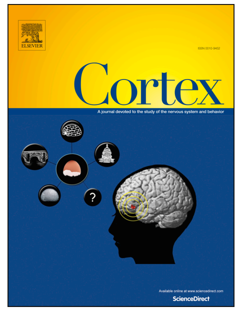
Journal Pre-proof



From action to abstraction: The sensorimotor grounding of metaphor in Parkinson's disease

Stacey Humphries, Nathaniel Klooster, Eileen Cardillo, Daniel Weintraub, Jacqueline Rick, Anjan Chatterjee

PII: S0010-9452(19)30323-5

DOI: https://doi.org/10.1016/j.cortex.2019.09.005

Reference: CORTEX 2727

To appear in: *Cortex*

Received Date: 2 August 2019

Accepted Date: 6 September 2019

Please cite this article as: Humphries S, Klooster N, Cardillo E, Weintraub D, Rick J, Chatterjee A, From action to abstraction: The sensorimotor grounding of metaphor in Parkinson's disease, *CORTEX*, https://doi.org/10.1016/j.cortex.2019.09.005.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Elsevier Ltd. All rights reserved.

Journal Pre-proof

From action to abstraction: The sensorimotor grounding of metaphor in Parkinson's disease

Stacey Humphries[1], Nathaniel Klooster[1], Eileen Cardillo[1], Daniel Weintraub[1], Jacqueline Rick[1], Anjan Chatterjee[1]

[1]University of Pennsylvania,
Philadelphia, Pennsylvania, 19104
USA

(corresponding author) Stacey Humphries: hstace@pennmedicine.upenn.edu

Nathaniel Klooster: klooster@pennmedicine.upenn.edu

Eileen Cardillo: eica@pennmedicine.upenn.edu

Daniel Weintraub: daniel.weintraub@uphs.upenn.edu

Jacqueline Rick: jacqui.rick@pennmedicine.upenn.edu

Anjan Chatterjee: anjan@pennmedicine.upenn.edu

Journal Pre-proof

Journal Pre-proof

From action to abstraction: The sensorimotor grounding of metaphor in Parkinson's disease

Stage 2 Registered Report
In Principle Acceptance granted: 8th September 2017 (CORTEX-D-17-00536)
Stage 1 version: https://osf.io/cwsg7/

Abstract

Embodied cognition theories propose that the semantic representations engaged in during language comprehension are partly supported by perceptual and motor systems, via simulation. Activation in modality-specific regions of cortex is associated with the comprehension of literal language that describes the analogous modalities, but studies addressing the grounding of non-literal or figurative language, such as metaphors, have yielded mixed results. Differences in the psycholinguistic characteristics of sentence stimuli across studies have likely contributed to this lack of consensus. Furthermore, previous studies have been largely correlational, whilst patient studies are a critical way of determining if intact sensorimotor function is necessary to understand language drawing on sensorimotor information. We designed a battery of metaphorical and literal sentence stimuli using action and sound words, with an unprecedented level of control over critical psycholinguistic variables, to test hypotheses about the grounding of metaphorical language. In this Registered Report, we assessed the comprehension of these sentences in 41 patients with Parkinson's disease, who were predicted to be disproportionately affected by the action sentences relative to the sound sentences, and compared their performance to that of 39 healthy age-matched controls who were predicted to show no difference in performance due to sensory modality. Using preregistered Bayesian model comparison methods, we found that PD patients' comprehension of literal action sentences was not impaired, while there was some evidence for a slowing of responses to action metaphors. Follow up exploratory analyses suggest that this response time modality effect was driven by one type of metaphor (predicate) and was absent in another (nominal), despite the fact that the action semantics were similar in both syntactic forms. These results suggest that the conditions under which PD patients demonstrate hypothesized embodiment effects are limited. We offer a critical assessment of the PD action language literature and discuss implications for the embodiment debate. In addition, we suggest how future studies could leverage Bayesian statistical methods to provide more convincing evidence for or against embodied cognition effects.

Keywords: Embodied cognition; Parkinson's disease; metaphor; simulation; abstraction.

1. Introduction

A central debate in cognitive science concerns the representational nature and organization of concepts in the brain. One view is rooted in computational approaches to cognition (e.g. Newell & Simon, 1976) and treats concepts as abstract, amodal symbols, divorced from the processes underlying their perception and acquisition. In contrast, embodied views of cognition (e.g.

