6 The New Evil Demon and the Devil in the Details

Mikkel Gerken

6.1 Introduction and Outline

I will argue that cases of massive deception, such as new evil demon cases, as well as one-off cases of local deception, present challenges to views according to which epistemic reasons, epistemic warrant, epistemic rationality, or epistemic norms are factive. In doing so, I will argue that proponents of a factive turn should observe important distinctions between what are often simply referred to as ‘bad cases’. Recognising epistemologically significant differences between deception cases raises serious challenges for proponents of a factive turn. Here is how I will proceed.

In Section 6.2, I consider various cases of warranted false beliefs. These include probabilistic cases and cases of one-off perceptual illusions. I note that resisting a factive turn in epistemology does not involve any commitment to epistemic internalism. On the contrary, a combination of moderate externalism in epistemology and philosophy of mind provides a unified externalist motivation for assuming that agents who are prone to one-off deceptions have a warranted false belief. I also critically discuss the frequently invoked terminology of ‘good cases’ and ‘bad cases’.

In Section 6.3, I argue that the appropriate epistemic diagnosis of massive deception cases depends on the specific aspects of them. Specifically, I argue that massive deception cases may be specified such that the agent is unwarranted according to the unified mind/epistemology externalism. Thus, I argue that it is consistent with a unified externalist methodology to accept that massively deceived subjects are unwarranted without adopting the more radical view that warrant is factive.

In Section 6.4, I note some general ramifications and raise some challenges for proponents of a factive turn. I conclude that insofar as

---

1 I will use the term ‘warrant’ as a broad label for epistemic rationality. Since I am a pluralist about epistemic warrant, I regard it as a genus that harbours internalist species – labelled ‘justification’ – as well as externalist ones – labelled ‘entitlement’ (Burge 2003; Gerken 2013, forthcoming a).
epistemology has turned to factivity, we need a U-turn back to an episteme-
ology in which the non-factive aspects of cognition figure centrally.

6.2 Some Basic Cases

New Evil Demon (henceforth ‘NED’) cases and other cases of massive
deception, such as brain-in-a-vat (henceforth ‘BIV’) cases, are innova-
tions by philosophers. They are highly abstract thought experiments that
can be useful in investigating the limits of our epistemic concepts or fringe
areas of epistemic reality. However, assessing such thought experiments is
a complex affair that requires both theoretical background and adequate
specification.

Since cases of massive deception are so hard, it is methodically sound to
begin with simpler and better empirically understood cases of local decep-
tion. Here I agree with Kripke’s methodological dictum ‘Hard cases make
bad law’ (Kripke 2011: 160). One advantage of beginning with cases of
one-off visual illusions is that we may draw on both vision science and on
philosophy of mind. Another reason for going from the comparatively
simple cases to the harder cases is that cases of massive deception are
philosophers’ attempt to generalise the familiar one-off illusion cases. If
the generalisation turns out to be problematic, that may reveal a lot of
interesting things. But it does not automatically reveal that the diagnosis
of the original one-off illusion case is misguided.

While these methodological remarks may seem plain, they lead to
serious challenges for proponents of a factive turn who argue from cases
of massive deception. In particular, I will argue that such theorists cannot
rest content with considering inadequately specified NED and BIV cases
but that they must also confront the simpler cases of one-off illusions. To
make good on this suggestion, let us consider some basic cases.

6.2.1 Probabilistic Cases

The relationship between probabilistic rationality and epistemic ration-
ality is a controversial affair. Nevertheless, a probabilistic case may illus-
strate an important distinction between normative judgments that focus
solely on the outcome of a decision or belief and normative judgments that
focus on rationality. So, I will start with a probabilistic case:

TICKET

Mr. Bigshot is an overconfident blowhard visiting Las Vegas. After drinking all the
bourbon on the Strip, he offers you a bet of $100 that a ticket, \( t \), will win a fair
lottery with a billion tickets. You like those odds, and you have $100 to spare. So,
you take the bet. Alas, ticket $t$ wins, and Mr. Bigshot boasts, ‘See, you shouldn’t have taken the bet.’

In one sense of the vernacular ‘should’ (or the more theoretical ‘ought’) – namely, the outcome sense – Mr. Bigshot is right. Given the outcome, things are now financially worse for you. But it is also clear that in the senses of ‘should’ and ‘ought’ that concern rationality, Mr. Bigshot is wrong. The odds were $1:1,000,000,000$, and *ceteris paribus* those odds made it rational to take the bet.\(^2\) Insofar as an outright belief may be epistemically warranted by probabilistic considerations, your false belief that ticket $t$ will not win was epistemically warranted. We do not need to require that probabilities translate straightforwardly into epistemic warrant to assume that it is epistemically rational to believe that $t$ does not win given the odds. While we should be cautious about reducing epistemic warrant to probability, we should also be cautious about divorcing the two entirely. So, although there is a sense – the outcome sense – according to which you should not have believed that $t$ loses, you should from an epistemological point of view believe that $t$ loses.

The epistemic rationality sense of ‘should’/’ought’ is reasonably well established in common sense and there is terminology reflecting it where it matters. For example, poker players have a notion of a *bad beat*. Roughly, a bad beat occurs when a play had the wrong outcome (i.e., the player was beaten) even though the player ought to have made it in order to maximise her chances of winning. It would be uncharitable to the poker players to think what is meant is merely that the player who suffers a bad beat is blameless in the outcome. Rather, the point is that she lost despite playing as she should given the aim of winning. Such quasi-technical terminology is also found elsewhere and indicates a common-sense appreciation of the distinction between the outcome ‘ought’ and the epistemic ‘ought’.

