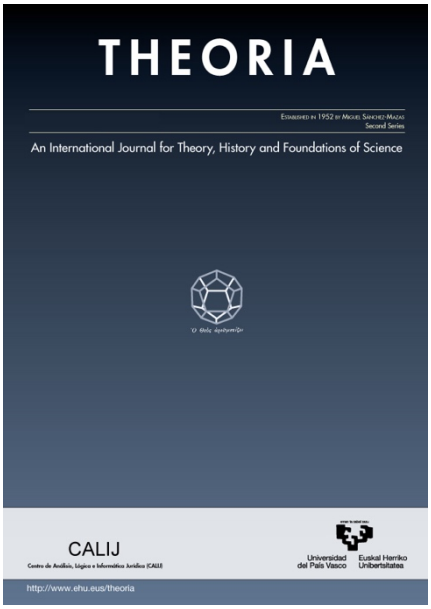
JOURNAL PRE-PROOF

Asymmetries between ‘you’ and ‘I’

Matheus Valente

DOI: 10.1387/theoria.26508

Received: 12/06/2024
Final version: 02/06/2025



ASYMMETRIES BETWEEN 'YOU' AND 'I'

Asimetrías entre 'tú' y 'yo'

Matheus Valente
University of Lisbon

***ABSTRACT*:** Can others grasp my first-person thoughts, or are such thoughts inherently private? Philosophers disagree: some argue that first-person thoughts are apprehensible only by their owners, while others contend that they can be shared through communication—expressible by 'you' as readily as by 'I'. In this paper, I set out to clarify the stakes of this age-long dispute. Taking J. L. Bermúdez's forceful defence of shareability as the backdrop of my discussion, I examine how the intersubjective availability of thoughts interacts with issues concerning the objectivity of thought, testimonial knowledge transmission, and rational action. The bulk of this paper is an elaboration of the Asymmetry Argument, which grounds the privacy of first-person thoughts in the need to explain how thinkers who believe and desire the same as each other might nonetheless have distinct reasons for action. If successful, the argument reveals how first-person thoughts cannot be shareable in a philosophically significant sense without compromising their fundamental connection to motivating reasons for action.

Keywords: thoughts, first-person, propositional attitudes, indexicals, action.

***RESUMEN*:** ¿Pueden otras personas comprender mis pensamientos de primera persona, o son estos pensamientos inherentemente privados? En filosofía no hay consenso: una corriente defende que los pensamientos de primera persona sólo son aprehensibles por sus poseedores, mientras que otra sostiene que pueden compartirse mediante la comunicación – siendo expresables tanto por «tú» como por «yo». En este artículo, me propongo aclarar los puntos clave de esta antigua disputa. Partiendo de la contundente defensa de la compartibilidad de los pensamientos de primera persona realizada por J. L. Bermúdez, examino cómo la disponibilidad intersubjetiva de esos pensamientos interactúa con cuestiones relativas a la objetividad del pensamiento, la transmisión del conocimiento por testimonio y la acción racional. El núcleo de este artículo es una elaboración del Argumento de la Asimetría, que fundamenta la privacidad de los pensamientos de primera persona en la necesidad de explicar cómo pensadores que creen y desean lo mismo podrían, no obstante, tener distintas razones para actuar. De tener éxito, el argumento revela cómo los pensamientos de

primera persona no pueden considerarse compartibles sin comprometer su conexión fundamental con las razones para actuar.
Palabras clave: pensamiento, primera persona, actitudes proposicionales, deícticos, acción.

SHORT SUMMARY: When I tell you, 'I'm tired', you may report what I've just said as, 'You're tired'. Intuitively: one thought, two perspectives. Yet if a thought's identity depends on its perspective, then 'you' and 'I' couldn't possibly express the same thought. In this paper, I examine the perspectival dimensions of thought and conclude that first-person thoughts are private, apprehensible only by their owners.

---

1. Introduction

When confronted with the peculiar features of first-person thought (those usually expressed with 'I'), many philosophers have felt compelled to hold that they differ from third-person thoughts in being apprehensible only by their owners.[1] Frege (1956) himself wrote that "everyone is presented to himself in a special and primitive way, in which he is presented to no one else", a remark often cited in defences of the thesis that first-person thoughts are private (henceforth, **Privacy**). Unsurprisingly, **Privacy** has not gone wholly unchallenged, even by authors of a broadly Fregean persuasion, such as José Luis Bermúdez.[2]

In a series of works spanning decades, Bermúdez has developed a systematic account of first-person thoughts among whose original features is a forceful defence of their objectivity and shareability. More specifically, Bermúdez argues for the so-called symmetry constraint - henceforth, **Symmetry** - which holds that "an account of the sense of 'I' must allow tokens of 'I' to have the same sense as tokens of other personal pronouns such as 'you' in appropriate contexts" (2017a, p. 61). The idea behind **Symmetry** is as simple as it is plausible: when you tell me something about yourself with 'I', I acquire knowledge which I could report back to you with 'you', suggesting that reciprocal uses of the first- and second-person pronouns not only might

[1] Authors who subscribe to the privacy of first-person thoughts include Evans (1981, 1982), Heck (2002), Morgan (2009), and Stanley (2011).

[2] The main works where Bermúdez discusses first-person thoughts and the sense of 'I' are: 2005, 2011, 2017a, 2017b, 2019. All Bermúdez's citations in this paper come from his monograph *Understanding "I"* (2017a), where his main ideas on the topic are given its most systematic presentation.

happen to co-refer but also to match in cognitive significance, and so, to underly the apprehension of a single thought.[3]

Despite extensive debate, the stakes of this dispute remain unclear. Some of this hinges on which theoretical roles thoughts are supposed to play. While most agree that thoughts are the cognitively significant contents of attitudes such as belief, it's unclear whether this alone suffices to adjudicate between **Symmetry** and **Privacy**. This paper aims to achieve more clarity on the significance of this dispute.

Sections 2 and 3 assess Bermúdez's two main positive arguments for **Symmetry**: *the argument from objectivity* and *the argument from testimonial knowledge*.[4] I argue that neither succeeds; indeed, the latter one might actually backfire, indirectly supporting **Privacy** due to issues with the assumption that thought sharing is necessary for the transmission of knowledge through communication. Section 4 turns to the interplay between thoughts and motivating reasons for action, introducing the concept of *confidants* to capture the relationship between ideally rational agents whose beliefs and desires are maximally shared through an idealised communicative exchange.[5] This sets the stage for the Asymmetry Argument (Section 5), which concludes that **Privacy** alone can explain how such agents might nonetheless have distinct reasons for action. After rebutting objections to the argument (Section 6), I conclude that the need to explain how agents who maximally agree with each other might diverge in their reasons for action gives rise to a direct argument for the privacy of first-person thoughts.

2. *The Argument from Objectivity*

Though deeply influenced by Gareth Evans (1981, 1982), Bermúdez departs form him on a few important issues, including on the shareability of first-person thoughts. One source of their disagreement centres on the objectivity of thoughts, a central Fregean commitment which both

[3] By 'the same thought' I mean 'thought-tokens of the same type'. In Bermúdez's terminology, these would be thought-tokens constituted by the same token-senses (Bermúdez, p. 80-97).

[4] A disclaimer: I discuss only a small fraction of Bermúdez's rich account of the sense of 'I'. Among its many invaluable contributions, we find: a defence of **Essential Indexicality** (roughly, the thesis that first-person attitudes must necessarily be involved in explanations of actions; Bermúdez, 2017a, chapter 1; 2017b), a careful reevaluation of the legacy of Evans' seminal work on the first-person (Bermúdez, 2017a, chapter 6), and an original account of immunity to error *via* misidentification and past-tense judgements based on autobiographical memory (Bermúdez, 2017a, chapter 7; 2011).

[5] The interplay between first-person thoughts and action has figured prominently in recent work. For a survey, see Ninan (2016; 2021). See also: Prosser (2005; 2023; *forthcoming*), Longworth (2013, 2014), Weber (2014), Recanati (2016; *forthcoming*), Torre (2018), Valente (2018), Torre & Weber (2021; 2022), Gray (2022), Lin (2022), Verdejo (2017; 2019; 2020; 2025).

authors endorse. While Evans argues that private thoughts can be objective, Bermúdez insists that first-person thoughts can only be objective if shareable.[6]

What is it for a thought to be objective? Most generally, a thought is objective if it is irreducible to the contents of a particular consciousness (i.e. distinct from, as Frege would put it, mere ideas [*Vorstellung*]). Evans and Bermúdez seem to agree that a thought is so irreducible when it fulfils at least one of the following conditions:

(1) The thought exists and has a truth-value independently of being apprehended;

(2) The thought is shareable (apprehensible by multiple thinkers);

Evans defended **Privacy** by appealing to (1): "an unshareable [private] thought can be perfectly objective – can exist and have a truth-value independently of anyone's entertaining it" (Evans, 1981, p. 313). Bermúdez disagrees, arguing that indexical thoughts (of which first-person thoughts are a key instance) are constitutively connected to the episodes of thinking whereby they are apprehended (p. 62-66):

> [T]he whole point of token-reflexivity, and indexicality in general, is that the identity of the thought is determined by the context in which it is thought—in which case, there is no thought without an episode of thinking. (p. 63)

On that basis, Bermúdez concludes that indexical thoughts can only be objective if they fulfil (2), i.e. if they are shareable. But why should the identity of indexical thoughts be constrained in the way which Bermúdez outlines? What follows is my interpretation of his reasoning.