Journal Pre-proof

Barsalou, 1999) propose that concepts are grounded in the same perceptual, motor and emotional processes involved in real-world experiences, via simulation. A major sticking point in the debate between these contrasting views is how embodied cognition can account for comprehension of abstract ideas (Chatterjee, 2010; Dove, 2009). One linguistic vehicle through which we understand the abstract is metaphor (Jamrozik, Mcquire, Cardillo, & Chatterjee, 2016), providing an ideal test case for theories of embodied cognition. Whilst considerable work has been devoted to clarifying whether or not we understand literal language about actions and sensory experiences in an embodied way (i.e., through a process of simulation), it remains unclear whether we understand metaphors in a similar fashion, by simulating the literal sensorimotor features of their components. Evidence that metaphors are processed in an embodied way, despite their abstract meanings, would provide strong support for theories of embodied cognition. To address this question, we tested the hypothesis that disruption to the motor system (in patients with Parkinson's disease) would impairthe comprehension of novel action metaphors.

The role of simulation is at the forefront of embodied and grounded theories of cognition (Barsalou, 1999, 2008). Theories of embodied semantics propose that cognitively representing a concept or word involves a reactivation of the same neural processes involved in the physical experience of the object or event denoted by that word. Support has come from neuroimaging studies, such as those using fMRI to demonstrate that reading action words and sentences associated with different effectors (the hand, foot and mouth) activates the same somatotopically organized motor regions of the brain that are activated during actual movement of those body parts (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Desai, Binder, Conant, & Seidenberg, 2009; O Hauk, Johnsrude, & Pulvermüller, 2004; Watson, Cardillo, Bromberger, & Chatterjee, 2014; Watson, Cardillo, Ianni, & Chatterjee, 2013). Behavioral studies also provide evidence for an interaction between the processing of action language and compatible or incompatible motor responses. For example, when asked to judge the sensibility of sentences by responding with a movement towards or away from the body, participants are faster to respond when the movement they make matches the direction of the movement implied by the sentence (e.g. "close the drawer" implies a movement away from the body); a phenomenon described as a the action-sentence compatibility effect (ACE) (Glenberg & Kaschak, 2002). The ACE has further been found to extend to specific actions such as rotations (Zwaan & Taylor, 2006), and to be sensitive to actions involving different effectors such as the hand, foot and mouth (Scorolli & Borghi, 2007).

Studies involving patients with damage to motor networks provide further compelling evidence for the involvement of motor systems in action-language comprehension. Patients with neurodegenerative conditions such as motor neuron disease (MND) (Bak & Hodges, 2004), Huntington's disease (García et al., 2017; Kargieman et al., 2014) and parkinsonian syndromes (Cardona et al., 2013) have demonstrated a specific deficit in action verb processing (Bak, 2013). Parkinson's disease (PD) patients are the most widely studied patient population in action semantics over the last decade, likely due to the prevalence of the condition and the generally better health of PD patients relative to MND patients (Bak, 2013). PD patients demonstrate deficits in generating action verbs (Péran et al., 2009; Piatt, Fields, Paolo, Koller, & Tröster, 1999; Signorini & Volpato, 2006), are impaired at making lexical and semantic decisions about action words (Bocanegra et al., 2015; Boulenger et al., 2008; Fernandino et al., 2013a), and

Journal Pre-proof

exhibit a diminished ACE (Ibáñez et al., 2013). An action naming impairment has also been found (Cotelli et al., 2007), which is modulated by the degree of motility implied by the verb (Bocanegra et al., 2017; Herrera, Rodríguez-Ferreiro, & Cuetos, 2012; Humphries, Holler, Crawford, Herrera, & Poliakoff, 2016). Furthermore, studies of natural discourse in PD patients show that whilst overall communicative output is unchanged, patients exhibit impairments in the spontaneous production of action-concepts in speech (Garcia et al., 2016) and gesture (Cleary, Poliakoff, Galpin, Dick, & Holler, 2011; Humphries et al., 2016). Most of these studies tested patients on their usual levodopa medication relative to controls, but some specifically compared the patients' own performance both on- and off-medication. In a lexical decision task, Boulenger et al. (2008) found that priming effects for verbs were significantly reduced in PD patients when they were off-medication relative to on, as well as when off-medication performance was compared to the control group performance, whilst medication status had no effect on the priming of nouns. Patients on-medication showed similar priming effects to controls. In another study examining verbal fluency, PD patients off-medication generated fewer words for action and phonological categories than they did on-medication, and relative to controls, whilst patients on-medication performed similarly to controls. Medication status did not affect fluency for other semantic categories (animals and shopping) (Herrera, Cuetos, & Ribacoba, 2012). Finally, in an action naming task, Herrera and Cuetos (2012) found that PD patients were slower to name high-motion actions than low-motion actions but only when they were off-medication relative to on. If motor and pre-motor regions contribute to the conceptual representation of action words, dopamine replacement may go some way to ameliorating PD patients' impairments in processing these words. However, it is notable that most studies find that PD patients are still impaired in action language processing even when on their usual medication, relative to controls, suggesting that just as dopaminergic medication may not completely alleviate motor symptoms, it may also not be able to completely restore access to conceptual action representations.