Although, one can recognise the distinction between the outcome ‘ought’ and the epistemic ‘ought’ by common-sense reflection, we do not always adequately adhere to the distinction. For example, the phenomenon of *outcome bias* is pervasive in human judgment. Roughly, this is the tendency to assess a decision or performance on the basis of its outcome rather than on the basis on which it was made (Baron and Hershey 1988). Empirical research has found outcome bias in many domains (Mazzocco et al. 2004).\(^3\) The fact that psychologists standardly regard this phenomenon as a *bias*

\(^2\) The *ceteris paribus* clause is due to the assumptions that we abstract away from, such as the assumption that you have $100 to spare.

\(^3\) Results that I think are reasonably interpreted as epistemic outcome bias have been unearthed by (Turri 2015b, 2016e). However, Turri himself takes these findings as evidence for factive epistemic norms. Since settling this issue depends on substantive methodological assumptions about the relationship between folk epistemology and epistemology, I will not attempt to do so here, but see Gerken (2017). Here I remain content with calling attention to outcome...
suggests that the assessment of the outcome of a fallible procedure should not be determined by its outcome on the particular occasion. Of course, one may argue that the presuppositions of science are misguided, and some factivist epistemologists appear to be committed to doing so (for an illustrative dispute, see Burge 2005, 2011; McDowell 2010, 2013b). But dismissing the scientific presuppositions on philosophical grounds requires substantive argument and engagement with the relevant sciences (Gerken 2017). Moreover, the point is familiar from externalist epistemology. According to epistemic externalists, a belief should be epistemically assessed, in part, in terms of whether it is a truth-conducive *type* rather than in terms of whether the particular *token* is true or false (Goldman 1979, 1986; Burge 2003).

I will not here argue against anyone who denies that you meet the epistemic constraints on taking the bet in the case *TICKET*. Some theorists think that no amount of probabilistic support less than 1 may provide epistemic warrant (Smith 2016). Theorists committed to the conjunction of the view that we cannot know lottery propositions and the view that a belief is warranted only if it is known have a similar commitment. Arguing against proponents of the latter view is beyond the scope of the present chapter. (But I would love to gamble with them all night long.) The probabilistic case makes vivid the distinction between the one-off outcome belief and the type of belief or its basis in abstraction from its outcome. It is epistemically rational to hold a belief of the type that has a $1:1,000,000,000$ chance of being true insofar as one upholds the belief in virtue of manifesting a cognitive competence — i.e., one that is properly sensitive to the probabilities. With the distinction between outcome and epistemic rationality in mind, let us move to a perceptual case.

### 6.2.2 Perceptual Cases

Let me try my hand at a case of a perceptual illusion in which the subject has no background indication that deception is going on:

**COIN ILLUSION**

In the not so distant future, scientists have developed the perfect hologram machine that generates holograms that cannot be visually discriminated from the entities they replicate. To prevent theft of a rare Persian gold coin, Penny acquires a hologram machine disguised as a glass display and programs it to make a hologram of her coin. When Penny is visited by her nephew, Silas, she does not tell him that the coin has been replaced with a hologram. The light emerging from bias and making the obvious suggestion that it is just that — a bias of our folk epistemology that should not be incorporated straightforwardly in our epistemological theorising.

---

4 For a conceptual framework for discussing these issues, see Gerken (2017: ch. 6).
the hologram hits Silas’s retina in a way that is photon-to-photon identical to the light that would have hit Silas’s retina had he seen the real coin. Thus, Silas forms the visual belief there is a coin in the display.  

**COIN ILLUSION** is a non-probabilistic case in which the belief is the product of the exercise of a truth-conducive visual cognitive competence. Again, there is a sense – the outcome sense – in which Silas ought not believe that there is a coin. After all, the belief is false. While that is an epistemically bad thing for Silas, there is an equally clear and genuinely epistemic sense in which Silas ought to hold the belief. Again, epistemic externalists have it that the sense of epistemic rationality is associated with the type of belief or type of basis for the belief rather than with the particular token (Goldman 1979; Burge 2003). Silas’s belief is of a truth-conducive type even though this token belief is false. Normally, a thing that looks precisely like a coin is a coin. Normally, Penny tells him the truth etc. So, normally, beliefs of the relevant type are true insofar as a truth-conducive cognitive competence is responsible for them. As above, it is beyond the scope of the present chapter to argue in a non-question-begging way that such one-off false beliefs are warranted. Rather, I seek to make two more modest points.

The first point is this: An argument that a subject is unwarranted in massive deception cases does not motivate the claim that the subject is unwarranted in one-off or local deception cases. This is important because it is sometimes suggested that massive deception cases are the central cases for a non-factivist epistemology. Thus Turri, for example, claims that the non-factivist argument is ‘epitomized’ by ‘thought experiments about radically deceived ‘brains in vats’” (Turri 2016e: 348). Other proponents of a factive turn simply argue against NED and BIV cases without addressing one-off illusion cases.

The second point is this: The assumption that the subject is warranted in the one-off illusion case in no way hinges on epistemically internalist assumptions. Rather, there is a basic epistemically externalist rationale for assuming that such a subject is warranted. Let us consider this point.

### 6.2.3 Unified Externalist Explanations

To illustrate the compatibility of warranted false belief and moderate (non-factive) epistemic externalism, I will sketch the contours of such a positive account. Although I cannot defend or develop it here, even this

---

5 I use underlining to mention concepts and thoughts.

6 If you harbour doubts about this case, substitute it with one in which you accept that things are typically what they seem.
rough sketch will put us in a better position for assessing the hard cases of massive deception, such as NED and BIV cases.

