vary from judge to judge or time to time.

Let me mention, finally, a kind of concept that may easily be confused with a judge-dependent concept. Some (at least possible) concepts depend for their correct application not on the response the judge would have under favourable conditions, but rather on the response an ideal judge would have under favourable conditions. The conditions of correct application of these concepts can be articulated like this:

(IJ)   For all subjects *s* and all objects *o*: it is correct for *s* to apply C to *o* if and only if an ideal judge would be disposed to have experiential response R to *o* under favourable conditions.

One difference one notes immediately is that concepts instantiating schema (IJ) will not be judge-dependent, because it no longer depends on a feature *of a judge*, whether it is correct for *that judge* to apply the concept. Rather, it depends on a feature of the ideal judge. This means, on the face of it, that (IJ)-type concepts are not judge-dependent.[9]

Another important difference between schema (IJ) and (JD) is that the link with rational motivation that arises if response R is something intrinsically desirable, is removed. If an object causes a pleasurable experience in

---

[9]   There is a way in which (IJ)-type concepts may nevertheless be judge-dependent, and this is if it depends on the judge what an ideal judge would be like.

me under favourable conditions, then this will give me some reason to prefer to encounter that object rather than others at least when conditions are likely to be favourable. If the disposition is not my own, but that of an ideal judge, then I no longer have that reason—unless perhaps I believe that I am similar to the ideal judge in the relevant respects, or I have reason to have preferences similar to those of an ideal judge. In any case, the reason would no longer be directly entailed by my judgement, but rather depend on further beliefs or preferences.

The concept of beauty, and perhaps other more narrowly aesthetic concepts, may appear to some to be of the (IJ) type. Remarks like: "This may be beautiful, but I don't appreciate it." or "I prefer the one that is less beautiful.", may suggest this. Moreover, we often have strong views about what sense of taste one ought to have, aesthetic apreciation has an important social role. This too may promote the impression that aesthetic concepts are of the (IJ) variety (more on this below).

Let me summarise, then, the results that I want to take from this section: aesthetic concepts are a kind of judge-dependent concept where the response in question is an experience of pleasure or displeasure of some sort. Moreover, judges regularly vary in their dispositions to have these experiences in response to objects, and as a result the range of correct application of aesthetic concepts may vary from person to person, or from time to time.

2. Three puzzling phenomena

The view outlined, that aesthetic concepts and judgements are judge-dependent, even though it can be motivated and supported in the way indicated, also faces some apparently disconfirming phenomena, i.e. phenomena that seem to conflict with what the view predicts. Some of these phenomena are quite well-known and need to be addressed. I shall here distinguish three such phenomena, two of them broadly taken from Hume, and one taken from Kant.

The view that aesthetic judgements are judge-dependent shares some key features with Hume's sentimentalism, in that the correctness of such a judgement will depend on the response-profile a judge happens to have.

Hume mentions two problems for this sort of view, in the form of tensions between the view and the way we commonly treat aesthetic matters.

First, if aesthetic judgements are governed by norms of correctness as specified in (JD), we should expect that no aesthetic judgement should be incorrect or wrong in itself, i.e. just in virtue of its content: if the judge has the appropriate dispositions, then an aesthetic judgement with any (contingent) content could be correct for that judge.[10] However, it seems as if we can easily come up with examples of aesthetic judgements that are, in Hume's words, "absurd and ridiculous" (Hume 1757, §9). One of Hume's examples is the judgement that there is "an equality of genius and elegance between Ogilvy and Milton" (Hume 1757, §9). The problem for Hume as well as for the proponent of the judge-dependence of aesthetic concepts is that for people with the right sentiments or dispositions, i.e. a disposition to favour Ogilvy in the relevant way, it should be correct to judge in this way. So there should not be contents of aesthetic judgement that automatically count as absurd or ridiculous.

The second problem is not specifically raised by Hume as a problem for his own sentimentalism, but much of his discussion in "Of the Standard of Taste" can be seen as an attempt to resolve this problem for his view. This second problem consists in the apparent existence of expertise in aesthetic matters, in other words the apparent existence of thinkers with a superior faculty of judgement concerning aesthetic matters. Again, if aesthetic judgements are Humean sentiments rather than judgements of the understanding, then it is not clear why anyone should qualify as an expert on aesthetic matters. The same goes for the view that aesthetic judgements are judge-dependent, as outlined in the last section: if the correctness of an aesthetic judgement depends on the judge's profile of dispositions to respond to objects of aesthetic appraisal, and if judges tend to differ significantly in their response profiles, then why should we elevate some judges with the predicate of "aesthetic expert," i.e. as someone superior in their ability to judge

---

[10]   Another question would be whether the dispositions required for the correctness of the judgement are normal or admirable. But that is independent of the correctness of the judgement, if it is judge-dependent in the way outlined. I will come back to evaluations of the response dispositions of judges below.

on aesthetic matters, someone on whom we can rely independently of our own experiences, and perhaps even against our own experiences.

A third puzzling phenomenon can be found in Kant's *Critique of Judgement*. Kant claims, not implausibly, that aesthetic judgements involve a certain claim to "universal validity." He says that

> when [someone] puts a thing on a pedestal and calls it beautiful, he demands the same delight from others. […] he *demands* this agreement of them. He blames them if they judge differently. (Kant 1790, 52)

This, too, suggests that aesthetic concepts can't be judge-dependent in the way I have been suggesting. For if the correctness of an aesthetic judgement with a certain content depended on features of a judge that he may not share with other judges, then it would seem unreasonable to expect or even demand of others that they judge in line with one's own judgement. Others may not have the same, or relevantly similar, response profiles and therefore a content that it is correct for me to accept may not be for others. So the claim to universal validity that Kant thinks we are making would seem to be quite confused.

To sum up, then, the three puzzling phenomena to be discussed are the following. First, we regard aesthetic judgements with certain contents as mistakes independently of the judge's experiential responses (e.g. "Ogilvy is better than Milton"). Secondly, some people seem to enjoy the status of experts in matters aesthetic. And finally, we seem to be making a claim to universal validity when making aesthetic claims.

3. Three ways of being a good aesthetic judge

In order to show how these phenomena can be explained (or explained away), let me begin by distinguishing two factors on which the correctness of an aesthetic judgement depends according to the thesis of judge dependence. Recall the schematic thesis of judge-dependence:

(JD)   For all subjects s and all objects o: it is correct for s to apply C to o if and only if s is disposed to have experiential response R to o under favourable conditions.

I claimed that aesthetic concepts are judge-dependent in the sense of being governed by instances of (JD) where the response R is some form of aesthetic appreciation (a kind of pleasure or displeasure), and moreover, that dispositions to response R can vary from judge to judge (and possibly from time to time). This means that the correctness of applying an aesthetic concept to a given object depends on two factors being aligned with one another: (a) the relevant (response-causing) features of the object and (b) the judge's dispositions to respond (at the time of applying the concept), i.e. the judge's response-profile. If the object has features that would elicit the required response in the judge under favourable conditions, then the concept was correctly applied.