Suppose I thought 'I am tired' last Sunday at noon.[7] The thought I thereby apprehend - call it 'T' - is referentially equivalent to (the thought which I could have expressed as) 'MV was tired on Sunday at noon'. However, their cognitive significance differs. For example, I could know that I'm tired without knowing that I'm MV or what time it is, a familiar point which shows that my first-person thought T is potentially different from its third-person analogue. One reason why that is so stems from the independent ways in which these thoughts' truth-conditions are fixed. In particular, T's truth-conditions are fixed by contextual features of its episode of apprehension such as the fact that its thinker is MV and that it's entertained on a particular Sunday at noon. Presumably, this entails that T would not so much as exist if it had not been apprehended in a

[6] Frege's emphasis on the objectivity of thoughts is arguably one of his central philosophical contributions (Dummett, 1980).

[7] For brevity, I'll often write "x thinks/believes 'I am F'" as short for "x apprehends/believes a thought which they could have expressed by uttering 'I am F'".

context with these particular features, which plausibly tells against Evans' contention that its objectivity could be grounded on fulfilment of (1).

However, we immediately face problems when trying to make Bermúdez's proposal more precise. Surely, Bermúdez does not mean to say that T could not possibly have been apprehended in any other context. Suppose, for example, that I knew on Saturday 8pm that I would be tired on Sunday at noon, and thought 'I will be tired in 16 hours'. Presumably, the thought I would then apprehend could be none other than T, after all, if I went on to track the passage of time for 16 hours and, finding myself tired just as I knew I would, thought 'I am tired', I would seem to be merely re-expressing the content of my earlier judgement as opposed to judging something new. To be sure, there are some who would challenge that intuitive claim, but these would be people who balk at the idea that an indexical thought could remain invariant across changes of perspective, and so, whose fundamental commitments are at odds with the core rationale behind **Symmetry**, i.e. that a thought expressible in the first-person perspective with 'I' could equally be expressed second-personally with 'you'. But if T could have been apprehended on Saturday, then it could have existed and have a truth-value independently of being apprehended on Sunday, which shows that its purported constitutive connection to a particular episode of thinking is less clearcut than it initially might seem.

Bermúdez could insist that T's existence depends on its being apprehended in *some* context, though not on any particular one. Notice, however, that some of these considerations appear to pertain more closely to the temporal indexicality of T as opposed to its involvement of the first-person sense of 'I'. I take it to be uncontroversial that a thought can be indexical in at least two ways: by involving the senses of personal pronouns like 'I' and 'you' or the senses of temporal indexicals like 'now' and 'soon'. T appears to be indexical twice over: not only it concerns myself as myself, it also makes indexical reference to time. In contrast, the thoughts 'I am tired at 12:00 GMT on Sunday, January 3rd, 2025' and 'today is Sunday' seem to be indexical in a single sense: the first is first-personal but not temporal, the latter is temporal but not personal. This raises a concern: in focusing on T to investigate the objectivity of first-person thoughts, we risk equivocating between issues which arise from temporal indexicality with possibly independent issues having to do with the first-person.[8] To avoid any mixup, we should turn attention to pure breed first-person thoughts, of which these are plausible examples:

[8] For a recent work addressing the peculiar differences between personal (*de se*) and temporal (*de nunc*) indexicality, see Morgan (2024).

On Sunday, MV thinks: (T2) I am tired at 12:00 GMT on Sunday, January 3rd, 2025.

On Sunday, MV thinks: (T3) I am tired at some/all times.

On Sunday, MV thinks: (T4) I am human *simpliciter*.[9]

As far as I can see, Bermúdez's criticism of Evans is harder to motivate on the basis of examples like T2, T3, and T4. Why, for example, should the identity of T2 - the thought whereby I self-attribute tiredness at a particular, non-indexically specified, time - be in any way constrained by when or where it is apprehended? Perhaps Bermúdez could insist that these thoughts' truth-conditions are fixed by an indexical act of reference to a particular individual - either via 'I' or, as **Symmetry** allows, via 'you' - which is subsidiary to an episode of apprehension. I agree that that something gets to be the referent of a token of 'I' or 'you' only if these tokens were indeed produced in the right way, but that strikes me as a trivial claim equivalent to saying that a token singular term ought to exist in order for it to have a referent. The crucial question is whether it makes sense to say that thoughts like T2, T3, and T4, could have existed and have a truth-value even if nobody had ever entertained them, and so far I have not been able to pinpoint a clear reason for saying that it does not. On the other hand, a reasoning like the following moves me into saying that it does make sense.

Suppose that I have known all along of this person who I fail to recognise as myself that she is tired at 12:00 GMT on Sunday, January 3rd, 2025. At some point in February, I realise that this is person is none other than myself, and so, discover that it is me who was tired at 12:00 GMT on Sunday, January 3rd, 2025, thus apprehending T2 for the first time around. Intuitively, the truth which I discovered, T2, would have remained true even if neither I nor anyone else had discovered it, after all, what else could it mean for something to be discovered in the first place? But Bermúdez must instead hold that my act of apprehending T2 for the first time is what brings T2 into existence, and so, what determines its being a true thought as opposed to nothing at all. This seems to get things the wrong way around. More generally: if first-person thoughts are neither true nor false before apprehension, then in which sense could their truth amount to a discovery?

Furthermore, in the scenario under consideration, I have apprehended all of the senses constitutive of T2 - the objectual sense of 'I' and the sense(s) of 'being tired at 12:00 GMT on Sunday, January 3rd, 2025' - prior to concatenating them as I do when I finally apprehend T2, after all, I have thought about myself and, independently, about someone who was tired on that specific date. Presumably, this entails that the constituents of T2 were available for thought prior to my

[9] Whenever F is an essential property such as being human, I assume that it makes sense to think of an individual as being F *simpliciter* (as opposed to: being F *now/in the past/in the future*).

apprehending it. If so, then Bermúdez seems committed to denying that a thought exists and has a truth-value even though all of its constituents have been separately apprehended in a meaningful way. But why should a thought's existence have to wait any longer?

As said before, the objectivity of a thought most fundamentally requires its being irreducible to mere ideas. Plausibly, an idea's subjectivity has at least partly to do with its fleeting, ephemeral nature; but thoughts like T2, T3, and T4, certainly can be entertained multiple times (even if only by a single subject), and so, wouldn't seem to be fleeting in a similar way. Interestingly, this fits with another traditional conception of objectivity, namely, that of existence unperceived. For example, one might reason by analogy and claim that, just like an objective particular is one which may be re-encountered after ceasing to be perceived, an objective thought is one which may be re-apprehended after previously having been so. Why, then, would objectivity require first-person thoughts to be apprehensible by multiple thinkers over and above requiring them to be apprehensible multiple times by a single one?

Rather than a set of knockdown objections, the above should be read as a plea for further clarification regarding the rules of this debate. As things stand, I'm not convinced that **Symmetry** and the shareability of first-person thoughts can be vindicated solely on the basis of their objectivity.

3. The Argument from Testimonial Knowledge

Bermúdez bases his other main argument for **Symmetry** on another broadly Fregean principle, namely, that "a Fregean theory of sense is a theory of understanding" (p. 67), and so, that there ought to be an "equation between understanding a sentence and grasping the thought that it expresses" (p. 27). On a natural interpretation, this requires thinkers who understand each other to apprehend the same thought. Bermúdez (p. 76) is careful to accept that not all forms of understanding might require thought sharing, but he explicitly claims that this must be so when knowledge gets transmitted via testimony. Thus, Bermúdez's main claim is that "knowledge through testimony typically requires the person acquiring the knowledge to think the very thought expressed by the speaker" (p. 76). If he's right, then **Symmetry** could be supported by the undeniable fact that one may acquire knowledge from another's testimony involving 'I', knowledge which one would naturally report back with 'you'.

However, Bermúdez's claim faces at least two serious problems. First, the relevant data on testimonial knowledge transmission appears amenable to an account which deploys no notion more sophisticated than referential contents and independently motivated epistemic concepts such

as safety and reliability. Second, so-called *non-transitivity cases* appear to show that knowledge transmission can occur even when thought sameness is absent. I now explain each of these issues.

A quick look at the recent literature on referential communication shows that the following is common ground among authors of virtually all persuasions: sameness of referential content is not sufficient for understanding and knowledge transmission via testimony.[10] For example, if I tell you 'Clark Kent is F' and you form the belief 'Superman is F', then, assuming that neither of us know that Clark Kent is Superman, your belief would have the same referential content as mine but wouldn't count as knowledge. A Fregean might say that your mistake was failing to apprehend the same thought as mine. But a non-Fregean - call her a Russellian - will judge this move *ad hoc*, for we seem capable of explaining your failure in purely referential-*cum*-epistemic terms. For example, "the Russellian might claim that the process involved in communication [knowledge transmission] must reliably produce coreferential thoughts, and so secure non-accidental, or not merely lucky, co-reference" (Goodman, forthcoming, p. 4). In other words, you failed because you got the right referential content by sheer luck as opposed to in a suitably reliable way. As far as I can see, there's nothing immediately unobjectionable about that kind of story (not to mention the fact that it's more theoretically parsimonious, after all, it can be endorsed independently of any prior commitments regarding the role of thoughts in interpersonal exchanges).