Epistemic externalists, such as process reliabilists, may point to the fact that Silas’s visual faculty is generally highly reliable although it is not infallible. The fallibility is manifest in *COIN ILLUSION* (Goldman 1979; Burge 2003; Graham 2012; Lyons 2013). A starting point for such an epistemically externalist account is the assumption that the subject is warranted in virtue of exercising a cognitive competence that is objectively truth-conducive in the subject’s normal environment (Burge 2003; Graham 2012; Gerken 2013).

This approach is epistemically externalist in a negative and in a positive sense. The negative sense is that the subject does *not* need to be aware of, or have any sort of epistemic access to, the fact that the competence is truth-conducive. Epistemic externalists standardly distinguish between the property of being warranted (e.g., as having an externalist entitlement) and the property of being able to justify a belief (Goldman 1979; for my take see Gerken forthcoming a).

The positive externalist assumption is this: The objective truth-conduciveness is explained in terms of the likelihood of true belief relative to the subject’s normal environment. In the present case, we may stipulate that as a matter of external fact, the subject’s normal environment is such that things that appear like coins typically are coins. This assumption about the role of the general features of the normal external environment is central to the epistemically externalist account. Had the subject’s environment been generally deceptive (e.g., counterfeit coins or holograms were very prevalent), he would not have been warranted on the basis of vision alone (Gerken 2009).

In sum, the account of why S is warranted in *COIN ILLUSION* is epistemically externalist by combining the two noted points: S is warranted partly in virtue of facts about general features of S’s external environment that S does not need to have any access to. For example, the environmental fact that coin-looking things normally are coins underwrites Silas’s warrant in his coin-belief even though he lacks any reflective access to this fact.

The epistemically externalist account may be integrated with externalism about mental state individuation. (I sometimes use ‘anti-individualism’ following Burge 1979.) Roughly, this is the view that an agent’s mental state types are partly individuated by patterns of relations holding between the agent and her external environment (see Burge 2010: 61 for a more precise formulation). Integrating epistemic externalism with externalism about mental states provides a way to constrain the notion of the normal environment. Roughly, the idea is that the epistemically normal environment
is constrained by the environment that plays the state-individuating role (Burge 2003; Sawyer and Majors 2005; Gerken 2013). Given that this approach unifies externalism in epistemology and mind, let us call it ‘unified externalism’.

According to unified externalism, the same environmental facts that help explain why Silas has a false coin-belief rather than a true hologram coin-belief also help explain why Silas’s belief is warranted. In each case, the explanation is, very roughly, that the coin-looking things that are prominent in Silas’s environment are normally coins (rather than, for example, holograms). This provides an externalist specification of the normal environment that differs from Goldman’s doxastic account of the normal environment as the one that is consistent with our general beliefs (Goldman 1986: 107). Where Goldman’s doxastic account of the normal environment is subjective, the unified externalist one is objective.

While I am sympathetic to this unified externalist account, I have argued elsewhere against an identification of the epistemically normal environment and the state-individuating environment (Gerken 2013). Although this criticism raises some difficult questions, it does not tell against the general strategy of integrating externalism about mental states (anti-individualism) and epistemic externalism. In particular, both anti-individualists and epistemic externalists emphasise how patterns of relations between the agent and general features of the environment bear on the type of belief and on its warrant.

Consequently, one-off illusions are regarded as exceptional cases that are explained by the fallibility of cognitive competences. As noted, this helps explain why the agent remains objectively warranted in a false belief. I highlight this unified externalism because proponents of the factive turn sometimes cite anti-individualists en passant as if they provided a congenial account. (E.g., Williamson (2000, forthcoming a) cites Burge 1979.) But for the reasons noted above, anti-individualism is no friend of the factive turn (see also Smith forthcoming; Gerken forthcoming b).7 Similarly, moderate epistemic externalism is directly at odds with factivist epistemologies. Moderate (non-factive) epistemic externalists regard it as

7 Of course, other more radical versions of externalism about mental states are more congenial to factivity in epistemology. For example, relational or disjunctivist versions that do not focus on the type of belief (McDowell 1994). There are plenty of philosophical and empirical problems for such views (Burge 2005). But given that an epistemic externalist account hinges on substantive assumptions in the philosophy of mind, proponents of a factive turn must clarify how their view relates to theories of attitude individuation pace, for example, Pritchard who seeks to remain neutral on these issues (Pritchard 2012). For an argument that the most influential factive epistemology – namely, knowledge-first epistemology – is committed to an extreme version of externalism about mental states, see Smith (forthcoming) and Gerken (forthcoming b).
an advantage of their view that it can account for warranted false belief in one-off cases such as COIN ILLUSION. A unified externalist account lends weight to this assumption.

6.2.4 The Good, the Bad, and the Technical Terminology

Perhaps it is worth making a terminological point. Often proponents of the factive turn deploy the terminology of ‘good’ and ‘bad’ cases (the trend goes back to Williamson 2000; see also Williamson forthcoming). This may be useful when context provides specification or when the terms work as anaphoric pronouns referring back to an antecedently specified case. But this terminology is far too coarse-grained to serve as technical terminology. To subsume radically different case-types in which the agent does not know under the heading ‘bad cases’ is to conflate things.

One important point is that conflating one-off illusion cases with cases of massive deception is illegitimate since epistemic externalists may grant that subjects are unwarranted in the latter cases by the same reasoning that motivates the view that the subject is warranted in the former cases. In each case, general features of the agent’s normal environment play a key role in the explanation. Furthermore, even within the category of massive deception scenarios, the substantive variation in details is so great that it is problematic to subsume them under the heading ‘bad cases’.