There are different methods for ensuring that one apply aesthetic concepts correctly, in line with (JD). In general, if one wants to find out whether an object has a certain disposition, i.e. would respond in a certain way under certain circumstances, one can put the object into (approximately) such circumstances, and see whether it responds in that way. Thus, to use a paradigmatic example of a disposition: to see whether a vase is fragile, I can drop it onto a hard surface and see whether it breaks. So, what I want to call the "canonical method" of arriving at an aesthetic judgement involves attempting to provoke the relevant response in oneself by exposing oneself to the object under approximately favourable conditions (cf. Lewis 1989). Thus, if I look at the painting in good lighting conditions when I myself am also in a favourable state (e.g. I am wearing my glasses, if needed, I am awake, not intoxicated, not impaired by distractions or prejudices, etc), then I can assess in a direct way whether I am experiencing the required response. In fact, I do not even need to become aware of my response and then *infer* that the concept can/cannot be correctly applied. Rather, judges will often use the canonical method in an "unreflective" way, merely coming to apply the concept reliably in line with the experience the object elicited in them, without any explicit thought about their own experiential response.

But there may also be non-canonical, more indirect methods. In general, in order to find out whether an object has a disposition to react in a certain way under certain conditions, one does not *need* to put the object into (approximately) those conditions. One can also apply general knowledge

about objects of this type to make an inferentially justified judgement. For example, one does not need to attempt to break a vase in order to find out whether it is fragile. Instead one can judge the vase to be fragile for example on the basis of a belief that it is made of porcellain or glass etc, and the general belief that vases made from these materials are fragile. Similarly, there are indirect ways to find out about those features of objects of aesthetic appraisal that are responsible for their affecting aesthetic judges in the way they do (see (a) above). To use a simple example, perhaps drinking coffee made from a certain type of blend of coffee beans, using a certain method of coffee-making, tends to cause gustatory pleasure in me in favourable conditions. So, I can rationally judge a particular cup of coffee to be delicious merely on the basis of my view that it was prepared in this way from this type of blend. Or perhaps paintings by a certain artist generally tend to cause a response of the right kind of disinterested pleasure in me. So I might judge that a particular painting is beautiful merely on the basis of my view that it was made by this artist, without having seen the painting.[11]

Now, this general picture of how the application of aesthetic concepts can be justified, i.e. how thinkers can ensure they judge correctly, leaves room for a variety of ways in which one can assess the competence of aesthetic judges, and corresponding ways in which one aesthetic judge can be superior to another. To begin with, we can measure the likelihood to which a judge's aesthetic judgements are correct, i.e. the likelihood that the judge will apply an aesthetic concept only to those objects which would elicit the appropriate response in him under favourable conditions. We can also make corresponding comparisons: one judge may be more likely than another to apply an aesthetic concept only to things that would in fact provoke the

---

[11]   There is a certain tradition in Aesthetics, following e.g. Wollheim (1980, 156) that denies the legitimacy and possibility of what I have called "non-canonical methods" in the realm of properly aesthetic judgement, or perhaps denies that the resulting judgements are genuine applications of aesthetic concepts. I will here assume without argument that this view is wrong. To the small extent to which my explanation of phenomenon P2 below depends on this assumption, my conclusions are merely conditional. See, e.g., Wollheim (1980); Hopkins (2001); Budd (2003); Franzén (2018) or Dinges and Zakkou (2020) for discussion.

right response *in him or herself* under favourable conditions. Let us call this way of assessing the reliability of a judge "likelihood of correct judgement."

This is an important dimension of assessment, however it does not *by itself* explain all aspects of our practice of treating some people as experts. Experts are people we can rely upon, but those who are likely to judge correctly need not therefore be people whose judgements we can simply rely upon. More precisely, we cannot just accept the contents they accept and thereby be likely to judge correctly ourselves. Such a conclusion would require some further assumption, for example that the reliably correct judge has dispositions to respond that are relevantly similar to our own.

However, the general picture allows us to distinguish further dimensions of assessing aesthetic judges. We can also assess the abilities of a judge to discern those features of objects of aesthetic appraisal that underlie our experiential response to them. For example, whether a given taster experiences gustatory pleasure upon tasting a dish causally depends on many overlapping factors, such as the composition of the food, the ingredients, their origin, their composition, the way the meal was prepared, by whom, using which tools and procedures, etc. Similarly, the response an aesthetic appraiser experiences when seeing a painting under favourable conditions causally depends on many underlying features of the painting, such as the distribution and structure of the paint on its surface, the composition of the painting, the painter who made it, the period in which it was made, the school of painting to which it belongs, etc. Independently of whether a judge is good at applying an aesthetic concept correctly, i.e. only to those things which elicit the relevant response in him or herself, a judge can be more or less reliable at recognizing and discerning these underlying factors. For example, a culinary expert may be very knowledgeable about various types of raw ingredients, their provenance and resulting qualities, the possibilities of using these ingredients, techniques for doing so and the resulting flavours, and she might combine this knowledge with the ability to detect all these features gustatorily, etc. An expert on paintings may be a superior judge on the provenance of paintings, on their paint distribution and the resulting experiential effects, on painters and their characteristic techniques and resulting experiential effects, on the cultural context (their own and the painter's) and its effect on the responses provoked by the painting etc.

If we introduce the umbrella term "underlying features" for all the causal factors (including extrinsic, contextual factors) that are responsible for judges' experiential responses to objects of aesthetic appraisal, then we can say that some thinkers are excellent aesthetic judges in the sense that they have excellent abilities to discern and recognize underlying features. The knowledge of, and ability to discriminate, underlying features can be sub-divided into many different types of such knowledge and abilities. There is general knowledge (e.g. paintings from this period are generally produced with such and such technique) and particular knowledge (e.g. this painting exhibits a composition typical of such and such period). There are also recognitional skills (such as the skill to recognize a type of composition by looking at a painting) and recognitional skills can be based on powers of discrimination with varying degrees of fineness (one judge may be able to discern differences that another judge is not able to notice). All the recognitional skills can be orthogonally divided into those that are based on direct experience of the objects of aesthetic appraisal (as in a direct visual appreciation of a painting, or sampling of a food), and those that are based on indirect kinds of evidence (e.g. via general knowledge, via measurement devices, etc). There is, finally, the associated skill of being able to exercise the above skills and the above knowledge reliably in the face of sources of distraction, such as prejudice, comparative environment, social pressure etc. Thus, the knowledge of, and abilities to discern, underlying features encompasses a very wide range of competence. Let me refer to them all summarily as "competence concerning underlying features."