A Fregean might retort that what the Russellian calls 'apprehending the right referential content in a reliable way' just is what she calls 'apprehending the same thought'. But there are problems with this response. For example, some have argued that thought sameness can itself be arrived at as a matter of luck (Byrne & Thau, 1996). Suppose that, by sheer coincidence, you and I independently introduce the name 'Winston' for the amnesiac in Room 101. I then tell you, 'Winston will never recover' and you form the belief you'd express by the same sentence. Plausibly, your belief won't count as knowledge: how do you know who I'm talking about? However, it's possible to argue that the thought which you believe is the same as that which I have expressed, after all, the senses you and I associate to our uses of 'Winston' appear to be the same, namely, a sense for the amnesiac in Room 101. If so, then even Fregeans will need to appeal to anti-luck considerations in order to characterise knowledge-supporting understanding, which again suggests that we might do better by following the Russellian in appealing to nothing but referential-*cum*-epistemic to accommodate the relevant phenomenon.

[10] Cases with this structure are often called 'Loar-cases' due to their origin in Loar (1976). For recent debate involving such cases, see Onofri (2019), Valente (2021), Prosser (forthcoming), Recanati (2016, forthcoming).

To respond, a Fregean needs to come up with principled argument for the importance of having a thought-involving account of the relevant data. Let's assume that this can be done. Still, it's not obvious whether this would result in any direct arguments for **Symmetry**. For one thing, knowledge-supporting understanding might require some suitable relation between the relevant thoughts without requiring them to be the same. Indeed, some have advanced proposals in that spirit. McDowell (1984, p. 290) is a good example:

> […] there is no obvious reason why [Frege] could not have held […] that in linguistic interchange of the appropriate kind, mutual understanding - which is what successful communication achieves - requires not shared thoughts but different thoughts which, however, stand and are mutually known to stand in a suitable relation of correspondence.

Therefore, even Fregeans can eschew the need for thought sameness as a condition on testimonial knowledge transmission. But why would any Fregean go down that route instead of sticking to the simpler view endorsed by Bermúdez? There might be multiple reasons why, but I'll resort myself to presenting that which I take to be the most general one: non-transitivity cases (see below) show that thought sameness not only isn't sufficient for knowledge transmission, it's not even necessary.[11]

Suppose that I meet Johannes on Monday and tell him that I'm human. I then meet him again on Tuesday and, again, tell him I'm human. It's certainly possible for Johannes to understand me both times, and so, to come to know all that I told him. Presumably, the thought I express to him is the same both times (T4 from Section 2). So, if Bermúdez's view is correct, then what Johannes apprehends and comes to know on Monday and Tuesday is thought T4. However, Johannes might have failed to recognise me on Tuesday as the same person who talked to him on Monday. Surely, this type of identification failure happens in real life all the time. But if he did, then the thoughts he apprehends on each day are cognitively distinct for him, i.e. he could in principle believe one while disbelieving the other (and rationally so). Of course, this just is to say that Johannes apprehends different thoughts each time, contradicting Bermúdez's assumption that knowledge transmission requires thought sameness.

To be sure, I admit that non-transitivity cases are anything but unobjectionable, e.g. one might suspect that Johannes' recognition failure tells against his coming to know what I tell him on Tuesday. I agree that recognition failures might sometimes lead to misunderstanding, but doubt that this needs always occur. It would be a stretch to claim that I cannot acquire knowledge from

[11] Non-transitivity and thought sharing has been a focus of recent debate by, among others, Goodman (forthcoming); Gray (forthcoming); Valente & Onofri (2023); Valente & Verdejo (2022), Cumming (2013), Fine (2007).

your utterance 'NN is F' if I fail to see that NN is the same person you were talking about some time before. As far as I can see, this is analogous to how I could look at an object and learn that it is red while later, failing to recognise the same object as before, look at the same object and learn that it is red, ending up with two pieces of knowledge that are based on similar perceptual experiences of the same thing.[12] So, if one agrees that Johannes could understand me regardless of not recognising me, then he can only come to know what I tell him on Tuesday if he apprehends a thought that is different from mine. In addition, one might claim that Johannes would have failed to understand me on Tuesday if he apprehended the same thought as mine, for then he would have been assuming that I'm the guy from yesterday in the absence of good reasons for doing so.

How would Bermúdez respond? In a couple of places, he suggests an argument as follows (cf. Bermúdez, pp. 75-77). Suppose a bystander eavesdropped on my Tuesday conversation with Johannes. If the bystander understood my 'I am F' as well as Johannes' ensuing report 'you are F', then it would be irrational for her to take contrasting attitudes to what Johannes and I have said. Doesn't this show that Johannes and I express the same thought by 'you' and 'I' on that occasion? I don't think so. This at most shows that the bystander couldn't have understood us if she associated distinct thoughts to each of our utterances. This puts demands on the bystander's thought(s) while leaving it open whether it/they ought to be same as mine, Johannes', or to both. More generally, rational cotenability is a relation between the thoughts of a single thinker at a time, and so, are bound to be of little help in assessing distinct thinkers' thoughts for identity or distinctness (Evans 1981).[13]

In conclusion, even if we grant the assumption that an account of knowledge transmission via testimony must be framed in terms of thinkers' thoughts, thought sameness might still neither be sufficient nor necessary for it to take place. For these reasons, I'm sceptical about whether any straightforward argument from knowledge transmission to **Symmetry** will be forthcoming. Given this negative conclusion, we're advised to look elsewhere in order to adjudicate between **Symmetry** and **Privacy**. In the next section, we turn to motivating reasons for action.

[12] Furthermore, the fact that these cases involve multiple contexts/conversations strikes me as irrelevant, for intra-contextual analogue cases could easily be constructed (e.g. if 'NN' is a very common name, you might wrongly but rationally take my series of utterances about NN in a conversation as being about distinct people with the same name; for related considerations, see Verdejo, 2025).

[13] This draws from Evans (1981, pp. 293-5), but see also Recanati (forthcoming) and Verdejo (2025). An anonymous referee points out further issues with the intuitive Fregean criterion for thought individuation, in particular, a version thereof which takes thoughts to be the same when contrasting attitudes towards them are not rationally cotenable. As the referee points out, this criterion over generates, for it seemingly entails that analytic or a priori statements express the same thought (e.g. "all planets revolve" and "all bachelors are unmarried"). This seems right. In addition, I may add, it has the unfortunate consequence of entailing that co-referential thoughts known to be co-referential are the same (e.g. if I know that I am NN, then I express the same thoughts by 'I am F' and 'NN is F').

4. *Attitudes, Reasons, and Confidants*

As Bermúdez reminds us, it's a "basic tenet of propositional attitude psychology that people act the way they do because of what they believe" (p. 47). Indeed, it's almost a truism that a rational agent who believes that Φing is the optimal way of fulfilling her most pressing desires, ideally, will be motivated to Φ (and so, will predictably try to Φ). This gives rise to a sort of platitude about ideally rational agents, namely, that their motivating reasons for action - those in light of which she intentionally acts (henceforth, her 'reasons')[14] - supervene on her attitudes (e.g. beliefs, desires, etc). This supervenience claim can be given an intra-personal diachronic formulation and an inter-personal synchronic one:

> **CHANGES**: an ideally rational agent's reasons cannot change if her attitudes remain the same.
>
> **TWINS**: ideally rational agents cannot have distinct reasons if they have all the same attitudes.[15]

CHANGES says of an agent at distinct times what **TWINS** says of distinct ones at the same time. Presumably, both principles stand or fall together. This becomes even more crystalline if we reframe **CHANGES** as the principle that time-slices of an ideally rational agent cannot have different reasons if they have the same attitudes. Absent reasons to think that the relation between the attitudes and reasons of time-slices of an agent differ from those of time-slices of distinct agents, **CHANGES** and **TWINS** just are two faces of the same coin.

Since **Symmetry** is an inter-personal principle, let's focus on **TWINS** (I'll have more to say about **CHANGES** in 6.1). I take it to be uncontroversial that **TWINS** (or something in the vicinity) is supported by our ordinary practice of explaining and predicting intentional action via the attribution of attitudes to agents under the assumption of their rationality. If two rational agents'

---

[14] Motivating reasons, i.e. reasons in light of which an agent acts, are typically contrasted with normative reasons, i.e. reasons that count in favour of actions independently of whether the agent is aware or acts in their light (Alvarez & Way, 2024). My discussion primarily concerns the former. For recent discussion on motivating reasons and similar proposals regarding their connection to agents' attitudes, see Singh (2019) and Wang (forthcoming).