I will consider the variations of massive deception cases below. But it may be worthwhile to quickly indicate the problems with labelling the variety of one-off deception cases as ‘bad cases’. For example, it is crucial to distinguish between one-off illusion cases in which the external environment is indiscernibly deceptive and one-off hallucination cases in which the false belief arises from a malfunction of the agent’s cognitive competences. Many epistemic externalists require that the token belief under consideration be generated or sustained by a well-functioning cognitive competence (Graham 2012; Gerken 2013). Thus, epistemic externalists may argue that in one-off hallucination cases, the agent is, as opposed to one-off illusion cases, not warranted. Of course, the issue is complex due to important differences within hallucination cases. For example, we must distinguish between veridical and non-veridical versions. Moreover, we must distinguish between hallucinations owing to systematic features, and mere performance errors. Other cases are even harder to categorise. For example, Williamson imagines a device that has an occasional but unsystematic ‘scrambling effect’ on brains (Williamson forthcoming a). Likewise, cases may differ in whether they involve belief-dependent or belief-independent processes, and they may be cast in terms of evidence or cognitive dispositions (Lyons 2013).
My point here is not to settle substantive issues about any of these complex cases but to urge recognition that the mentioned complexities are epistemologically significant. It may be convenient to gloss such cases simply as ‘bad cases’. But doing so encourages substantive equivocations. Since no one would want to rest an argument for a factive turn on equivocations, the ‘good’ and ‘bad’ terminology should be replaced with either less course-grained technical terms or with specified cases.

The issue becomes even more complex once we turn from one-off cases of false belief to cases of massive deception. In particular, NED and BIV cases leave so much room for variation that further specification is paramount.

6.2.5 Preliminary Upshots

One upshot is that the verdict that S is warranted in one-off illusion cases does not hinge on any epistemically internalist assumptions. I have not sought to argue pre-emptively that there is no possible argument from one-off deception cases to a factive turn in epistemology. Rather, I have argued that such an argument from one specified case-type must be given. The differences between one-off, local, and massive deception cases and the differences within each category are epistemologically significant. We can sum up this basic point as follows:

Cases and Cases
An argument that a subject is not warranted in a massive deception case falls short of an argument that a subject is not warranted in a one-off deception case.

While Cases and Cases articulates a basic methodological point, it is not always adequately respected in practice. Likewise, unifying epistemic externalism with externalism about mental state individuation requires specification of the cases under investigation. Such a specification matters for the individuation of the subject’s mental state types which, in turn, matters for the epistemological account. To illustrate these points, I will consider a variety of cases of massive deception.

6.3 Massive Deception Cases

The NED scenario was originally set forth as an objection to reliabilism. More specifically, it was set forth as a counterexample to the thesis that a belief-generating process generates warranted beliefs only if it is de facto reliable (Cohen and Lehrer 1983; Cohen 1984). According to Cohen and Lehrer, the individual’s beliefs remain warranted in such a scenario, and
this speaks against the idea that de facto reliability of cognitive competences is necessary for epistemic rationality

6.3.1 Specification Required

To see how massive deception scenarios are underspecified in epistemically important ways, let us consider an original statement of the NED scenario by Lehrer and Cohen:

Imagine that unknown to us, our cognitive processes, those involved in perception, memory and inference, are rendered unreliable by the actions of a powerful demon or malevolent scientist. (Cohen and Lehrer 1983: 192)

A number of features of the NED scenario are unspecified. These include the duration of the deception but also the mode of deception. I will not attempt to resolve the challenge raised by the NED scenario or similar massive deception scenarios such as BIV cases. But I will consider how some specifications of massive deception scenarios may be addressed by extending the unified externalist normal circumstance approach that explains cases of one-off illusion (Burge 2003; Sawyer and Majors 2005).

The point about inadequate specification is not meant as thought experiment–bashing. We can learn from thought experiments even when they involve metaphysically impossible aspects. So, I do not criticise the extravagant or far-fetched aspect of massive deception scenarios. That said, far-fetched thought experiments raise multiple methodological issues. If, for example, we are dealing with a scenario, such as a demon scenario, that is arguably metaphysically impossible, this raises the bar in terms of how much specification we need. In an empirically recognisable case, background empirical knowledge may ‘fill in’ missing details. The ability to specify the details of the scenario to be reflected upon is a luxury of constructing a thought experiment, but it is also an obligation. When we reflect on a purely thought experimental scenario, that scenario is all that we have to reflect on. Perhaps, then, the following methodological guideline is a reasonable one:

---

8 For example, I think that Twin Earth thought experiments can serve their argumentative purpose even though they are impossible. The thought experiment has it that there is no water (H\textsubscript{2}O) on Twin Earth and also that Twin Oscar’s body is molecule-to-molecule identical to Oscar’s body (Putnam 1975). But Oscar’s body is partly constituted by water (H\textsubscript{2}O). This flaw of the thought experiment has rightly been regarded as irrelevant. Likewise, the impossibility of demons may be irrelevant to the force of NED considerations. But that is no excuse for failing to specify the scenario.
Thought Experiment Guideline

The less empirically well understood a thought experimental scenario is, the more incumbent it is upon the thought experimenter to include a specification of details of the scenario.