Excellent competences concerning underlying features may or may not be accompanied by the tendency to apply aesthetic concepts correctly (i.e. in line with one's own dispositions to respond), even if, presumably, there will be many correlations. However, a judge with superior competences concerning underlying features will be a valuable source of information to others. It is quite obvious how such a judge's general and particular knowledge of underlying features can be useful when articulated in non-aesthetic (i.e. non-judge-dependent) terms, for example by saying "Painter X tends to use technique Y." But even if such a judge's discernments of underlying features are articulated in judge-dependent terms (e.g. "Painter X paints better still lifes than painter Y."), an audience who is less competent concerning

underlying features may nevertheless rely on the superior judge's detection of a relevant difference.

I have distinguished two dimensions of assessing aesthetic judges: likelihood of correctness and competences concerning underlying features. Both these types of assessment are arguably objective: it does not depend on one's aesthetic standard or one's taste whether one can correctly regard a judge as more or less likely to be correct, better or worse at telling apart underlying features.[12] The third type of assessment that I want to distinguish differs in this respect. I am speaking about an evaluative assessment of a judge's response profile itself, i.e. their dispositions to respond to objects of aesthetic appraisal. For example, one might regard a disposition to respond more favourably to Hip Hop than to Heavy Metal as an inferior response profile. Another might regard a taste that evaluates either of the two positively as as a bad standard of taste. These evaluations of the response profiles, or tastes, of judges seem themselves to depend on evaluative standards. However, the point of distinguishing this third type of assessment does not depend on this. The important point is rather that an evaluation of a judge's response profile, i.e. his or her taste, is independent of an evaluation of that judge's likelihood of correctness and that judge's competences concerning underlying features. For example, one might recognise someone as excellent in the latter two respects while regarding their taste as inferior.

Whether or not evaluations of response profiles are an objective matter, they do not affect the conceptual competence of judges. According to the thesis of judge-dependence, correct application of the concept requires that one's judgements are in line with one's dispositions to respond. But this does not require any particular type of disposition to respond. A judge could have any response profile whatsoever and still be a fully competent user of aesthetic concepts. If we regard one response profile as superior to another that is not an evaluation of a judge's competence as a concept user, or of his or her likelihood of using these concepts correctly.

Humean sentiments are spontaneous and beyond rational control. So, in Humean terms, a judge's response profile, their dispositions to respond to

---

[12]   Except perhaps because likelihood assessments are judge dependent, as claimed by subjectivists about probability, e.g. Keynesians or Bayesians.

objects of aesthetic appraisal, is something beyond direct voluntary or rational control. At best, I can deliberately undergo a training programme that will gradually alter my dispositions. Just as I cannot directly control the phenomenal quality of my visual experiences when looking at certain objects, I cannot directly control my aesthetic responses to objects of aesthetic appraisal.

Perhaps this hard Humean line needs to be softened somewhat when we are talking about more refined aesthetic responses. These may be more susceptible to cognitive influences than, for example, the brute phenomenal responses involved in sensory perception. One might argue, for example, that a thinker's dispositions to respond to a work of art may change if they are told that it was made with a certain technique, or even that it is merely a very good copy of the original. We may reply that this is because the technique (and the status as original) is one of the underlying features that are causally responsible for the characteristic experiential response. Of course, to the extent to which we admit such influences, we are construing the "experiential" response as more "intellectual." If we assume such cognitive influences, i.e. that a judge's response profile is not merely a Humean sentiment beyond all rational control, we may be able to construe our evaluations of judges' response profiles partly in terms of their level of informedness.

However, even with this concession of a limited rational evaluability of response profiles (as more or less informed), the comparative evaluation of response profiles remains independent of an evaluation of judges' likelihood of correctness or their competences concerning underlying features.

Before moving on, it is worth pointing out the importance most of us attach to comparing the response profiles of different judges, i.e. their taste. People's aesthetic responses seem to be associated with certain social groups, and members of such groups may construe their own "identity" partly in terms of their taste. For example, some seem to regard a taste that favours paintings by Jack Vettriano as vulgar, while others regard those as snobbish who do not appreciate Vettriano's art.

To sum up: we can distinguish three ways of assessing aesthetic judges: (a) likelihood of judging correctly; (b) competences concerning underlying features; and (c) the quality of the response profile (which may itself be

a judge dependent issue). Despite some correlations between the three, they are independent of one another.

4. The puzzles addressed

Let me briefly recapitulate the three puzzling phenomena that threaten the response-dependence account, before I begin addressing them:

P1   Aesthetic judgements with certain contents just seem wrong ("absurd and ridiculous"): we regard them as mistakes no matter what the personal features of the judge may be.

P2   There seem to be experts regarding aesthetic questions (wine experts, art critics, etc).

P3   When making aesthetic claims or judgements, we seem to be making some kind of claim to universal validity (expecting and demanding of others that they judge likewise).

How can phenomenon P1 be explained, i.e. the fact that aesthetic judgements with certain contents seem mistaken for any judge? Hume's example is the judgement that Ogilvy is equal in elegance and genius to Milton. What would be a contemporary example? The judgement that *Conan the Barbarian* is equal to *Ulysses* perhaps? The sphere of gustatory taste might provide better examples: the judgement that rotten eggs are delicious seems to be a mistake for any judge. Or perhaps the judgement that Florence is an ugly city can serve as another example.

It seems to me that such examples may seem to be convincing for at least two different types of reason. On the one hand, they are judgements that it would be correct to make only for a judge with a response profile that we would evaluate negatively. Thus, we may think that any response profile which yields a negative aesthetic response to Florence, or a gustatory response of delight to rotten eggs, must be a sick or perverted response profile. If this is the source of the impression that such judgements are mistaken independently of who the judge is, then the judge dependence account can clearly cope: the basis of the assessment as mistaken is a negative assessment of reponse profiles of a certain type (namely those relative

to which the judgement would be correct). Perhaps no normal human, for biological reasons, will respond with gustatory delight to rotten eggs (although I believe there may be exceptions). Perhaps any aesthetic standard that rates Florence as ugly (or that equates the quality of *Conan the Barbarian* with *Ulysses*) must be condemned as perverted. If so, this would explain the view that such judgements are always mistaken: either the judgement is not in line with the response profile, or the response profile itself is a mistake: it is not a set of dispositions one should have. This explains why some may have the impression that there must be a mistake in regarding Milton and Ogilvy as equal, whatever the dispositions of the judge are. Once we concede their negative evaluation of any standard of taste that would treat Ogilvy and Milton as equal, the impression is correct: either a judge has an ok response profile, in which case he judged incorrectly (in the sense of not judging in line with the norms spelled out in the relevant instance of (JD)). Or she has a response profile that it is a mistake to have, in which case, perhaps, she may be judging in line with the norms of (JD), but is guilty of having the wrong response profile. So, the impression that judgements with such a content are a mistake to make, whoever the judge may be, is correct.

It is important to stress, however, that this take on P1 still leaves open that some judges might have the condemned response profiles, so that they would be applying the relevant aesthetic concepts correctly if they made these judgements. Thus, perhaps a judge with a strong aversion to terracotta and marble in a city would be judging correctly if judging Florence to be ugly. Fungus the Bogeyman will correctly judge rotten eggs to be delicious because they produce in him the required gustatory delight. Thus, our negative evaluations of the response profiles that would make such judgements correct do not show that these judgements cannot be made correctly in the sense of being made in line with the conceptual norms articulated in the relevant instance of (JD).