[15] My formulation of **TWINS** is derived from Hunter's *Inter-Personal Reasons*: "There can be no difference in what two people have reason to do without a difference in what they believe or desire" (Hunter, 2017, p. 693; 2022, p. 158; see also Cappelen & Dever 2013, p. 52). It's also inspired by the law-like generalisation which Ninan refers to as *EXPLANATION* (Ninan, 2016) and as *Thesis (IV)* (Ninan, 2021). However, there are some important differences. First, Ninan's generalisations quantify over actions instead of reasons. Second, they include a *ceteris paribus* clause ("if all else is equal") instead of being restricted to ideally rational subjects. Analogous considerations apply to Prosser's (forthcoming) *Weak Thought-Action Principle*. While **TWINS**'s more restricted scope makes it more artificial, it also makes it more precise, which pays off when it is deployed in the argument laid out in the next section.

being of the same mind did not correlate with their intentionally acting alike, then the very idea "that people act the way they do because of what they believe" would seem out of place. To be sure, **TWINS** encapsulates a form of internalism about reasons that is not universally accepted. However, it's an internalism of the most plausible kind for it explicitly concerns motivating reasons for actions, and so, shouldn't be incompatible with most forms of externalism or objectivism about normative reasons (Singh 2019, Wang, forthcoming).

A few caveats about **TWINS** are in order before proceeding. First, it's a principle of ideal rationality. I will offer no precise definition of an ideal reasoner, but they must at least be perfectly aware of which attitudes they hold as well as being maximally compliant with the norms of epistemic and practical rationality. This vague characterisation suffices to avoid a few distracting objections (e.g. cases involving reasoning mistakes due to cognitive or temporal limitations, failure to keep up with one's evidence, forgetfulness, behaviour motivated by irrationally-held attitudes, akrasia, weakness of the will, etc). In addition, notice that **TWINS** doesn't presuppose that it's possible for distinct agents to have all the same attitudes, after all, doing so would seemingly beg the question against proponents of **Privacy**. It's then helpful to introduce a label applying to agents who share each other's attitudes to the maximal possible degree which is compatible with there being attitudes towards private thoughts.

Let agents x and y be *confidants* if (1) they are mutually acknowledged ideally rational agents, (2) they have engaged in an idealised communicative exchange whereby they have mutually expressed all of their attitudes, and, consequentially, (3) whose post-communication attitudes are the result of updating on everything which they have just learned from each other.[16] In other words, confidants are mutually acknowledged rational peers who pool their information together via communication, and so, end up in attitudinal states which result from factoring in everything they have learned from each other. So, confidants have common knowledge of each other's attitudes. But not only that: since they take each other as peers in all epistemic and practical rationality-relevant respects, they will be moved towards calibrating their own attitudes in light of what they learn from one another. By doing so, their resulting attitudes will be maximally convergent: not only will they know each other's beliefs and desires, but, to the extent that this is possible, they will generally believe and desire the same as each other too.

[16] The concept of *confidants* is an adaptation of Caie's (2018) concept of *epistemic confidants*, which he deploys in an independent – but not wholly unrelated - investigation on the limits of objective disagreement between agents who have common knowledge of each other's doxastic state. One key difference is that, given my focus on rational action, I'm as much concerned about non-doxastic attitudes like desire than about doxastic ones like belief - my *confidants* not only believe the same but desire the same as well.

There's a great deal of simplification in the above story, some of which I'll address in Section 6.3, but this much is enough for us to draw a few important conclusions. First, the degree to which the third-person attitudes of confidants can differ is minimal. Given their status as ideal reasoners and the fact that they will end up in possession of all the same evidence, it cannot be possible for them to harbour much disagreement about objective matters of fact (if any). Plausibly, this holds as much for their third-person doxastic attitudes as it does for their third-person non-doxastic ones, at least to the extent that we may assume third-person desires (e.g. my desire that candidate A win the election) to be subject to similar rational constraints as their doxastic counterparts.[17] Indeed, if we grant the assumption that all considerations relevant to the adoption of third-person attitudes are of a type which can be unproblematically transmitted via communication among ideal rational peers, then it's safe to say that confidants will just have the same third-person beliefs and desires as each other.

While the shareability of third-person attitudes shouldn't pose too much of a problem, indexical attitudes make matters more complicated. So as not to beg the question against **Privacy**, we must be open to the possibility that confidants will differ in their indexical attitudes even if they don't differ in their third-person ones. Still, proponents of both **Symmetry** and **Privacy** can agree with the following: the indexical attitudes of confidants are *maximally coordinated*, where this roughly means that there's a one-to-one mapping which takes attitudes expressible by one with 'I' to attitudes expressible by the other with 'you' (and vice-versa). If, for example, one believes 'I am F', the other believes 'you are F'; if one has a desire expressible as 'I desire to travel abroad', the other has a desire expressible as 'I desire *you* to travel abroad'. And so on.

Thus, confidants have *maximally coordinated* indexical attitudes, i.e. attitudes which differ only by a systematic permutation of indexically-encoded perspectives ('I' for 'you', 'my' for 'your', etc).[18] Those sympathetic to **Symmetry** will be advised to hold that *maximally coordinated* indexical attitudes just are the same, after all, these will be attitudes towards the same thoughts. Proponents of **Privacy**, on the other hand, will deny that, claiming that *maximally coordinated* indexical attitudes,

[17] Unfortunately, the literature on the limits of rational non-doxastic disagreement and interpersonal rational deliberation is scarce, so I might be glossing over some important difficulties. To err on the side of caution, bear the following qualification in mind: even if rationality permits confidants to somehow differ in their third-person attitudes (e.g. maybe it's just as rational to be a compatibilist than an incompatibilist), it's much less plausible that rationality could *require* confidants to so differ. If the only types of disagreement which confidants are permitted to harbour are ones which they have no rational obligation to harbour, then all the premises in the Asymmetry Argument (Section 5) remain correct given some minimal precisifications, some of which I'll address in Section 6.3. I'm grateful to Quentin Ruyant for pointing out some of these issues to me.

[18] Since I exclusively consider confidants who are in each other's presence at a time, there's no obvious reason to think that their other indexical attitudes will be any different. I'll come back to related issues in Section 6.3.

as intimately related as they may be, are attitudes towards distinct thoughts. The dispute between **Symmetry** and **Privacy** can thus be reframed as a disagreement about whether *maximally coordinated* indexical attitudes ought to, for the purposes of propositional attitude psychology, be identified or not. Granting that both camps accept that confidants have the same third-person attitudes, the deeper manifestation of their disagreement turns on whether we should take confidants to have all the same rational attitudes (third-personal and indexical) or not. This is the core rationale of the Asymmetry Argument, to which we now turn.

5. The Asymmetry Argument

The Asymmetry Argument can be framed as a three-premise inference to the negation of **Symmetry**:

(1) Ideally rational agents cannot have all the same attitudes if they have distinct reasons [**TWINS**].

(2) Some confidants have distinct reasons.

(3) If **Symmetry** is true, then confidants have all the same attitudes.

Therefore, **Symmetry** is false (from 1, 2, and 3).[19] [20]

Since (1) and (3) have been initially motivated in Section 4, I now comment on (2) before turning to consider objections to each premise in Section 6.

To establish (2), we need only a single case involving confidants who are motivated to act distinctly from each other. Purely general considerations show that this should be an easy task. Suppose, for example, that after completely pooling their information, two agents realise that something terrible will happen unless they immediately change the lightbulb above them, a task which requires one to hold the stairs for the other to climb it. Clearly, there's a sense in which what they will have to do is different - the lightbulb won't be changed unless they perform different sub-tasks. So, there's a sense in which even cooperative agents who maximally agree with each other about what needs

[19] In elaborating the Asymmetry Argument, I've been influenced by many prior works. The argument's roots arguably trace back to Perry (1977; 1979). More recently, related reasoning has been developed by Ninan (2016; 2021), García-Carpintero (2017), Valente (2018), Torre (2018), Torre & Weber (2021, 2022), Lin (2022), and Prosser (2005, 2023, forthcoming). Other works discussing the rational asymmetries between different indexicals include: Stalnaker (2008), Recanati (2016, forthcoming), Verdejo (2017, 2019, 2020), Bozickovic (2021), Gray (2022).

[20] Does **Privacy** follow from the negation of **Symmetry**? A more comprehensive analysis of the Asymmetry Argument would have to consider the alternative conclusion that first-person thoughts do not have absolute truth-values, and so, should be modelled as e.g. self-ascribed properties or centred contents (Lewis, 1979; also Ninan, 2016, 2021; Shaw 2019; Torre & Weber 2021, 2022). I discuss a related "relativistic" idea in 6.3.

to be done and about how that ought to be done might be required to act distinctly.[21] More generally, it's a truism that cooperation often requires coordination (of distinct actions), not imitation. Under the assumption that agents have distinct reasons when they're not motivated to perform the same action, then the above is sufficient to support (2).