Unfortunately, this guideline is not always heeded when massive deceptions are cited as reasons for a factive turn in epistemology (Littlejohn 2009b; Logins 2014; Williamson forthcoming a). Proponents of a factive epistemology are certainly not alone in failing to adequately distinguish between cases. But this failure may encourage the misguided thought that a subject in a one-off deception case is unwarranted because a subject in a massive deception case is unwarranted. So, failing to distinguish between cases may encourage an erroneous motivation for a factive epistemology.

In contrast to NED cases, the hologram scenario, for example, is underwritten by what we know empirically about vision and the laws of optics. BIV scenarios tend to be situated between hologram cases and NED cases. They are not as empirically well understood as hologram cases, but they may improve on the demon scenarios by approaching nomological possibility and by some specification of the mode of deception.

However, even BIV scenarios require specification, for example of the time and duration of the envatment (Putnam 1981; Brueckner 2012). Consequently, I will stick with the modus operandi of beginning with less extreme deception scenarios and then consider how far such cases may be generalised before the subject should no longer be regarded as warranted. So before moving to the NED and BIV versions of massive deception cases, I will consider an attempted generalisation of the one-off case of perceptual illusion.

6.3.2 Massive but Non-Global Recent Deception

As mentioned, the NED scenario may be seen as an attempt to generalise the one-off illusion scenario. So, to provide a candidate case for a massive but non-global deception, I will set aside demons and supercomputers and stick to holograms:

MUSEUM OF ILLUSIONS

On his fifteenth birthday, Jorge is sentenced to a lifetime of service at a public museum. He is to guard the museum’s vast, ever-changing collections during its closed hours of 16:00–12:00. The remaining eight hours, he spends sleeping in a completely dark guard’s chamber at the museum. But, unbeknownst to Jorge, the collections have been sold and replaced with indiscernible holograms. So, for the rest of his life, Jorge wanders the halls in solitude under the impression that he is guarding ancient treasures in their glass displays. He takes pleasure in his job and forms many visual beliefs (that is a vase from the Qing dynasty; here is a statue that
is made of bronze; there is a double-edged sword). On Jorge’s seventy-fifth birthday, then, the majority of the visual beliefs that he has formed during his lifetime have been false.

What is the status of Jorge’s beliefs? In this case, the unified externalist account may regard Jorge as entitled (externalistically warranted) to his visual beliefs under discussion. Jorge is removed from his normal environment and placed in an indiscernible illusory one. But his visual competence remains truth-conducive relative to the environment in which he was originally embedded in a manner such that it played a part in type-individuating his mental states. Moreover, Jorge is prone to an illusion rather than a hallucination. His cognitive competences are not malfunctioning and had he been placed in a normal museum, his beliefs would have been true. So, the unified externalist account appears to be compatible with assuming that someone can be entitled to the majority of his beliefs.

However, from the perspective of a unified externalist, the verdict that Jorge remains entitled is already far more controversial than the verdict about the one-off visual illusion. It raises complex questions about the nature of a normal environment and the individuation of perceptual beliefs.

For example, some might regard the case as a ‘slow-switch’ case in which the museum environment becomes the environment that type-individuates Jorge’s mental states (Burge 1988; Boghossian 1992; Comesaña 2002; Littlejohn 2009a). I doubt that this is the right account. But the rationale for this verdict requires more philosophy of mind than I can include here (but see Gerken 2013: ch. IV for an extensive discussion of slow-switching). What I will note is that the features of Jorge’s new environment are artefacts made with the intention of deceiving the spectators. The holograms are look-alikes of the real things. So, given that we regard Jorge as continuously deceived, we should resist saying that he forms true beliefs with new contents (involving new concepts, such as $\text{vase}^H$ and $\text{sword}^H$). Rather, he continuously forms false vase and sword beliefs type-individuated by the normal (non-deceptive) environment. If so, we may describe his visual belief-generating processes as de facto unreliable but reliable relative to his normal environment. Thus, some unified externalist normal environment accounts regard MUSEUM OF ILLUSIONS as exemplifying a case of massively mistaken entitled belief.

---

9 Following the notation introduced in Gerken (2013), I add superscripts to indicate the concept type – in this case $^H$ for ‘Hologram’. The type individuation is often not transparent to the subject. So, the superscripts are added for the reader’s convenience.
These remarks do not amount to a full response to *MUSEUM OF ILLUSIONS* and much less to more dramatic cases such as NED or BIV cases. A full response requires a full account of the epistemically normal environment, of the mental state-individuating environment and of the relationship between them.\(^{10}\) Rather, the case is a candidate illustration of the idea that (visual) entitlement is compatible with massive, albeit not total, deception. However, it does not speak directly to even more radical NED scenarios.

Note that even if it is denied that Jorge remains warranted in his visual beliefs in the *MUSEUM OF ILLUSIONS* scenario, or upheld that he is (due to slow-switching) warranted in type-distinct beliefs, there is no slippery slope towards the view that Silas is not warranted in his belief in the one-off *COIN ILLUSION* scenario. Consider, for example, an externalist about both warrant and mental state individuation who thinks that the museum eventually constitutes a new epistemically normal – but still deceptive – environment relative to which Jorge’s cognitive competences are unreliable.\(^{11}\) Such an externalist will conclude that Jorge’s beliefs are unwarranted. But such a theorist must do so in terms of the same type of explanation that explained why Silas’s false belief in the one-off case is warranted. That is, the explanation must appeal to general features of the external environment rather than merely to the truth or falsity of the token beliefs.