Let us move on to P2: some thinkers undoubtedly enjoy the status of aesthetic experts, as do for example wine experts or art critics. Does the judge dependence account outlined leave room for this to be a deserved status? I believe it does, though we ought to be careful to distinguish the various dimensions of assessment of aesthetic judges that I distinguished.

The status of expert can be warranted on the basis of superior competence concerning underlying features (which in turn can take many forms). But it can also be warranted to treat someone as expert because they have a type of taste (i.e. response profile) that one regards as superior and worth emulating. Superiority in terms of greater likelihood of correctness, however, can play only an indirect role in justifying expert status. Let us look at some examples.

Wine experts are usually superior to others in their competences concerning underlying features, and this is usually all that matters in this case. This superiority, as mentioned, can take many different forms, from superior abilities to discriminate flavours in canonical ways to superior general knowledge of wines (usually restricted to certain regions). Often, a judge with superior wine-tasting abilities will also have a more differentiated response-profile. But I presume that the expert status is in these cases owed usually to their competences concerning underlying features. They might also be good sources of advice because they have general knowledge of what others' palates are like.

By contrast, expert art critics, I suspect, typically acquire their status due to a combination of enhanced competences concerning underlying features (both in their discriminatory skills and in their general and applied knowledge) and due to a response profile, i.e. a taste in art, that is regarded as superior—at least by those who regard them as experts. Thus, art critics will often be more perceptive or more knowledgeable than others, but they might also be admired for their taste. To the extent to which their response profile is emulated and admired, their role of expert is similar to that of a guru: they serve as a model for others, who would like to be like them.

Likelihood of correctness will only play an indirect role when it comes to the expert status: if one thinker believes that she has a response profile similar to that of another person, who in turn has a high probability of being correct in her aesthetic judgements, then relying on their expert judgement seems to offer a quick shortcut to achieving their rate of correctness without needing any competence concerning underlying features. Thus, the usefulness of one thinker's correctness to another thinker, who wants to rely on the first's testimony, depends on the assumption that the response profiles or tastes are relevantly similar. There are, of course, further

assumptions that would allow one to draw inferences from the fact that someone with a high probability of correctness has made a certain aesthetic judgement: e.g. explicit assumptions about the thinker's response profile. But these are not a basis for an explanation of the status of aesthetic expert in the sense of P2.

Phenomenon P2, then, can be explained within the judge-dependence framework: when we regard certain judges as experts in aesthetic matters, we regard them either as especially competent concerning underlying features, or we value their taste (their set of dispositions to respond to objects of aesthetic appraisal) as worth emulating, or both.

This leaves us with phenomenon P3, the universality that Kant claims to pertain to aesthetic judgements. Does the judge dependence of aesthetic concepts make room for this claim to universal validity? I will argue that there is no room for the phenomenon exactly as stated by Kant. However, the judge-dependence account offers a ready way to explain some nearby claims that may well be the basis for thinking (erroneously) that Kant's strict universality claim is correct. This will again make use of the idea that we can assess judges in the three ways I outlined in §3.

One straightforward way in which *certain* aesthetic judgements might come with a claim to universal validity would concern the negations of the judgements that figured in phenomenon P1. The judgement that Florence is not ugly, or that rotten eggs are not delicious, are plausibly taken to be correct for any judge, simply because no-one is likely to have a response profile relative to which these are not correct. However, this will not do justice to the phenomenon as it was intended by Kant: the idea of P3 is that we expect and demand others to judge as we do not only in these special cases, that actually command widespread agreement, but that we expect and demand this *whenever* we make aesthetic judgements.

A more adequate explanation might be given in terms of our evaluations of tastes (response profiles). Even if it is largely an involuntary matter which response profiles we have, we do nevertheless attach great social significance to them. We regard people as admirable or despicable, as suitable friends or bad company, merely on the basis of their taste. We identify ourselves strongly with our own taste and regard it as an essential feature that defines who we are, and which socially significant groups we belong to.

As a result, we will regard people who change their taste too quickly, or for no good reason, as inauthentic. This sheer level of significance can explain, to some extent, why it seems that we expect and demand that others judge aesthetically as we do: on the one hand, we expect and demand them to assess the underlying features as we do. But on the other hand, we also have a strong social expectation that they have a taste similar to ours, so that if they meet the expectation, and they judge as we do, then their judgements will be correct just if ours are.

One limitation of this explanation is that it does not really account for us having this expectation, or making this demand, completely universally on everyone. Rather, the account predicts that we should not have these expectations and demands whenever we expect others—perhaps members of a different group—to have quite different tastes. So, this explanation requires us to treat P3 as involving an exaggeration.

A third way to try to do justice to P3 takes seriously the idea that the phenomenon concerns the explicit *claims* we make rather than concerning (potentially unexpressed) judgements. As Kant says, "when [someone] puts a thing on a pedestal and *calls* it beautiful, he demands the same delight from others. […] he *demands* this agreement of them." When we assert a proposition, we are proposing to our conversational partners to accept this proposition for the purposes of the conversation. They can veto the proposal, by denying what we said, but if they let it pass, then the proposal becomes accepted, the proposition accepted for the purposes of the conversation. Many models of conversation acknowledge this basic picture (Stalnaker 1978, 2002; Lewis 1979b; Brandom 1983) The picture is controversial for perspectival or centered propositions (see e.g. Torre 2010; Ninan 2010a, 2010b; Stalnaker 2014; Kindermann 2018), but can be defended for that case (Kölbel 2013; Dinges 2017). If we accept the basic picture of the conversational effect of an assertion, we have the beginning of an explanation of phenomenon P3: we are making a demand on anyone who might be or become a party to this conversation that they accept, at least for the purposes of the conversation, what we have asserted. This is a demand they may not comply with, i.e. they may deny what we have asserted, thereby preventing the asserted proposition from becoming mutually accepted.

Now, it is important here to contrast aesthetic predicates, such as "beautiful" with predicates that wear their judge-dependence on their sleeve, such as "is my favourite painting." When I call a painting "beautiful," I may expect and demand the agreement of others in the sense of expecting and demanding that they allow this to become accepted in our conversation. This will not be so when I call a painting "my favourite painting": even though the concept of one's favourite painting may well be judge-dependent too, this particular form of words would not bring about acceptance of the painting in question as one's favourite painting in the conversation, even if no-one objects. When judge-dependent concepts are expressed through indexical language, the conversational effect will be different: at best it will become accepted that the painting is the favourite painting of that participant who just called the painting her favourite painting.[13]

I conclude that while phenomenon P3 cannot perhaps be saved exactly as stated by Kant, we can at least "explain it away": the account predicts some phenomena that are similar to P3 and may well have been the basis for the attraction of Kant's universality claim. The defender of judge-dependence needs to deny that an aesthetic judge who is minimally informed about the occurrence of interpersonal differences in taste can reasonably expect and demand of everyone unrestrictedly that they agree (in this sense, P3 is denied). But they can accept that we expect and demand of anyone whose taste we are presuming to share, that they agree with our judgements as to which things are beautiful. They can also accept that everyone who asserts that something is beautiful ("calls it beautiful") thereby makes a proposal to accept for the purposes of the conversation, that that thing is beautiful, and thereby expect and demand that their conversational partners do accept this by not rejecting the assertion they have made.