Recent work has begun to uncover the neural mechanisms underlying this impairment. When processing action verbs, healthy people exhibited functional connectivity between M1 and the inferior frontal gyrus, and within subcortical basal ganglia structures. In contrast, PD patients demonstrated reduced connectivity within the basal ganglia, and increased connectivity between M1 and posterior regions, which was positively correlated with the degree of atrophy of the basal ganglia, suggesting a compensatory process (Abrevaya et al., 2017). In sum, embodied cognition theories are bolstered by the fact that motor-impaired patients exhibit a specific impairment in comprehending and producing action language.

Whilst the involvement of motor systems in the comprehension of literal action language is well-documented, the question of whether or not abstract language is also grounded in sensorimotor systems is more controversial. Metaphors employ objects, actions and other concepts in non-literal ways, allowing us to reason about abstract ideas by reference to more concrete and familiar concepts (Bowdle & Gentner, 1999; Gallese & Lakoff, 2005; Gentner, Bowdle, Wolff, & Boronat, 2001; Gibbs, 1994, 2005, 2006; Lakoff & Johnson, 1999, 2008). Nominal (noun based) metaphors, such as "fear is a roadblock", liken two dissimilar semantic domains, one of which (the target) is typically more abstract than the other (the vehicle). Predicate metaphors, by contrast, involve metaphorical extensions of verbs (e.g. "The stock soared.") Proponents of embodied cognition have argued that processing metaphors involves simulating the literal sensorimotor features of the vehicle (roadblock) to apply to the target (fear) in nominal

Journal Pre-proof

metaphors (Gallese & Lakoff, 2005; Gibbs, 2006; Lakoff & Johnson, 1999, 2008). Similarly, comprehending predicate metaphors would require simulating the literal sense of the verb that is being used metaphorically. However, evidence in support of this claim is mixed. Some studies show that motor regions are activated when people read both literal and figurative uses of action verbs (Boulenger, Hauk, & Pulvermüller, 2009; Boulenger, Shtyrov, & Pulvermüller, 2012), that the extrastriate body area is active when reading literal and metaphorical sentences about body parts (Lacey et al., 2017), and that gustatory regions respond to taste metaphors as well as literal taste sentences (Citron & Goldberg, 2014). In contrast, other studies observe sensorimotor activation only in response to literal, but not figurative sentences (Aziz-Zadeh et al., 2006; Chen, Widick, & Chatterjee, 2008; Raposo, Moss, Stamatakis, & Tyler, 2009a). In a study of German verbs, abstract verbs (such as *begreifen*, to comprehend) built on motor stems (*greifen*, to grasp) did not activate the motor system any more than abstract verbs built on abstract stems (Rüschemeyer, Brass, & Friederici, 2007), disputing the idea that abstract concepts are grounded in sensorimotor systems. In fact, a previous study in PD patients found that whilst they were slower to respond to literal action sentences than abstract sentences relative to controls, this effect was not found for metaphorical action sentences (Fernandino et al., 2013b).