How far may this sort of case generalise? I tentatively take *MUSEUM OF ILLUSIONS* to be in the range of cases of massive deception that exemplify warranted false beliefs. But as indicated, even this case raises hard questions about mental state individuation, competence individuation, and epistemically normal circumstances. As we generalise further towards cases in which the deceived subject is so deceived that she has never held a true belief, it becomes harder to uphold the view that the subject has as warranted beliefs as the rest of us (Brueckner 2012; Lyons 2013). The patterns of relations to the external world that partly constitute the subject’s perceptual entitlement are lacking (Gerken 2013). To see this, I will consider and compare some more radical enfratment cases.

---

\(^{10}\) The case illustrates why it is problematic to reduce epistemic normality to frequency. As noted, I have argued contra Burge, and Sawyer and Majors, that the epistemically normal environment is merely *constrained* by but not *identical to* the state-individuating environment (Gerken 2013). So, plenty of questions about epistemically normal environments remain open. However, just as a purely modal characterisation of warrant requires an account of modal nearness and a relevant alternatives theory requires an account of epistemic relevance, so talk of normal circumstance requires an account of epistemic normality.

\(^{11}\) I myself suspect that his diagnosis is problematic for the reason noted above. The assumption of deceptiveness indicates that the relevant environment is the one Jorge grew up in.
6.3.3 Attempted Generalisations – Warrant and Mental Representation

To further indicate how the specification of the massive deception case matters for the unified externalist account, let us compare two envatment cases – an original and a recent envatment case. The difference has been recognised since Putnam and is widely discussed in debates over scepticism (Putnam 1981; Brueckner 2012; Graham 2012):

ORIGINAL ENVATMENT

An odd quantum-event occurs, and the result is a configuration of a ‘brain’ – let’s call it ‘Brian’ – that is connected to a ‘super-computer’ which produces stimulations that yield a phenomenology in Brian that is indiscernible from the phenomenology of the perceptual representations of an adult human.

From the perspective of externalist philosophy of mind, it may be questioned whether Brian has any representational mental states. Indeed, it may be reasonably doubted whether Brian is a brain and whether the entity delivering the inputs is a computer. Thus, Graham about the entity in a similar case: ‘It’s a sophisticated lump of grey cells that bears no interesting constitutive relations to anything else’ (Graham 2012: 468).

Is the life of Brian one without mental representations? An affirmative answer may be motivated by focusing on function, history, causal connection, etc. But if Brian is not even capable of entertaining representational mental states, it is natural to also deny that he is warranted. Moreover, even if Brian does have beliefs, it is problematic to regard them as entitled insofar as the relevant belief-forming competence does not appear to be truth-conducive relative to any normal environment. But again the issue is complex since it might be argued that Brian has reliably true beliefs about the relevant stimuli (Chalmers 2005). If Brian is assumed to entertain beliefs, further complications arise for pluralists about epistemic warrant – i.e., theorists who think that there are both internalist and externalist types of epistemic warrant (Gerken forthcoming). Given that Brian is ascribed beliefs, it may be tempting for an epistemic pluralist to regard them as justified – i.e., as internally warranted (Miracchi forthcoming a). For example, Brian’s assumed beliefs may serve as reasons.

Again, my aim is not to resolve these complex issues but to indicate that their complexity is epistemologically significant and to note that the rationales for denying that Brian is warranted do not extend to the one-off illusion case. The epistemological significance of the issues in the

---

12 For examples, see Putnam (1981), Millikan (1996), Burge (2010), and Graham (2012).
13 Of course, some positions in philosophical space allow for such warrant. One might say that although Brian is propositionally or personally warranted. For the latter notion, see (Littlejohn 2009b).
philosophy of mind is most apparent by comparison, so let’s compare ORIGI
NAL ENVATMENT with a recent envatment case:

**RECENT ENVATMENT**
Ina is anaesthetised, kidnapped and connected to a super-computer that produces neural stimulations that yield a phenomenology in her brain that is indiscernible from the phenomenology of the perceptual representations that Ina used to enjoy.

I am inclined to regard this as a limiting case in the grey-zone between warrant and lack thereof. It is noteworthy, however, that the case for ascribing warrant to envatted-Ina does not derive primarily from the fact that the computer stimuli are indiscernible from worldly stimuli. Likewise, the case for ascribing warrant does not hinge on Ina being blameless or excused for having false beliefs. These are at most necessary but insufficient conditions on warrant. The core of the explanation has to do with the fact that Ina is in epistemically abnormal circumstances. In the normal environment, the environment that grounds the assumption that recently envatted-Ina has false beliefs, Ina would have had true beliefs. This is the assumption characteristic of the unified externalism. However, as above, **RECENT ENVATMENT** raises many questions – for example about attitude and competence individuation – that may lead a unified externalist to deny that Ina is warranted. Crucially, however, these questions derive from the generalisation and do not arise in the case of one-off illusions.

The relationship between the ascription of mental states and warrant is highly complex. Whereas it is problematic to ascribe warrant without the capacity for representational mental states, it may be reasonable to ascribe mental representation without the capacity for warrant to a subject. However, it is hard to come by a clear example of this phenomenon. But here is a stab at a candidate case:

**JELLYFISH**
Scientists discover a cognitively sophisticated species of jellyfish. The sensory organs of these jellyfish can discern nutritious items – zooplankton, eggs, larvae etc. – from sediment particles. However, the scientists find a way to make the sediment particles look like the nutritious items and feed the jellyfish by adding liquid nutrients to the water in the tank where they are kept. After a while – a long while – in this deceptive environment, the jellyfish evolve dramatically and become capable of basic perceptual belief. Although the jellyfish have undergone an extreme cognitive development, the discriminative abilities of their sensory system remain unchanged. In this case, a jellyfish, Mr. J., may form a belief that there is a cloud of eggs immediately below it when there is, in reality, only a cloud of fool’s egg sediment particles.
It is hard to accurately diagnose this case, and further specification is likely required. In particular, it is unclear whether it is a reasonable stipulation that Mr. J. really can be said to have a perceptual belief. However, even if this is granted, there are reasons to regard Mr. J. as unwarranted. After all, Mr. J. has never had any true beliefs about eggs. Indeed, his assumed capacity to form such beliefs was developed in a deceptive environment. So, arguably, the relevant cognitive competence is not truth-conducive in the environment that plays the state-individuating role.