5. A parallel case for the judge dependence
of moral judgement?

In this final section, I want to argue for a certain asymmetry between moral and aesthetic judgement. While some of the motivations for a judge-

---

[13]   See Kölbel (2013) for elaboration.

dependence account transfer straightforwardly from the aesthetic to the moral case, there is also evidence against the judge-dependence of moral concepts. In my view, this evidence forces us to draw one of two uncomfortable conclusions: either that moral concepts are incoherent, or that the users of moral concepts are persistently confused about these concepts.

I take it that aesthetic concepts (at least the ones that I have been discussing here—not all aesthetic concepts mentioned by Sibley in his famous 1959 article) are a species of evaluative concept. Another species is that of moral concepts. However, some evaluative concepts seem not to be specifically related to any narrowly *experiential* response. Moral or normative concepts may impose restrictions on the preference structure of judges without being associated with any specific sensory response. However, we already saw in the case of aesthetic concepts that it is possible to conceive of the response mentioned in instances of (JD) as experiential in a wide sense. Aesthetic appreciation may be a partly intellectual response. Similarly, the responses associated with moral concepts—if treated on the (JD) model— may include experiential responses more or less closely tied to perception or other basic responses. Perhaps they are quasi-sensory experiences produced by a quasi-sensory "moral sensibility." For simplicity, let us call the relevant responses in the moral case "moral approval" and "moral disapproval."

We might then propose various instances of (JD) as articulations of the norms governing various moral concepts:

(G)   For all subjects s and all objects o: it is correct for s to apply *good* to o iff s is disposed, under favourable conditions, to respond with moral approval to o.

(B)   For all subjects s and all objects o: it is correct for s to apply *bad* to o iff s is disposed, under favourable conditions, to respond with moral disapproval to o.

(O)   For all subjects s and all objects o: it is correct for s to apply *ought* to action type a iff s is disposed, under favourable conditions, to respond with moral disapproval to a failure to perform a.

One advantage of this type of approach is that it offers a clear explanation of the motivational nature of these evaluative concepts. Motivational internalism in metaethics has remained controversial, but there seems to be

at least a very convincing suggestion that competence with evaluative concepts requires having a corresponding motive when applying them. Those who claim not to have the motive while also claiming to be applying the concept (so the internalist claims) have not really understood the concept or are applying a related but different concept (i.e. an "inverted commas use"). A judge-dependence account has no difficulty explaining this whenever the response mentioned in the relevant instance of schema (JD) is an experience that is appropriately motivating. If pleasure is intrinsically desirable, and correct application of the concept of beauty requires of the judge that he or she be disposed to experience a certain kind of pleasure in response to that object, then judging an object to be beautiful will have immediate consequences for rational motivation. If such a judgement does not, ceteris paribus, motivate me to choose the object over others not judged to be beautiful, then something will have gone wrong: be it my competence with the concept of beauty or my processes of rational deliberation. The same goes for the proposed instances of (JD) for moral concepts: if moral approval or disapproval is a response associated with the application of moral concepts, then this explains the link between the genuine application of moral concepts and motivation. It is rational to prefer, *ceteris paribus*, good things to those that are not. Why?—because it is rational to prefer *ceteris paribus* what one approves of.

The second type of support for a judge-dependence account of aesthetic concepts came from the observation that we regard it as a requirement for competence that users of these concepts apply them in line with their dispositions to respond. In the moral case, analogous observations support an analogous conclusion: we tend to require that competent users of moral concepts apply these concepts in line with their own responses of moral approval or disapproval. It is hard to imagine a case where we classify a thinker as competent with moral concepts on the basis of their applying them to the right range of objects, but criticize them for morally approving or disapproving out of line with their moral judgements. Perhaps this observation is just as controversial as the purported observations supporting internalism. But let that pass. My aim here is to point out an asymmetry and difficulty that would arise for a judge-dependence account of moral concepts even if we accept this evidence.

The asymmetry shows up when we consider a piece of evidence that can be adduced to support the view that aesthetic concepts are judge dependent in addition to being response dependent, i.e. that the relevant response profiles can vary from judge to judge. This is supported by the fact that we have no difficulty in conceding that someone else is correctly applying the concept of *beauty*, say, to a given case, yet refuse to apply the concept ourselves, since it would not be correct for us. Thus, it seems coherent for me to concede that someone else is complying with all the conceptual norms when calling something beautiful, yet to deny that that thing is beautiful, since this is what the conceptual norms require of me.[14]

In the moral case, by contrast, there seems to be something very odd about conceding that someone else is applying a moral concept correctly, yet to refuse to apply it oneself. Thus, even if the instances of (JD) proposed above were correct, there seems to be an additional requirement, on pain of incompetence, that different thinkers morally approve and disapprove alike.

A similar asymmetry affects the reasons that can be adduced against an ideal judge account of aesthetic concepts on the model of (IJ) above in §2. If aesthetic concepts followed the (IJ) model, then persistent divergence from applying these concepts in line with an ideal judge would count as a manifestation of incompetence. However, as long as these uses of a concept are plausibly in line with the user's own aesthetic responses, we draw no such conclusion. Thus, the (IJ) model seems to be the wrong model in the case of aesthetic concepts. The parallel case, however, cannot be made for moral concepts: if a thinker applies moral concepts perfectly in line with their own sentiments of moral approval or disapproval, but they diverge sharply from the way we think an ideal moral judge would apply the concepts, then we might conclude that the thinker is not competent with the moral concepts—at least we are more likely to do so than in the aesthetic case. The conceptual requirements seem to include both that the thinker judge in line with their own responses of approval and disapproval, and also that they judge in line with the responses they would have if they were ideal judges.

---

[14]   NB: this feature is directly related to the restrictions the judge dependence account is subject to in connection with phenomenon P3 above.

584 Max Kölbel

Thus, in the moral case there seems to be more pressure than in the aesthetic case to say that correctly performing judges will judge alike. If we wanted to follow the approach taken in (G), (B) and (O), i.e. accept that these articulate correct conceptual norms for the concepts of *good*, *bad*, and *ought*, then I can see only two fairly undesirable options: either we class the extra requirement of agreement amongst correctly performing judges as conceptual, or we don't. On the first option, moral concepts are deficient, on the second option the concept users are.