Still, it'll be helpful to have a concrete case in mind. Consider this familiar scenario from Perry (1977, p. 494):

> **Bear Attack**: Ann and Bill are walking in the woods when a bear starts chasing Ann. Ann and Bill both realise that the bear is about to attack Ann (Ann thinks 'I'm about to be attacked by a bear' and Bob thinks 'You're about to be attacked by a bear'), and they both want to save Ann from the attack. They act differently. Ann curls up into a ball and plays dead. Bill, who witnesses the attack from a distance, runs to get help. (Adapted from Lin 2022, p. 2)

Bear Attack has been put to many uses in the literature. For present purposes, it should be seen as a case involving agents who, though broadly of one mind, act distinctly from each other (and rationally so). However, it's one thing to say of two agents that they're broadly in agreement with each other (or that they agree about everything that seems to be of present relevance) and another to say that they're confidants. If Ann and Bob aren't confidants, then their third-person attitudes might differ in respects which, though seeming irrelevant at first sight, could possibly explain why they don't react in the same way upon discovering the same fact (e.g. perhaps, as Lin (2022) suggests, Ann and Bill disagree about what is the best action). If that were the case, then **Bear Attack** would be a completely uninteresting case: it would just illustrate how agents who agree about some things but disagree about others might react distinctly in the same circumstances. I'm not sure how tenable that hypothesis is (more on this in 6.2), but since I don't have a principled way for demarcating which attitudes are relevant and irrelevant, I'll simply evade it by stipulating that Ann and Bill are confidants. Suppose, then, that they spotted the bear right after leaving a communication chamber wherein they communicated in the manner proprietary of confidants. Among other things, they talked about how they ought to react in various cases of bear attack, and even considered the very same scenario at which they found themselves later - having whole-heartedly agreed that, were this scenario to become actual, then whoever is under attack would need to curl up into a ball while the other would run for help. With those assumptions in place, we can be sure that Ann and Bill's reactions to the bear attack won't be explainable by their having any conflicting attitudes towards any third-person thought. Seen under this light, **Bear Attack**

[21] To be sure, there's another sense in which what they ought to do is the same, namely, change the lightbulb. More on this in 6.2.

effectively becomes an example of confidants who have distinct reasons. This concludes my case for (2).

6. *Against Asymmetry*

In the following, I look into objections to each of the Asymmetry Argument's premises.

6.1. AGAINST (1)

(1), i.e. **TWINS**, is susceptible to at least two types of criticism. First, one may argue that reasons do not supervene *solely* on attitudes. Second, one may accept the supervenience of reasons on attitudes at the intra-personal level (**CHANGES**) while denying that it must also hold inter-personally (**TWINS**).

Consider the hypothesis that reasons not only supervene on agents' attitudes but also on something else. What could that something else be? A natural candidate is what the agent is capable of doing, that is, which actions they can in principle perform. If so, then agents of the same mind might have distinct reasons due to differing in the actions which are available to each of them (Cappelen & Dever, 2013, pp. 49-56).[22] I think this proposal has already been successfully refuted (e.g. Valente, 2018; Torre, 2018; Lin, 2022), but rehearsing those arguments can still be worthwhile, for they point to a general reason to be sceptical of any structurally similar proposal.

For starters, consider confidants x and y in each other's presence at an arbitrary time, both of whom know that x has a reason to perform action α at that time, that x is capable of performing α, and that y is not so capable (perhaps due to physical limitations, lack of know-how, etc). Given the plausible assumption that no ideally rational subject can have a reason to perform an action which they judge impossible for them to perform, it follows that y does not have the same reason as x, i.e. a reason to perform α. So, a defender of **Symmetry** could suggest, the fact that x and y differ in their capacity to perform α could by itself explain why they have distinct reasons compatibly with the assumption that they have all the same attitudes (*contra* **TWINS**).

To see what is wrong with that hypothesis, imagine that, unbeknownst to both x and y, whatever impediments were the reason for y's being incapable of performing α had somehow been

[22] Interestingly, the hypothesis under consideration is reminiscent of proposals by authors who notoriously are sceptical of the significance of 'I' and irreducibly indexical thoughts (Cappelen & Dever, 2013; Magidor, 2015). It would be surprising if the shareability of first-person thoughts turned out to depend on an argument frequently associated to so-called *de se scepticism*. This would suggest that the shareability of first-person thoughts is at odds with their having distinctive essentially indexical or first-personal features.

eliminated from their case. For example, if α were the action of calling the police and only x could perform it because only his phone had service, then what we should imagine is that, unbeknownst to x and y, service had just been restored to y's phone. For another example, if α were the action of clapping one's hands and y could not do it due to being handcuffed, then we suppose that y's handcuffs had magically just been unlocked. Would these changes in y's practical capacities entail corresponding changes in y's reasons? Plausibly, they would not, for y would still wrongly but rationally believe that she cannot perform α, and we have supposed that an ideally rational agent cannot have a reason to do what she thinks is impossible for her to do. This is enough to motivate a fairly general claim, namely, that our reasons are only sensitive to facts about what we can do intermediately via being sensitive to our beliefs about what we can do. As I hope is clear enough, this is just to concede that agents' reasons do in fact supervene only on their attitudes, exactly as **TWINS** proclaims.

For an objection, one might raise the hypothesis that α is a type of action which can in principle only be performed by x - perhaps because the logical form of that action is something like *the action that x performs a*. I'm not sure if this proposal is as much as coherent in and of itself; in any case, it certainly should not get sympathies from a defender of **Symmetry**. For one thing, it effectively appears to invite a notion of privacy to the realm of actions in a way that dangerously resembles how a proponent of **Privacy** applies that notion to a subset of thoughts. Indeed, both the hypothesis under consideration and **Privacy** share the assumption that a fundamental element of propositional attitude psychology and inferential reasoning is private, disagreeing only about whether this element pertains more closely to the reasoning's premises (thoughts) or to their outcomes (actions). Second, even if we granted the assumption that α is an action performable only by x, the same reasoning as in the previous paragraph could be used to show that y could have a reason to perform α if she wrongly but rationally believed that she is x. Just like agents' reasons don't seem to be sensitive to brute facts about what they can do (as opposed to their attitudes about what they can do), they also don't seem to be sensitive to brute facts about their identity (as opposed to their attitudes about who they are). More generally, for any candidate type of fact on which agents' reasons might additionally supervene, it seems possible to tweak the relevant agents' attitudes about these facts in such a way as to show that, ultimately, their reasons exclusively supervene on their attitudes.

For an alternative, and perhaps more radical, challenge against **TWINS**, one could argue that the supervenience of reasons on attitudes holds only for a single agent, not for multiple ones (Hunter 2017; Hunter 2022, pp. 157-9). In our terminology, the hypothesis is that we can adequately accommodate what Bermúdez calls a basic tenet of propositional attitude psychology by accepting

CHANGES while rejecting **TWINS**. I believe that the structural analogy between **TWINS** and **CHANGES** is enough to show that this "mixed" strategy is untenable: why would time-slices of a single agent be different from multiple agents with regards to the independence between their reasons and their attitudes? In any case, I'll advance a more general challenge to anyone aiming to save **Symmetry** by means of this strategy. In a nutshell, my point is that acceptance of **CHANGES** can by itself give rise to an argument for a type of thought unshareability that in many ways resembles the workings of the Asymmetry Argument. With one exception, however: the new argument concerns temporally indexical thoughts, not first-personal ones.

To recap, **CHANGES** says that one's reasons cannot change if one's attitudes remain the same. Well, either it's possible for all of an agent's attitudes to remain the same from a time to another or it's not. If it's not, then it's because there are attitudes which by their very nature cannot be retained across time, a conclusion which will invariably motivate the introduction of thoughts that are *temporally private*, i.e. apprehensible only at some times. Presumably, this wouldn't sit well with any defence of **Symmetry**, so, instead suppose that an agent's later self could in principle have retained all of her earlier self's attitudes. A good candidate example would be an idealised agent with flawless memory who does nothing from a moment to another but track the passage of time in a wholly predictable way (think of someone watching their favourite movie for the hundredth time and thinking in succession, e.g. 'the bad guy will soon be killed', 'the bad guy is being killed now', 'the bad guy has just been killed'; Titelbaum 2013, p. 232). Now, notice how the intimate relationship between this agent's time-slices' attitudes resembles that between the attitudes of confidants: intuitively, they're just the same attitudes *modulo* a shift in their indexically-encoded perspectives - e.g. across time, 'soon' gets switched for 'now', across agents, 'you' gets switched for 'I', etc. Finally, notice that a single agent's reasons can change even when her attitudes shift only in this minimal sense. For example, suppose that the movie-watcher has a ritual where she always claps when the bad guys falls dead on the floor. She would then have a reason to clap when she thinks 'the bad guy has fallen dead [now]' but not when she earlier thought 'the bad guy will soon fall dead'. Absent reasons to think that her attitudes have changed in any other way over and above being temporally shifted - e.g. the switch from thinking of the bad guy's death as present instead of future -, the change in her reasons will ultimately require distinguishing between temporally indexical attitudes expressed with distinct indexicals even when coordinated via flawless tracking to the passage of time.

The argument above can be framed in a way that highlights its structural similarity to the Asymmetry Argument. Let **Synchrony** stand to temporally indexical thoughts as **Symmetry**

stands to first-person ones (i.e. both principles stating that the corresponding type of indexical thoughts are shareable across times/time-slices or agents). The Asynchrony Argument then is:

> (1*) **CHANGES** [an ideally rational agent's reasons cannot change if their attitudes remain the same].
>
> (2*) An agents' reasons can change even when she does nothing but perfectly keep track of the passage of time in a wholly predictable way.
>
> (3*) If **Synchrony** holds, then the attitudes of an agent who does nothing but perfectly keep track of the passage of time in a wholly predictable way remain all the same.
>
> Therefore, **Synchrony** does not hold (from 1*, 2*, and 3*)

I hope it is easy to see that the Asynchrony Argument supports the (temporal) privacy of temporally indexical thoughts in virtually the same way as the Asymmetry Argument supports the privacy of first-person ones. However, it's hard to see how one could have a good rationale for defending **Symmetry** independently of a more general ambition to vindicate the shareability of thoughts as a whole. If that's right, then there's no promise for a mixed strategy which minimises the damage of rejecting **TWINS** on the basis of accepting **CHANGES**. Either both principles must be taken on board - in which case one ought to challenge the arguments 'other premises - or both must be rejected - contrary to a basic tenet of propositional attitude psychology, namely, that reasons supervene on attitudes both for a single agent and for multiple ones.