The unified externalist approach does not merely pursue reasonable diagnoses about individual cases but also underlying principles. While this is a formidable task, it is illustrative to note a trend suggested by reflection on these case variations: The more detached a creature, in the broadest sense of the term, is from the external environment, the less plausible is it that the creature is warranted in beliefs about the external environment.

Brian (the brain-like creature) fails to have warrant for mental representations concerning the external world in virtue of being entirely detached from it. In turn, Mr. J. (the evolved jellyfish) may be sufficiently connected to the external environment to be ascribed representational mental states, but is not a good candidate for being ascribed warrant for them. Finally, Ina (the recently envatted person) is at least a candidate for someone who has false warranted beliefs.

I reemphasise that these matters are so complex that further deception cases must be considered in conjunction with what we know empirically about the phylogenesis of cognitive competences. But the cases highlight why specification is crucial.

Moreover, these case-variations indicate a contrast between a factive turn and the unified moderate externalist approach. As I have sketchily indicated, proponents of the latter may pursue a principled fine-grained diagnosis of massive deception cases by reflection on the relationship between the nature of mental representation, cognitive competences, empirical background conditions and epistemic warrant. In contrast, it is unclear how proponents of a factive turn may pursue such fine-grained explanations. Elsewhere I have criticised the most prominent brand of factivist epistemology – namely, knowledge-first epistemology – on similar grounds (Gerken 2017, forthcoming b). The principled fine-grained diagnoses sketched above crucially appeal to the types of patterns of relations that normally hold between the agent and the environment. A factivist

---

14 The qualification is included to anticipate objections that it is an analytic truth that an entity is a creature only if it is a representational being.
epistemology that focuses on the token occurrences appears to be expla-
natorily impoverished in comparison.

6.3.4 Mode of Massive Deception – Illusion and Hallucination

It is hard to assess a massive deception scenario without some specifi-
cation of how the subject is deceived. Even if we assume the possible
existence of demons, it should be specified how the demon constructs
the deceptive environment. One crucial distinction among many is that
between illusion and hallucination versions of the cases. To illustrate the
significance of this variation, I will consider illusion and hallucination
versions of NED cases beginning with the latter:

**RECENT NED HALLUCINATION**

A middle-aged man, Sam, manages to irk a demon who, in turn, decides to punish
Sam by deceiving him as much as possible. The demon has the ability to affect
Sam’s brain so as to make him hallucinate. So, it puts that ability to use in making
all of Sam’s beliefs about his nearby surroundings arise from a hallucination rather
than from perception.

In this hallucination scenario, the deception occurs in the agent’s mind
(or, for physicalists, nervous system) rather than in the external environ-
ment. Consequently, there are stronger reasons to reject that the agent is
warranted. Or, at least, there are reasons to reject that the agent is as
warranted as her counterpart in an ordinary veridical case. In *COIN
ILUSION*, an important feature is that Silas forms his belief by employing
a well-functioning cognitive competence that would have led to a true
belief had the circumstances been epistemically normal.

In the massive hallucination case, in contrast, Sam does *not* exercise a
well-functioning cognitive competence that is truth-conducive in his
normal environment. Further specification matters for whether we should
hold that he exercises a type-distinct and altogether unreliable cognitive
competence or whether his ordinary cognitive competences are continu-
ously malfunctioning. But in either case, moderate epistemic externalists
have reason to deny that Sam is warranted. However, none of these
reasons support a factive turn. Sam may be blameless in consistently
forming false beliefs. But, contrary to what proponents of a factive turn
often appear to presuppose, moderate epistemic externalists agree that
blamelessness is insufficient for warrant (Burge 2003; Graham 2012;
Gerken 2013: 278–281). Let us contrast this case with one of illusion.

**RECENT NED ILLUSION**

A middle aged man, Sim, manages to irk a demon who, in turn, decides to punish
Sim by deceiving him as much as possible. The demon has the ability to make a
force field around Sim that consists of holograms and equally illusory auditory and tactile inputs. So, it puts that ability to use in making all of Sim’s beliefs about his nearby surroundings arise from illusion rather than from veridical perception.

The present case postulates holograms that emit sounds and generate force fields that yield indiscernible tactile stimuli. This raises questions that we do not encounter in more simple one-off cases of visual illusion. How does the subject eat? How is an illusion of this process generated? An illusion version of the NED scenario in which all the subject’s perceptual beliefs are illusory requires that a lot of things be specified. While we can learn from reflecting on extreme thought experiments, doing so requires serious work in specifying all the relevant details of the scenario. Failing to do so is to leave open aspects of epistemological significance.

Setting aside these complications, the illusion version of the NED case is a better candidate case of warrant persisting in a deceptive environment than the hallucination version. In the illusion case, we do not attribute a malfunctioning cognitive competence to Sim. Had Sim exercised the relevant cognitive competences in the environment in which they were established, he would have formed true beliefs. So, there are moderate epistemically externalist reasons to regard Sim as warranted rather than merely as blameless or excused.