On the first option, the above instances of (JD) are not the only norms governing moral concepts. In addition, there are requirements that resemble instances of (IJ): concept users are required to judge in line with the approval and disapproval patterns of an ideal judge. But these conceptual requirements impose inconsistent requirements to the extent to which concept users diverge in their approval responses from ideal judges. In an exceptional world, in which moral education is perfectly uniform as well as perfectly successful, the two norms may be consistent. But not in any other case.

The effect of incoherent conceptual requirements may not always be detrimental. Some argue that vague concepts are governed by an ultimately incoherent tolerance principle (a principle that is both analytic and false, argues Eklund 2002). Some argue that compliance with the equivalence schema is part of the conceptual requirements on the concept of truth, yet when we discover the liar paradox, we will refrain from accepting the requirement in certain cases (Horwich 1998). Nevertheless, they claim, we are using these concepts fruitfully by exercising caution with the problematic principles: we simply refuse to apply the principles in problematic cases. Perhaps in the moral case it is harder to brush the incoherence under the carpet: both (G) and an appropriate (IJ) requirement would seem to be quite central demands. Thus perhaps in this case the inconsistency in the conceptual rules can be seen as driving us towards reaching a situation where the rules do not conflict, i.e. a case where our approval patterns are those of an ideal judge.

On the second option, the extra requirement of agreement amongst correctly performing judges is not construed as conceptual, and (G), (B) and (O) are the sole conceptual norms governing these moral concepts. On this view, our reluctance to say "He/she is applying the concept correctly to this object, but it would not be correct for me to apply it." is a reflection not

just of conceptual norms, but also of certain non-conceptual requirements. Correctness in applying the concept encompasses two components, only one of which is conceptual. The conceptual part requires the thinker to judge in line with (G), (B) or (O). However, full correctness also requires them to have the approval patterns of an ideal judge. If this is not the case, the application is not fully correct.

The disadvantage of this option is that there seems to be no evidence in the behaviour of concept users that would support the view that some of the norms they bring to bear regularly, and which seem to be an integral part of educating others in the use of moral concepts, are not in fact conceptual.

I conclude that a judge dependence account of moral concepts, while sharing some of the advantages of a judge-dependence account of aesthetic concepts, also suffers from difficulties not faced by the aesthetic analogue. However, it is not clear that any better account is available on which neither our moral concepts nor our patterns of use are deficient.

References

Brandom, Robert. 1983. "Asserting." *Noûs* 17 (4): 637–50.
https://doi.org/10.2307/2215086
Budd, Malcolm. 2003. "The Acquaintance Principle." *British Journal of Aesthetics* 43 (4): 386–92. https://doi.org/10.1093/bjaesthetics/43.4.386

Dinges, Alexander. 2017. "Relativism and Assertion." *Australasian Journal of Philosophy* 95 (4): 730–40. https://doi.org/10.1080/00048402.2017.1284248

Dinges, Alexander and Julia Zakkou. 2020. "Taste, Traits and Tendencies." *Philosophical Studies*, online first. https://doi.org/10.1007/s11098-020-01470-7

Egan, Andy. 2007. "Epistemic Modals, Relativism and Assertion." *Philosophical Studies* 133 (1): 1–22. https://doi.org/10.1007/s11098-006-9003-x

Egan, Andy. 2012. "Relativist Dispositionalist Theories of Value." *Southern Journal of Philosophy* 50 (4): 557–82. https://doi.org/10.1111/j.2041-6962.2012.00136.x

Eklund, Matti. 2002. "Inconsistent Languages." *Philosophy and Phenomenological Research* 64 (2): 251–75. https://doi.org/10.1111/j.1933-1592.2002.tb00001.x

Franzén, Nils. 2018. "Aesthetic Evaluation and First-Hand Experience." *Australasian Journal of Philosophy* 96 (4): 669–82. https://doi.org/10.1080/00048402.2018.1425729

Hopkins, Robert. 2001. "Kant, Quasi-realism, and the Autonomy of Aesthetic Judgement." *European Journal of Philosophy* 9 (2): 166–89. https://doi.org/10.1111/1468-0378.00134

Horwich, Paul. 1998. *Truth.* Oxford: Oxford University Press. https://doi.org/10.1093/0198752237.001.0001

Hume, David. 1757. "Of the Standard of Taste." In *Four Dissertations.* London: A. Millar.

Kant, Immanuel. 1790. *Critique of Judgment.* Page reference to trans. Meredith. Oxford: Oxford University Press 1928.

Kindermann, Dirk. 2018. "Varieties of Centering and *De Se* Communication." In *De Se Thought and Communication*, edited by Manuel García-Carpintero and Stephan Torre, 307–39. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198713265.003.0012

Kölbel, Max. 2013. "The Conversational Role of Centered Contents." *Inquiry* 56 (2/3): 97–121. https://doi.org/10.1080/0020174X.2013.784456

Kölbel, Max. 2014. "Agreement and Communication." *Erkenntnis* 79 (1): 101–20. https://doi.org/10.1007/s10670-013-9447-2

Kölbel, Max. 2015. "Relativism 1: Representational Content." *Philosophy Compass* 10 (1): 38–51. https://doi.org/10.1111/phc3.12190

Lewis, David. 1979a. "Attitudes De Dicto and De Se." *Philosophical Review* 88 (4): 513–43. https://doi.org/10.2307/2184843

Lewis, David. 1979b. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8 (1): 339–59, reprinted in his *Philosophical Papers*, vol. 1, Oxford: Oxford University Press, 1983. https://www.jstor.org/stable/30227173

Lewis, David. 1989. "Dispositional Theories of Value." *Proceedings of the Aristotelian Society Supplementary Volume* 63 (1): 113–37. https://doi.org/10.1093/aristoteliansupp/63.1.89

Ninan, Dilip. 2010a. "Semantics and the Objects of Assertion." *Linguistics and Philosophy* 33 (5): 355–80. https://doi.org/10.1007/s10988-011-9084-7

Ninan, Dilip. 2010b. "De Se Attitudes: Ascription and Communication." *Philosophy Compass* 5 (7): 551–67. https://doi.org/10.1111/j.1747-9991.2010.00290.x

Sibley, Frank. 1959. "Aesthetic Concepts." *Philosophical Review* 68 (4): 421–50. https://doi.org/10.2307/2182490

Stalnaker, Robert. 1978. "Assertion." In *Syntax and Semantics* 9: *Pragmatics*, edited by P. Cole, 315–22.

Stalnaker, Robert. 2002. "Common Ground." *Linguistics and Philosophy* 25 (5/6): 701–21. https://doi.org/10.1023/A:1020867916902

Stalnaker, Robert. 2014. *Context*. Oxford: Oxford University Press.

Torre, Stephan. 2010. "Centered Assertion." *Philosophical Studies* 150 (1): 97–114. https://doi.org/10.1007/s11098-009-9399-1

Wollheim, Richard. 1980. "Art and Evaluation." In Richard Wollheim, *Art and Its Objects*: *Second Edition with six supplementary essays*, 152–61. Cambridge: Cambridge University Press.

Wright, Crispin. 1992. *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press.