6.2. Against (2)

According to (2), some confidants have distinct reasons. As we've seen, (2) can be supported by general considerations having to do with cooperative action as well as by looking into particular cases like **Bear Attack**. Given the simplicity of these arguments, it strikes me as the most difficult premise to deny. One way to do so is to question our intuitive verdicts about sameness of action and, as a consequence, about sameness of reasons.

No doubt there's a sense in which Ann and Bob do *not* act alike in **Bear Attack**. However, there might also be a plausible sense in which they do so, after all, both act in light of a common goal - e.g. protecting Ann from the bear attack. If asked to explain why they acted as they did, for example, both could give similar answers: 'to protect Ann/her/myself'. So, who is it to say that the concept of action and reasons relevant for propositional attitude psychology is one which, applied to **Bear Attack**, entails that Ann and Bob act distinctly as opposed to alike, or have distinct reasons as opposed to the same ones?

Bermúdez (pp. 48-9, pp. 105-6) makes a few observations that point towards the reasoning outlined above. The context is a critical discussion of Perry (1977), who Bermúdez charges of confusing between intentional actions and the bodily movements via which they are implemented:

> Perry is assuming that if our bodies move in comparable ways then we are behaving in the same way [...] But psychological explanations explain actions rather than bodily movements, and actions are (at least partially) individuated by their goal. From this perspective it is not obvious that my rolling up in a ball is the same action as your rolling up in a ball. (Bermúdez, 2017, p. 48-9)

One of Bermúdez's point which strikes me as clearly correct is that the concept of intentional action relevant to propositional attitude psychology admits of distinct action-types having tokens which look just like one another, as plausibly the case when the relevant tokens have radically distinct goals (e.g. an agent walking home might look just like another performing a contemporary dance piece). However, Bermúdez's contention that actions ought (at least partially) to be individuated by their goal suggests more than that, namely, that action-tokens which are radically unalike can be grouped under the same type when they have the same goal. This more general point is echoed in further remarks by Bermúdez elsewhere in his book:

> What I am doing is saving myself from the bear. This action-type can have other tokens, such as, for example, my shooting the bear. (p. 57)

> The real issue [...] is not who is performing the action, but what the goal of the action is. What matters for the thought that *I myself am being pursued by a bear* is that it should give rise to actions with the aim of extricating *me* from my predicament. Those might be actions that I perform—standing my ground and making noise (recommended), or running away (not recommended). Or they might be actions that someone else could perform—distracting the bear or attacking it with bear spray. (Bermúdez, 2017, p. 105)

On one way of reading the above, Bermúdez's suggestion is that, for the purposes of propositional attitude psychology, we neither need to differentiate the actions performed by Ann and Bob nor their underlying motivating reasons. Regardless of the oddity of saying that Ann's curling up into a ball and Bob's running for help are tokens of the same action-type, I don't think this is an absurd claim. Still, absurd or not, it cannot suffice to motivate the rejection of (2).

A general hypothesis suggested by Bermúdez's remarks is that confidants need not be attributed distinct reasons because, though there might be superficial differences in how they act, their acting in light of the same goal entails that their actions (and reasons) can appropriately be grouped under the same type. An initial question is: how should goals be individuated (and why should that be easier than individuating actions/reasons)? Surely, if one is creative enough, one can readily find a general enough action-type under which to group any two arbitrary action-tokens that have little

to do with each other. And surely, in the absence of a more precise theory of goal individuation, one's creativity will allow one to readily do the same for agents' goals as well. One then wonders whether shifting from action/reason-talk to goals-talk will be much of an improvement.

Another question is whether propositional attitude psychology sits well with the grouping together of action-tokens whenever they have the same goal. One obvious issue with that idea is that it would restrict the scope of psychological generalisations in a dramatic way, entailing that whatever differences there are between, for example, Ann and Bob's reactions to the approaching bear, cannot be given an explanation in attitudinal terms. In other words, Ann and Bob's attitudes would at most allow us to explain why their goal is to protect Ann whilst being completely silent about why Ann ought to do so by curling up into a ball and Bob, by running for help. This conflicts with the general idea that, in practical reasoning, an agent's attitudes stand to their actions just like the premises of a theoretical inference stand to its conclusion.[23] To spin the same point in yet another way: if the practical implications of being in Ann's attitudinal state were the same as that of being in Bob's, then their attitudinal states would underdetermine whether they ought to curl up into a ball or to run for help. How, then, could Ann rationally explain why she curled up into a ball as opposed to running for help if even a full disclosure of her attitudinal state wouldn't justify anything more specific than that she ought act with the goal of protecting Ann from the bear?

To be sure, all the above is compatible with Bermúdez's fair remark that any sufficiently interesting concept of intentional action must be compatible with there being permissible variations on the manner in which it can be performed. Perhaps we wouldn't need to say that Ann acted distinctly if, instead of curling up into a ball head first, she started by curling up her legs. However, this weaker claim is surely not enough to entail anything as strong as the type-identity between Ann and Bob's actions. I'm also tempted to agree with Bermúdez that no two action-tokens, however similar to each other, ought to be grouped under the same type if they uncontroversially have distinct goals (at least not when our primary concern is attitudinal explanations of action; cf. Evans, 1982, p. 203-4). But this is compatible with there sometimes being a need to distinguish between action-tokens which, intuitively, have the same goal, and so, to attribute distinct reasons to agents even when their actions are perfectly orchestrated as part of a joint plan. If so, then (2) remains in good-standing.

[23] As I read him, Bermúdez (p. 47) explicitly endorses this general idea and goes as far as deploying it in an argument against Perry.

6.3. AGAINST (3)

Since (3) involves a particular interpretation of what the shareability of first-person thoughts entails, it looks like the most promising target for proponents of **Symmetry**. To recap, (3) holds that, if **Symmetry** is true, then confidants have all the same attitudes. I take (3) to follow from the following four premises:

(i) Two agents have all the same attitudes if and only if they apprehend all the same thoughts and adopt the same attitude-types towards them

(ii) If **Symmetry** is true, then confidants apprehend all the same *first-person* thoughts and adopt the same attitude-types towards them

(iii) Confidants apprehend all the same *third-person* thoughts and adopt the same attitude-types towards them

(iv) Attitudes are either third-person or first-person

Therefore, if **Symmetry** is true, then confidants have all the same attitudes.

Given the above, a defender of **Symmetry** is advised to challenge at least one of the four premises supporting (3). For ease of exposition, let's focus on those two attitudes of Ann and Bob which most see as the most relevant in **Bear Attack**:

(AA): Ann's attitude expressible as 'I am about to be attacked by a bear'

(AB): Bob's attitude expressible as 'you are about to be attacked by a bear'[24]

(i) can be challenged by arguing that there's more to sharing an attitude than adopting the same attitude-type towards the same thought. For example, attitudes like belief might be triples involving the attitude-type of belief, a thought, and some third-ingredient under which the thought is believed (e.g. a mode of presentation, perspective, guise, etc).[25] If so, then AA and AB might be equal with respect to the first two elements while differing in their third one. Alternatively, one might accept that attitudes are pairs of attitude-types and thoughts but deny (ii) by holding that coarse-grained attitude-types such as belief and desire are too simplistic, and so, must give way to fine-grained attitude-types which correspond to different ways of believing and desiring the same

[24] Though AA and AB are doxastic attitudes, my reasoning below could have equally been framed in terms of non-doxastic indexical attitudes like desire.

[25] The view under consideration is reminiscent of Perry's (1977) and has been recently developed by Prosser (forthcoming). Since Perry is one of Bermúdez's (p. 46-50) main targets of criticism, it's fair to assume that he's not the least sympathetic to it.

thought.[26] For example, AA and AB might differ because the former involves a first-person type of belief but not the latter. Clearly, both strategies are structurally analogous in an important sense: they attempt to save **Symmetry** by holding that AA and AB are different attitudes regardless of involving the same thought. For this very reason, both strategies suffer from the same problem with respect to their compatibility with **Symmetry**.

Suppose AA and AB are different attitudes due to the denial of either (i) or (ii). Faced with this objection, a proponent of **Privacy** might ask: but are attitudes shareable or private on this new view under consideration? Surely, a defender of **Symmetry** must hold that they're shareable, for any view where some attitudes can in principle only be held by some agents is anathema to the spirit of thought shareability. But if they are, then it must in principle be possible for Bob to have AA in **Bear Attack** (and Ann to have AB). By assumption, Bob doesn't have AA. Furthermore, he's maximally rational. So, the reason why Bob doesn't have AA is because it would have been irrational for him to do so. If that's the case, then rationality appears to require Ann to have AA while forbidding Bob to do so. Given that they're confidants, this can only be because the rationality of holding an attitude towards a thought is an agent-relative matter which can vary across agents even when all else remains the same. In other words, the proposals are committed to:

> **Relativism**: though confidants could in principle have all the same attitudes towards the same thoughts, it could be irrational for them to do so.