Nevertheless, the massive illusion scenario raises vexed questions. I have already mentioned the issue of mental ‘slow-switching’ concerning whether Sim’s beliefs remain of the same type. But most of all I want to reemphasise that in the absence of specification of the basic nature of the massive illusion, such NED-style cases may be at the periphery of what we can reach a sound verdict about. There are empirical reasons to assume that our intuitive judgments about such far-fetched cases are less robust (Wright, 2010; Gerken 2017). Furthermore, the theoretical considerations that underwrite the considered judgments about COIN ILLUSION and MUSEUM OF ILLUSIONS apply less clearly. In both of these cases, we can specify facts about the visual system that underwrite the judgment (Burge 2005, 2010). But a NED case in which all the subject’s perceptual beliefs are supposed to be deceived does not, without further specification, provide an adequate basis for a considered judgment. That is why I am inclined to regard RECENT ENVATMENT as a limiting case that abstracts away from a number of issues that are in fact of epistemological significance.

6.3.5 Some Lessons

I have only scratched the surface of the complicated relationship between massive deception cases, the nature of mental representation and the
nature of epistemic rationality. But this fact illustrates a central point: Massive deception cases, such as NED and BIV cases, are dramatically different from one-off illusion cases and from one another. These differences are epistemologically significant. Even in cases where the final verdict is the same, it may be on the basis of very different diagnoses. For example, the unified externalist explanation of why S is unwarranted in cases involving recent hallucination differs greatly from the unified externalist explanation of why S is unwarranted in cases involving massive illusion. Consequently, the cases cannot simply be addressed as one group of bad cases. Rather, proponents of a factive turn in epistemology are as obliged as everyone else to respect the epistemologically significant differences.

6.4 Conclusions and Challenges

Responding to problems of massive deception may advance our understanding of some of the knottiest epistemological issues. I have sketched how an account that unifies externalism in epistemology and in philosophy of mind has resources to explain why the subject is unwarranted in some cases of massive deception and warranted in others.

From a methodological point of view, this conclusion indicates the significance of specifying the scenarios invoked in substantive arguments. Of course, there are grey-zone scenarios in which it is unclear whether the subject should be regarded as epistemically warranted. Nevertheless, addressing well-specified massive deception scenarios may further progress on such hard cases. Overall, the discussion indicates a number of challenges for the proponent of a factive turn. So, I will conclude by highlighting some of these challenges.

Throughout, I have argued that it is crucial to articulate arguments from specific cases. Variations in the details of the scenario matter for factors (mental state individuation, nature of cognitive competence, etc.) that impact the subject’s epistemic status. These arguments suggest the following challenge:

1. Challenge from Specificity

Proponents of a factive turn should argue from specific cases rather than from a generic notion of ‘bad cases’.

Furthermore, I have argued that moderate non-factive epistemic externalists explain the lack of warrant in some massive deception cases by invoking the same unified mind-epistemological externalist principles that explain why warrant persists in one-off perceptual cases. This suggests the following challenge.
2. Challenge from Moderate Epistemic Externalism
Proponents of a factive turn should argue against moderate externalist accounts rather than merely against epistemically internalist accounts.

I have argued that arguments that the subject is not warranted in massive deception cases do not entail or even suggest that the subject is not warranted in local or one-off deception cases. So, proponents of a factive turn who do not admit of any non-factive epistemic notions of warrant, rationality or reason face the following challenge.

3. Challenge from One-Off Cases
Proponents of a factive turn should account for local and one-off deception cases rather than merely for massive deception cases.

I opened the chapter with a probabilistic case in which a subject could be argued to be epistemically warranted in a false lottery belief. This type of case raises the following challenge:

4. Challenge from Lottery Cases
Proponents of a factive turn should explain the asymmetry between one-off perceptual cases and probabilistic cases, or argue against probabilistic cases.

I have sketched how non-factive epistemologists may pursue an account that is unified in the sense of integrating epistemic externalism and externalism about the individuation of mental state types (anti-individualism). This raises the following challenge.

5. Challenge from Externalism in the Philosophy of Mind
Proponents of a factive turn should clarify its relation to the aspects of philosophy of mind that pertains to the individuation of types of mental states and cognitive competences.

Challenges (1)–(3) are adequacy constraints on the case for a factive turn in epistemology. Challenge (4) from lottery cases is a more general desideratum for an epistemological theory, but one that may be particularly hard to meet for proponents of a factive turn. Likewise, Challenge (5) from the philosophy of mind is merely a desideratum. But it is nevertheless urgent to address it given that opposing moderate epistemic externalist accounts may provide an approach that is unified with a moderate brand of externalism in the philosophy of mind (namely, anti-individualism). As things stand, the non-factive opposition may adopt a unified approach that provides more nuanced diagnoses of specific deception cases than proponents of a factive turn appear to be able to deliver.

I set forth these challenges to provide a broader methodologically oriented perspective and in the hope that proponents of a factive turn in epistemology will address them head on. However, my more restricted
aim has been to argue that we have yet to see a sound argument from cases of massive deception to the conclusions that epistemic reasons, epistemic warrant, epistemic rationality, or epistemic norms are factive. I conclude that the quarters of epistemology that may have undergone a factive turn should take a U-turn.\textsuperscript{15}

\textsuperscript{15} The chapter was presented at the Factive Turn in Epistemology Workshop at the University of Vienna in May 2015. Thanks to the audience for a helpful session, to Veli Mitova and Arturs Logins for subsequent correspondence and to a Cambridge University Press referee for very helpful comments.