BOOK REVIEW

Galen Strawson: *Things That Bother Me. Death, Freedom, the Self, Etc.* New York: New York Review Books, 2018, 240 pages

Jacek Jarocki*

Galen Strawson is certainly one of the most original contemporary philosophers. And he has—perhaps as every original philosopher has—given rise to many controversies. The theories he defends, such as panpsychism or the conception of sesmets—extremely short-lived selves—may indeed prompt an 'incredulous stare' (as Peter van Inwagen once observed) and are usually rejected by philosophical orthodoxy. But this, often too hasty, criticism causes Strawson to express his views even more boldly; it is not surprising, then, that his thinking might be seen as lifelong training in philosophical rebellion. It is for that very reason that *Things That Bother Me* is so important. It enables one to understand that his motives are in fact quite the opposite: Strawson turns out to be a humanist, concerned with the sempiternal questions that strike every clever man.

The book consists of nine popular papers, i.e. meant for non-philosophers. The first, 'The Sense of the Self,' is the oldest one—it was written in 1996. Its main subject, clearly stated in the title, is the phenomenology of the self: the way we feel our inner I. Strawson mentions seven features that 'capture the core of the ordinary human sense of the self' (p. 30). However, not all of them are parts of a genuine experience of ourselves; and in some cases, features such as personality, activity and long-term continuity are absent. This 'thinned' sense of the self involves being a thing, being something mental, being subject of experience, singleness and distinctiveness. The vast part of this chapter is devoted to familiarising readers with the idea that longevity of the self—a feature that seems essential to many of us—might not be (and, indeed, is not to some)

*   John Paul II Catholic University of Lublin

     🖉   Institute of Philosophy, Faculty of Philosophy, John Paul II Catholic University of Lublin, Al. Racławickie 14, PL 20-950 Lublin, Poland

     ✉   jacekjarocki@kul.pl

a part of the sense of the self. Although it is difficult to say whether Strawson succeeds, he at least gives some notable examples to suggest that we should take his idea seriously.

The second chapter—'A Fallacy of our Age'—is one of many of Strawson's papers criticising narrativism, a conception that currently enjoys great popularity in almost every area of the humanities. Strawson distinguishes two versions of the view: according to the Psychological Narrativity Thesis, 'human beings typically experience their lives as a narrative or a story of some kind' (p. 45); by contrast, the Ethical Narrativity Thesis holds that a 'richly narrative outlook on one's life is a good thing, essential to living well' (p. 46). Strawson is one of a few philosophers to reject both theses. The reason for this follows on from the previous paper: some people experience their selves as short-lived (transient) and 'have no particular tendency to see their life in narrative form' (p. 48). As before, Strawson does not propose any formal argument but simply stresses the diversity of human phenomenology.

The third paper, 'I Have No Future,' addresses the problem of death expressed in a question: Is my death bad for me? Strawson defends a non-deprivation view, according to which death does not deprive the one who died of any good, and therefore it cannot be bad. He calls his view 'No Loss (of the Future)' (p. 72), although—under different names—it has been widely accepted throughout the ages, most notably by Epicurus. What is truly original about the version defended by Strawson is the justification: he argues that one loses nothing when one's death is—as it is often said—untimely. There is no such thing as a lost future of someone who is already dead. This point seems to be clear, despite its counter-intuitive consequences discussed by Strawson (at pp. 84–86). Still, it does not make death any less frightening: even if it does not deprive us anything, the thought of eternal non-existence is—at least to some—emotionally unbearable.

The next two papers are devoted to the problem of free will. Throughout his career, Strawson has argued against the possibility of freedom and (ultimate) moral responsibility. His argument rests on the assumption that being free and morally responsible require one to be *causa sui* (a cause of oneself). But being *causa sui* is incoherent, so freedom and moral responsibility are logically impossible. This is exactly the core argument that can be found in the paper 'Luck Swallows Everything.' However, despite its clarity, it is impossible for the reasoning to convince us that we are not free. We cannot help but believe that we are responsible agents, for we experience freedom in almost every moment of

our life. In effect, free will turns out to be both (metaphysically) impossible and (phenomenologically) necessary. One of the questions Strawson does not answer is why we are doomed to the experience of freedom. In his *Slaughterhouse-Five*, Kurt Vonnegut describes an alien race called Tralfamadorians that abducts humans from time to time. 'I've visited thirty-one inhabited planets in the universe, and I have studied reports on one hundred more. Only on Earth is there any talk of free will' explains one of the aliens. The question is: why is this so? Is there something special about our evolution? If yes, what role does this deceptive experience play? It is a shame that Strawson does not even try to tackle these questions at length.

The next chapter, titled 'You Cannot Make Yourself the Way You Are,' is a conversation between Strawson and Tamler Sommers. Here the former repeats—but in an even less formal manner—his pivotal views on free will. However, Sommers asks him a couple of troubling questions that lead Strawson to draw some inconvenient consequences. He is forced to admit, for example, that neither Adolf Hitler nor Joseph Stalin can be held responsible for their deeds. Even if we are able to accept this fact cognitively, for we have *a priori* proof, we still feel emotional resistance to treating genocide in the same way as we treat an earthquake: as something that just happened through nobody's fault. Strawson may be right, then, that 'the impossibility of radical free will, ultimate moral responsibility, can be proved with complete certainty' (p. 111); however, the consequences of adopting this view could be catastrophic for our everyday lives.

The next two papers consider the problem of consciousness. The first, titled—quite significantly—'The Silliest Claim,' discusses the views of the Deniers: the philosophers who deny the existence of conscious experience, also known as 'what's-it-like' (or qualia, although Strawson finds the last term 'uncomfortable' [p. 131]). Few of them do so explicitly. Their more common strategy is to redefine 'consciousness' in terms of behaviour, function or brain process. Strawson, who firmly believes that this view is mad, tries to find the roots of this claim. He believes that the first of them is metaphysical behaviourism—a mutation of a much weaker, methodological thesis about the object of psychology. The second, however, is far more important: it is the belief that we know exactly what the fundamental nature of the physical world is. This is—Strawson concedes—plainly false. But contemporary philosophers have made this fatal mistake even worse: they have assumed that this fundamental nature is also utterly non-experiential. Contrary to common opinion, nor is physics is much of a help,

for it is just a bunch of equations and pointers readings and so remains silent on the nature of the physical. Why, then, do some people endorse the Denial—'the silliest view ever held in the history of human thought' (p. 151)? Strawson offers a psychological explanation: we are tempted to believe in literally anything that is universally acclaimed.

So, what might be the fundamental nature of the world? Strawson tries to answer the question in the next paper, 'Real Naturalism,' in which he argues for panpsychism. Most philosophers assume that the stuff the physical world is made of is utterly different from our experience. This assumption leads straightforwardly to so called 'mind–body problem,' i.e. the question of how to combine experiential occurrences with the non-experiential nature of the physical world. Strawson reverses the initial assumption. We know, he argues, that experience exists and what it is, but we know nothing about the physical. What follows is, first, that the mind–body problem is a pseudo-problem, for its main question is similar to questions such as 'Have you already stopped bullying your wife?'; and second, that it is at least plausible that what we call the physical is indeed experiential. For, Strawson asks, what non-experiential nature could the physical have? This is a tricky question, because by definition we know nothing about anything non-experiential. In that light, panpsychism—physicalist panpsychism—is indeed the best explanation.