One possible explanation for this discrepancy is that studies differ in the types of figurative language they have used as stimuli (Schmidt, Kranjec, Cardillo, & Chatterjee, 2010; Yang, Edens, Simpson, & Krawczyk, 2009). Some used highly familiar sentences such as conventional metaphors or idioms, whereas others used more novel metaphors. Whilst novel metaphors must first be understood in reference to their literal senses (even if only to inhibit those irrelevant concrete features), the initially novel figurative extension of a word might become abstracted over many encounters, such that over time this new word sense is completely lexicalized and can be accessed without reference to its literal features (Jamrozik et al., 2016). Cardillo et al. (2012) investigated the neural basis of this abstraction process by repeatedly exposing participants to novel metaphors such that they became more familiar. They found that the conventionalization of metaphor meanings resulted in decreased neural load: activation in a left-lateralized semantic network decreased as metaphors became more familiar. Furthermore, Desai et al. (2011) found that the familiarity of action verb metaphors was negatively correlated with the activity they elicited in primary motor cortex. A "weak embodiment" view of the embodied approach to language comprehension has therefore emerged, suggesting that only novel metaphors require sensorimotor simulations of their literal elements for comprehension. Whilst the study of action metaphor processing in PD patients by Fernandino et al. (2013b) distinguished between idioms (highly conventionalized non-literal phrases) and metaphors, the metaphor stimuli they employed were still highly familiar (e.g. "The war raised the price of wheat and rice."). It is therefore possible that participants understood them directly in terms of their abstracted, conventionalized sense, which could explain why PD patients did not demonstrate impaired action metaphor processing. The "weak" version of embodiment is consistent with both the Language and Situated Simulation (LASS) theory (Barsalou, Santos, Simmons, & Wilson, 2008) and the Embodied Conceptual Combination (ECCo) theory (Lynott & Connell, 2010). Both argue that simulation is not always necessary for language comprehension, and that multiple systems interact in the representation of conceptual knowledge. LASS and ECCo emphasise two conceptual processing systems: a statistical, distributional, linguistic information system, and a situated, modal, simulation system. It has been proposed that both systems are activated when a word is perceived, but that the linguistic system peaks earlier than the simulation system.

Connell and Lynott (2013) show that the faster linguistic system can provide a shortcut to processing when only shallow conceptual representations are required, but that the simulation system is necessary for deeper conceptual processing where the linguistic system does not suffice. In the context of metaphor processing, when encountering a highly familiar metaphor ("The roommates clicked with their new neighbor."), information from the statistical patterns in language might support comprehension of the metaphorical sense of "clicked", because this sense has been encountered in similar contexts many times previously. However, for the comprehension of a more novel metaphor ("The textbooks snored on the desk."), the linguistic system would not suffice if this new metaphorical sense of "snoring" (i.e. where snoring implies sleeping, and suggests that the books are going unused) has not been encountered before. The simulation system would then be required to allow for deeper conceptual processing to resolve this new metaphor.

A further confounding factor in the previous mixed results of embodied metaphor processing is variability in the extent to which studies have attempted to characterize and control psycholinguistic features of the stimuli they used (Cardillo, Schmidt, Kranjec, & Chatterjee, 2010). In addition to familiarity, both metaphorical and literal sentences might vary in their syntactic complexity, naturalness, imageability, length, interpretability, figurativeness, frequency and concreteness, any or all of which may contribute to ease of comprehension and neural demands (Balota, Yap, & Cortese, 2006; Constable et al., 2004; Friederici, Fiebach, Schlesewsky, Bornkessel, & Von Cramon, 2006; Olaf Hauk, Davis, & Pulvermüller, 2008; Just, Newman, Keller, McEleney, & Carpenter, 2004). These factors of non-interest can create differences in difficulty between conditions, potentially confounding experimental manipulations. This problem was demonstrated in an fMRI study showing that varying factors like concreteness in metaphors affected task difficulty and resulted in different patterns of neural activation (Yang, Edens, Simpson, & Krawczyk, 2009). To meaningfully test embodiment hypotheses about metaphor, it is therefore critical to ensure that sentences in different conditions are as closely matched on these psycholinguistic features as possible.

To address these discrepancies, we conducted a Registered Report study to investigate the comprehension of novel metaphors in PD patients (without dementia) in a highly controlled task. We compared the speed and accuracy with which PD patients comprehended novel action metaphors relative to novel sound metaphors on a task requiring deep semantic processing. An advantage of comparing action metaphors to sound metaphors, rather than to abstract sentences (a common practice), is the elimination of psycholinguistic differences likely to impact their neural processing. Features such as concreteness and imageability are likely to be significantly lower in abstract sentences than metaphors, but can be more closely matched between metaphors referring to different sensory modalities. Based on the hypothesis that novel metaphor processing is "embodied", we predicted that PD performance would be impaired on action metaphors but not sound metaphors. An additional test compared performance on literal action and sound sentences, which were matched to the metaphorical sentences in terms of the base term (e.g. metaphor: "The test review was a quick jog.", literal: "The racecourse was an easy jog."). In addition, half of the sentences were of a nominal form, involving metaphorical extensions of event nouns (e.g. metaphor: "The puzzle was a logic cartwheel.", literal: "The gymnastics stunt was a cartwheel.") and half were of a predicate form, involving metaphorical extensions of verbs (e.g. metaphor: "The frank speaker sailed towards a finish.", literal: "The boat sailed towards the

Journal Pre-proof

sandy shore.”). This design allowed us to address syntactic structure (nominal and predicate) as orthogonal to sensorimotor semantics (motion and sound), where these have previously been conflated in the literature (Cardillo et al., 2010).