Whatever the merits of this principle, I'm confident that **Relativism** undercuts the fundamental motivation for anyone aiming to defend **Symmetry** in the first place, and so, has no place in a view which emphasises the shareability of thoughts. To see why, suppose that **Relativism** is true, and so, that Ann and Bob apprehend the same indexical thoughts as each other regardless of being required to do it in distinct ways. Now, consider: why would the shareability of thoughts allowed by that type of view amount to a vindication of **Symmetry** over **Privacy** as opposed to a pyrrhic victory?

My worry is that a proponent of **Privacy** would be justified in taking the conjunction of **Symmetry** and **Relativism** (henceforth, **Symmetry-Relativism**) as, at best, a close cousin to their own view, at worse, a notational variant thereof. For one thing, defenders of **Privacy** and of **Symmetry-Relativism** agree that Ann and Bob cannot rationally share AA and AB - be this because AA and AB involve private thoughts or because they involve thoughts which, though

[26] A similar idea is elaborated for temporally indexical attitudes by Shaw (2019).

shareable, cannot rationally be apprehended in the same way by anyone. If they disagree at all, they do so only with respect to whether AA and AB involve the same thought. But it's not clear whether this purported disagreement amount to anything substantial. Having rejected the standard view according to which attitudes are pairs of coarse-grained attitude-types and thoughts, the defender of **Symmetry-Relativism** doesn't appear to mean anything by, 'AA and AB involve the same thought', over and above the claim that Ann's holding AA and Bob's holding AB involves them apprehending thoughts which could exemplify a successful instance of knowledge transmission via testimony. However, as we've seen in Section 3, this is a claim that a proponent of **Privacy** will be more than happy to embrace. A defender of **Symmetry-Relativism** might retort that her view fares better due to eschewing any hint of privacy in propositional attitude psychology. But her opponent could simply point out that **Relativism** does introduce a type of privacy, after all, it attributes a sort of agent-relativity to the rationality of holding an attitude which obviously resembles the asymmetry in the conditions of thoughts' apprehension countenanced by **Privacy**. Against this, proponents of **Symmetry-Relativism** might insist that they're the only ones able to maintain that attitudes expressed with 'you' and 'I' like AA and AB might literally be said to be attitudes towards *the same thought*. However, they can do so only by accepting that this true shareable thought is not one which Ann and Bob can be said to believe in the same way. Surely, the proponent of **Privacy** will exclaim, "ways of believing a thought" are neither less nor more mysterious theoretical posits than their favoured private thoughts. Furthermore, **Symmetry-Relativism** is committed to there being attitudes which, though accessible to anyone in principle, are not *rationally* accessible to everyone. Indeed, the proponent of **Symmetry-Relativism**'s claim that Ann and Bob could in principle share AA and AB - though not rationally - reminds me of W. W. Jacobs' horror story The Monkey's Paw, where a man who asks a demon for £200 is granted his wish only through receiving the insurance compensation for the sudden death of his loved one. Shareable, yes, but at what cost?

Finally, one who defends **Symmetry-Relativism** will still be hard-pressed to explain how the rationality of holding an attitude could be agent-relative in that sense. If AA is a doxastic attitude towards a true thought which Ann would be irrational not to hold, then, presumably, Ann's holding AA amounts to her being in possession of a piece of knowledge. But since Bob would be irrational for holding AA, then Bob's having AA could not amount to his being in possession of a piece of knowledge - for how could rationality forbid Bob from having knowledge? This is indication that **Symmetry-Relativism** will ultimately turn out to be committed to a particularly nasty form of epistemic relativism wherein some pieces of knowledge are in principle accessible only to particular agents. How can any similar conclusion be an acceptable corollary of **Symmetry**, a view

predicated on the need to vindicate the objectivity and shareability of thoughts across multiple perspectives?

In summary, by rejecting (i) or (ii), the proponent of **Symmetry** will have to buy onto **Relativism**. As a consequence, they will ultimately have to accept that propositional attitude psychology and knowledge are infused with a degree of agent-relativity that at the very least is as worrisome as any implications of postulating private thoughts. I take all this to suggest that **Symmetry-Relativism** and **Privacy**'s purported disagreement on the shareability of first-person thoughts is less substantial than it might initially appear. If so, then the denial of (i) and (ii) is less a way to save **Symmetry** than of conceding that distinct agents cannot in principle adopt the same attitudes towards the same thoughts, as proponents of **Privacy** have been defending all along, though not in these exact terms.

Let us now consider (iii) and (iv). To deny (iii), one must hold that confidants can differ in their third-person attitudes. On a weaker reading, the claim is that confidants are rationally permitted to hold conflicting attitudes towards some of the same third-person thoughts. On the stronger reading, they not only have permission to disagree but are required to do so. I take the stronger reading to be strongly implausible. Even those with highly permissive views of rationality are careful in saying that, due to evidential underdetermination or related phenomena, a single body of evidence might *permit* more than a single attitude towards the same thought. However, if confidants are permitted (but not required to) harbour third-person disagreements, then they are equally permitted to agree about these same matters. Thus, we need only stipulate that, through rational deliberation, confidants like Ann and Bob have decided to adopt a fully conciliatory approach to any matters about which rationality is permissive, and so, have preempted the possibility of their ever making different choices in cases where rationality allowed them multiple options. In any case, it's just immediately implausible that Ann and Bob's distinct actions in **Bear Attack** are the product of their holding permissibly distinct attitudes towards a single third-person thought; the fact that they ought to act differently from each other is plausibly not a product of a decision which they could rationally have not taken - it's just an inevitable consequence of their need to coordinate their actions in light of their shared goal. So, there's no reason to think that permissible disagreements might underly Ann and Bob's distinct actions.

Finally, one might deny (iv) by holding that some attitudes are neither third-person nor first-person. This would allow one to hold that Ann and Bob don't have the same attitudes regardless of having the same third-person and first-person ones. But which other attitudes could those be? One hypothesis is that they differ in their demonstrative attitudes, those naturally expressible using

'this' or 'that'. However, we may simply stipulate that Ann and Bob are acquainted with all the same objects, and so, able to demonstratively pick out the same things as one another. Of course, a defender of **Symmetry** won't find solace in that view once defended by Bertrand Russell whereby we can only be directly acquainted with ourselves and the inner objects of our own consciousness. This view would surely entail that Ann and Bob don't have the same demonstrative attitudes, but only in detriment to the shareability of demonstrative thoughts.

Alternatively, one might raise the hypothesis that the relevant difference between Ann and Bob lies in their being in distinct perceptual states. For example, Ann might be seeing the approaching bear's face while Bob can only see its back. Surely, this cannot be the whole explanation of why they act differently from each other. In any case, we might circumvent this complication by tweaking **Bear Attack** in such a way as to minimise any perceptual discrepancies between the agents (e.g. picture Ann and Bob in a sensory deprivation chamber, being telepathically informed that something horrible will happen unless Ann mentally counts from 1 to 20, and Bob mentally counts from 20 to 1). Additionally, one might take their relevant difference to be a matter of their emotions or affective states, after all, isn't Ann bound to be much more scared of the approaching bear than Bob is? That might be so, but cannot we easily come up with a boring case involving confidants who ought to act differently due only to their sense of duty, and so, who comply with their reasons in the utmost blasé manner?[27] There surely are other hypotheses I could consider - e.g. their values, their preferences, their habits, their tastes - but I take the above to sufficiently justify the conclusion which from the outset struck me as the most intuitive one, namely, that Ann and Bob act differently from each other because of asymmetries in their attitudes expressible with 'you' and 'I', those very asymmetries which a proponent of **Privacy** is well-positioned to accommodate by taking some of them as targeting private thoughts which can only be apprehended by their very owners.

7. Conclusion

I have claimed that the shareability of first-person thoughts doesn't sit well with the rational asymmetries between 'you' and 'I' showcased by the Asymmetry Argument. Given that confidants, agents whose attitudes agree with one another's to the maximal degree, can have distinct motivating reasons for action, it seems that we have no better way to explain how that is possible without accepting that thoughts and attitudes are subject to some degree of privacy and

[27] The irrelevance of perceptual and emotional discrepancies in **Bear Attack** has been, correctly in my view, pointed out by Gray (2022).

unshareability. In that context, **Privacy** emerges as a natural candidate to account for the relevant data: since first-person thoughts are private, then even confidants will differ in which first-person attitudes they hold. I looked into multiple ways for trying to make **Symmetry** compatible with these findings, but none turned out to be ultimately tenable. Some saddled **Symmetry** with the rejection of a basic tenet of propositional attitude psychology (6.1), some with an implausible picture of how actions and reasons ought to be individuated (6.2), and others with an unstable commitment to **Relativism** which appears anathema to the idea that 'you' and 'I' can underly the holding of the same attitude (6.3). I might have failed to identify a further sense in which the shareability of first-person thoughts would both be philosophically significant and incompatible with **Privacy**, but I expect to have said enough to show that the burden of proof lies on defenders of **Symmetry**'s side. For the time being, my conclusion is that either we must accept that first-person thoughts are private or accept that they are only are shareable in a weak sense which any defender of **Privacy** could be happy to embrace. If I'm right, then the thesis that first-person thoughts are shareable is either false or uninteresting.[28]

Acknowledgements

I would like to express my sincere gratitude to Víctor M. Verdejo, Valen Simpson, Quentin Ruyant, and two anonymous referees for invaluable feedback on this paper. I'm also thankful to audiences at the Valencia Philosophy Lab, LOGOS, Indexical Dynamics colloquium (Collège de France), Language & Mind group (Arché, St Andrews), University of Aberdeen, and 6th PLM Workshop (University of Vienna). The research leading to this paper has been funded by: FCT Researcher Fellowship (University of Lisbon, Center of Philosophy, LanCog), Margarita Salas fellowship (Universitat de Barcelona), and the projects Philosophy of Hybrid Representations (PID2020-119588GB-I00) and Eliminativism, Fictionalism and Expressivism (PID2019-106420GA-I00).