The next paper, 'The Unstoried Life,' goes back to the critical examination of narrativism. Once again, Strawson expresses his conviction that 'some of us are naturally—deeply, positively—nonnarrative' (p. 179). He scrutinizes the view of Marya Schechtman, who holds that we constitute our selves by weaving stories, that our life is indeed a process of life-writing. In contrast, Strawson cites those whose experience is totally different. He puts special emphasis on the fact that 'there are deep human differences' (p. 195) that we cannot ignore.

In light of the previous chapter, it may be a bit surprising that the last paper, titled 'Two Years' Time,' is autobiographical. However, Strawson clarifies, there is no contradiction between being a non-narrativist and writing one's own memoire. (Probably the more serious obstacle to Strawson writing a full biography would be the fact that—as he concedes at p. 34—he has 'a little interest in [his] own past'.) In this short paper, Strawson recalls the late 1960s—the final years of his time at Winchester College, a hitchhiking trip across the Middle East and his early years at Cambridge University. It is a very personal, sometimes painfully sincere yet beautiful story of growing up during the short

but revolutionary era of 'flower power,' full of rebellion, drugs, music and love. All in all, it is a story of—or a manifesto for—the whole generation.

As one can easily see, what binds these papers together is certainly not their consistency. Why, then, has Strawson decided to publish them in one volume? The answer appears obvious: he represents a small minority of living philosophers whose aim is to construct a sort of wider metaphysical frame: a kind of metaphysical system. And within that system there is a place for all the problems he tackles in the book. However, what is particularly remarkable about the project is its mainly aporetic character. Of course, Strawson's positive ideas, such as panpsychism, are certainly worth our attention; but the arguments he proposes against universally held philosophical views such as libertarianism (and compatibilism), narrativism or (non-experiential) naturalism are in many cases lethal. This becomes even more obvious in these popular papers, for they get things as straight as possible and do not split hairs.

But this attempt to explain why Strawson is so vividly interested in all these topics is not enough. The final answer to that question is hidden in the title of the book: all the problems Strawson takes on have bewildered him one way or another. In this way he reminds us of an old yet rarely expressed truth: that the basic stimulus to do philosophy is astonishment. Strawson's astonishment stems from two sources. The first are certain philosophical views he finds hard or even impossible to believe. In one of his papers, 'The Depth(s) of the Twentieth Century'—not included in this collection—he enumerates nine philosophical views that were commonly accepted when he was an undergraduate (e.g. on meaning and understanding, feeling and emotion or consciousness) and confesses that they 'seemed to me […] the wrong way round.' Then he adds: 'This led me to feel very insecure about my ability to do philosophy.' It is simply inexplicable—Strawson says—that some people believe that 'the self is an illusion generated by an improper use of language' (p. 23) or that conscious experience does not exist. So, one of the basic sources of Strawson's puzzlement are philosophical views.

Yet the second—and much more important—cause of Strawson's astonishment are not the answers to certain central philosophical problems but the problems themselves. It is obvious that many of them engage Strawson emotionally, which is sufficient to explain the title of the book. One might ask whether detached reflection—a view from nowhere—would not have been better than personal involvement? Strawson, though, proves this question to be ill-posed: his personal preferences give him at most the initial impulse to take on

certain issues, but justification of his views is independent of emotional commitments.

Still, the fact that one has to defend against the charge of personal involvement says a lot about the intellectual climate today and—at the same time—reveals the truth: that Strawson does not belong to the party of fully-fledged analytic philosophers. The latter often ignore the fact that, historically speaking, philosophy is mostly a discipline that does involve personal commitment. Pierre Hadot in his classical book *La Philosophie comme manière de vivre* [*Philosophy as a Way of Life*] shows that philosophical reflection has been a kind of a spiritual exercise, usually complementary to—but sometimes also substitutive for—religion. One of the hallmarks of the analytic tradition is that it has abandoned that meaning, as Thomas Nagel convincingly argues in his essay 'Secular Philosophy and Religious Temperament.' By contrast, Strawson tries to reintroduces philosophy understood as an outlook one is personally committed to; *Weltanschauung* as German idealists called a world view, a notion wider than philosophy yet still respecting the rules of reason. As a person who 'had a non-religious upbringing' (p. 13), Strawson seeks an alternative way of coping with deeply bothering '"cosmic" things' (p. 15) without reference to a transcendental being. Perhaps that is why he seldom introduces definition and argues for his views—rather, he 'want[s] to record […] reflections' (p. 73)—and even more rarely tries to convince his opponents. This might initially disappoint or even upset some analytically oriented readers. There are also *prima facie* contradictions, e.g. between Strawson's 'episodism'—his experience of being a short-living self—and besetting him from early childhood, his fear of death—'the first of the things that deeply bothered' him (p. 15). It is worth noting, however, that in a previous version of the essay 'I Have No Future' (published in a volume *The Subject of Experience*), Strawson replies to that charge: 'All that I can say is that this is a truthful report of how I feel—even if it involves a sort of inconsistency.' In this way Strawson gives expression to a general notion that people may be internally incoherent, especially when it comes to what they know and what they feel. Sometimes philosophy can help us to sort things out, but in some cases—as in the case of free will or fear of death—any systematic reflection is helpless. Consistency at all costs, although often very desirable, may move us away from the truth.

We must bear in mind, then, that *Things That Bother Me* is not in any case a philosophical treatise. On the other hand, it is something more than just a popular book for non-philosophers (although casual readers may benefit

considerably from it). It is, rather, the testimony of a humanist: someone inter-
ested in human nature. And the first object of inquiry for Strawson is Strawson
himself. For that reason, the book tells a very personal story. Strawson writes
about his bewilderments, about his severe depression and difficult youth but
also about his favourite reads and bands. The book reminds us that a true
philosopher has never been—and should never be—an office worker who leaves
his desk at 4.30pm. Rather, it is someone who, constantly bothered by certain
problems, tries to answer them or at least pose them clearly in a rational man-
ner.

It is no accident that *Things That Bother Me* was published by New York
Review Books. It is a sort of intellectual autobiography, full of confessions and
memories; quite similar to Michel de Montaigne's *Essais*, so frequently quoted
by Strawson. Besides, it is an excellently written book; witty, engaging and
sometimes provoking. In short, it is a brilliant piece of literature that easily lives
up to any expectations one might have upon seeing Ian McEwan's words printed
on the back cover: 'Galen Strawson is one of the cleverest men alive.'

Contents

RESEARCH ARTICLES

DISCUSSION NOTES

BOOK REVIEWS