[28] One interesting avenue of research I haven't explored concerns the implications of Bermúdez's commitment to **Essential Indexicality** (see footnote 4) and **Symmetry**. *Prima facie*, **Essential Indexicality** draws a sharp divide between attitudes expressed with 'I' from all others, including those which, according to **Symmetry**, are attitudes to the same thought. The dissonance between **Essential Indexicality** and **Symmetry** might mean one of two things: either it shows the existence of a deep internal conflict in Bermúdez's account, or it hides the key to dissolving the arguments I presented in this paper. I wouldn't be surprised if the second option turns out to be the correct one.

REFERENCES

Alvarez, M. & Way, J. (2024). Reasons for action: Justification, motivation, explanation. *The Stanford Encyclopedia of Philosophy* (Fall 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2024/entries/reasons-just-vs-expl/>.

Bermúdez, J. L. (2005). Evans and the sense of "I". In J. L. Bermúdez (Ed.). *Thought, reference, and experience: Themes from the philosophy of Gareth Evans* (pp. 164-94). Oxford University Press.

Bermúdez, J.L. (2011). Self-knowledge and the sense of "I". In A. Hatzimoysis (Ed.). *Self-knowledge* (pp. 226-45). Oxford University Press.

Bermúdez, J. L. (2017a). *Understanding "I": Language and thought*. Oxford University Press.

Bermúdez, J. L. (2017b). Yes, essential indexicals really are essential. *Analysis,* 77(4), 690-694.

Bermúdez, J. L. (2019). First person thoughts: Shareability and symmetry. *Grazer Philosophische Studien,* 96(4), 629-638.

Bozickovic, V. (2021). *The indexical point of view: On cognitive significance and cognitive dynamics*. Routledge.

Byrne, A. & Thau, M. (1996). In defence of the Hybrid View. *Mind*, 105(417), 139-149.

Cappelen, H. & Dever, J. (2013). *The inessential indexical: On the philosophical insignificance of perspective and the first person*. Oxford University Press.

Caie, M. (2018). Agreement and updating for self-locating belief. *Journal of Philosophical Logic*, 47(3), 513-547.

Cumming, S. (2013). From coordination to content. *Philosophers' Imprint*, 13(4).

Dummett, M. (1980). *The interpretation of Frege's philosophy*. Harvard University Press.

Evans, G. (1981). Understanding demonstratives. In H. Parret (Ed.). *Meaning and understanding* (pp. 280–304). Clarendon Press.

Evans, G. (1982). *The varieties of reference*. Oxford University Press.

Fine, K. (2007). *Semantic relationism*. Blackwell.

Frege, G. (1956). The thought: A logical inquiry. *Mind,* 65(259), 289-311.

García-Carpintero, M. (2017). The philosophical significance of the de se. *Inquiry*, 60(3), 253-276.

Goodman, R. (forthcoming). Shared thought and communication. In J. Bermúdez, M. Valente & V. M. Verdejo (Eds.). *Sharing thoughts: Philosophical perspectives on intersubjectivity and communication*. Oxford University Press.

Gray, A. (2022). Minimal fregeanism. *Mind*, 131(522), 429-458.

Heck, R. K. (2002). Do demonstratives have senses?. *Philosophers' Imprint*, 2(2), 1-33.

Howard, N. (2024). Convergence and the agent's point of view. *Belgrade Philosophical Annual*, 37(1), 145-165.

Hunter, D. (2017). Practical reasoning and the first person. *Philosophia*, 45(2), 677-700.

Hunter, D. (2022). *On believing: Being right in a world of possibilities*. Oxford University Press.

Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review,* 88(4), 513-543.

Lin, L. (2022). Attitudes and action: Against de se exceptionalism. *Inquiry*. Advance online publication: https://doi.org/10.1080/0020174X.2022.2158126

Loar, B. (1976). The semantics of singular terms. *Philosophical Studies*, 30(6), 353-377.

Longworth, G. (2013). IV—Sharing thoughts about oneself. *Proceedings of the Aristotelian Society*, 113(1pt1), 57-81.

Longworth, G. (2014). You and me. *Philosophical Explorations*, 17(3), 289-303.

Magidor, O. (2015). The myth of the de se. *Philosophical Perspectives*, 29(1), 249-283.

McDowell, J. (1984). De re senses. *Philosophical Quarterly*, 34(136), 283-294.

Morgan, D. (2009). Can you think my 'I'-thoughts?. *Philosophical Quarterly,* 59(234), 68-85.

Morgan, D. (2024). Islands of perspectival thought: A case study. *Ergo: An Open Access Journal of Philosophy*, 10: 44.

Ninan, D. (2016). What is the problem of de se attitudes?. In S. Torre & M. García-Carpintero (Eds.). *About oneself: De se thought and communication* (pp. 86-120). Oxford University Press.

Ninan, D. (2021). De se attitudes and action. In H. Geirsson & S. Biggs (Eds.). *The Routledge handbook of linguistic reference* (pp. 482-498). Routledge.

Onofri, A. (2019). Loar's puzzle, similarity, and knowledge of reference. *Manuscrito*, 42(2), 1-45.

Perry, J. (1977). Frege on demonstratives. *Philosophical Review*, 86(4), 474-497.

Perry, J. (1979). The problem of the essential indexical. *Noûs,* 13, 3-21.

Prosser, S. (2005). Cognitive dynamics and indexicals. *Mind and Language*, 20(4), 369-391.

Prosser, S. (2023). Tense and emotion. In K. M. Jaszczolt (Ed.). *Understanding human time* (pp. 11-29). Oxford University Press.

Prosser, S. (forthcoming). Sharing egocentric thoughts. In J. Bermúdez, M. Valente & V. M. Verdejo (Eds.). *Sharing thoughts: Philosophical perspectives on intersubjectivity and communication*. Oxford University Press.

Recanati, F. (2016). *Mental files in flux*. Oxford University Press.

Recanati, F. (forthcoming). Shared modes of presentation. In J. Bermúdez, M. Valente & V. M. Verdejo (Eds.). *Sharing thoughts: Philosophical perspectives on intersubjectivity and communication*. Oxford University Press.

Shaw, J. R. (2019). De se exceptionalism and frege puzzles. *Ergo: An Open Access Journal of Philosophy*, 6, 1057-1086.

Singh, K. (2019). Acting and believing under the guise of normative reasons. *Philosophy and Phenomenological Research*, 99(2), 409-430.

Stanley, J. (2011). *Know how*. Oxford University Press.

Stalnaker, R. (2008). *Our knowledge of the internal world*. Oxford University Press.

Titelbaum, M. G. (2013). *Quitting certainties*. Oxford University Press.

Torre, S. (2018). In defense of de se content. *Philosophy and Phenomenological Research*, 97(1), 172-189.

Torre, S. & Weber, C. (2021). What is special about de se attitudes?. In H. Geirsson & S. Biggs (Eds.). *The Routledge handbook of linguistic reference* (pp. 464-481). Routledge.

Torre, S. & Weber, C. (2022). De se puzzles and frege puzzles. *Inquiry*, 65(1), 50-76.

Valente, M. (2018). What is special about indexical attitudes?. *Inquiry,* 61(7), 692-712.

Valente, M. (2021). On successful communication, intentions and false beliefs. *Theoria*, 87(1), 167-186.

Valente, M. & Onofri, A. (2023). A puzzle about communication. *Review of Philosophy and Psychology*, 14(3), 1035-1054.

Verdejo, V. M. (2017). De se content and action generalisation. *Philosophical Papers,* 46(2), 315-344.

Verdejo, V. M. (2019). The second person perspective. *Erkenntnis*, 86(6), 1693-1711.

Verdejo, V. M. (2020). Rip Van Winkle and the retention of 'today'-belief: A puzzle. *Res Philosophica*, 97(3), 459-469.

Verdejo, V. M. (2025). Thoughts about oneself to share in context: Meeting Bermúdez's challenge. *THEORIA. An International Journal for Theory, History and Foundations of Science*.

Wang, L. (forthcoming). Taking motivating reasons 'deliberative role seriously. *Philosophical Studies*.

Weber, C. (2014). Indexical beliefs and communication: Against Stalnaker on self-location. *Philosophy and Phenomenological Research*, 90(3), 640-663.

MATHEUS VALENTE is a postdoctoral researcher at LanCog (Centre of Philosophy, University of Lisbon) who specializes in the philosophy of mind, language, and epistemology. He has published about indexical and self-locating attitudes, communication, reference, and demonstrative thoughts.

ADDRESS: Centre of Philosophy, University of Lisbon, Lisboa, Alameda da Universidade, 1600-214. Email: matheusvalenteleite@gmail.com – ORCID: 0000-0001-6380-2623