Our main question of interest was whether PD patients were impaired in the comprehension of action metaphors compared to sound metaphors. However, the design of the experiment allowed us to test secondary hypotheses to tease apart several kinds of impairments that PD patients might exhibit. Alongside a modality specific impairment, previous work has suggested that frontostriatal executive dysfunction contributes to a more general metaphor impairment in PD (Berg, Björnram, Hartelius, Laakso, & Johnels, 2003; Lewis, Lapointe, Murdoch, & Chenery, 1998; Monetta & Pell, 2007). On this account, we may have found a general metaphor comprehension impairment that was correlated with performance on executive function and working memory measures. In addition, the design of the present study allowed us to address the potential effect of syntax on metaphor comprehension in PD. The first sets of studies reporting action language impairments in PD examined performance on verbs compared to nouns, and reported a verb-specific impairment (Bertella et al., 2001; Boulenger et al., 2008; Péran et al., 2003). More recent studies have revised this thinking to emphasize that the PD impairment is in action semantics rather than verbs per se (Bocanegra et al., 2015; Cardona et al., 2013). Since our noun and verb-based metaphors both extend action words metaphorically, we expected PD patients to be impaired on both types, though the cognitive process involved in comprehending these sentences is likely to be different. Nominal (noun) metaphors might be understood through a process of analogy, feature mapping, or categorization between the two conceptual domains (Bowdle & Gentner, 1999; Gentner et al., 2001; Glucksberg & Keysar, 1990; Glucksberg et al., 1997). In predicate metaphors, in which a verb is extended metaphorically, there is no explicit comparison being made between two concepts. Instead, predicate metaphor comprehension may involve a shedding of the irrelevant sensorimotor features of the verb (Chen et al., 2008). Though these two types of metaphors appear to involve different processes in their comprehension, we found no differences between the neural substrates underlying the processing of predicate metaphors and nominal metaphors using nominalized verbs as the vehicle (e.g., “a slump”; Cardillo et al., 2012). Assuming that the semantics of the metaphor vehicle are more critical than the metaphor’s syntax, we did not expect to find differences in nominal and predicate metaphor performance in PD patients. The sentence stimuli we used were extensively characterized in terms of their psycholinguistic properties, matched extremely closely across conditions and normed in a non-elderly adult population, providing greater precision to test embodied hypotheses about metaphor than in previous studies.

To summarize, we tested the following hypotheses:

1. **Figurative motor concept:** Motor impairment impairs comprehension of figurative language that use action concepts as a vehicle. We predicted that PD patients would be impaired in the comprehension of novel action metaphors compared to novel sound metaphors. Control participants were predicted to show no difference in performance between the two conditions.
2. **Literal motor concept:** Motor impairment impairs comprehension of literal action concepts. We similarly predicted that PD patients would be impaired in the comprehension of literal action sentences compared to literal sound sentences. Again,

Journal Pre-proof

control participants were predicted to show no difference in performance between the two conditions.

3. **Syntax:** Motor impairment does not affect comprehension based on syntactic processing. Consequently, we predicted that neither group would show a difference in comprehension performance between nominal and predicate metaphors.
4. **Figurativeness:** Metaphorical language is more difficult to comprehend than literal language and relies more on executive functioning and relational reasoning ability. We predict that both groups would be worse at comprehending metaphors than literal sentences, but the magnitude of the difference would be greater in PD patients.

2. Method

2.1. Sampling Plan

We analysed the data using Bayesian methods, which permit optional stopping. This allowed us to maximize the efficiency of our sampling, which is particularly advantageous when conducting patient studies. Prior to data collection, we planned to calculate Bayes factors sequentially as participant numbers increased until we reached sufficient evidence for either the null or the alternative hypothesis, or until we exhausted the potential subject pool (see Data Analysis section below for further details). Before collecting any data, we estimated that the maximum number of PD patients we would be able to recruit was approximately 50. To ensure that this number of participants could feasibly address our hypotheses, we conducted a Bayesian reanalysis of summary statistics reported in similar previous studies which had comparable sample sizes (note that in several cases this was not possible because only p values and not test statistics were reported). To be convincing, our study would need to find either evidence for the alternative hypothesis (Bayes factor (BF) > 3) or evidence for the null (BF < 1/3) for all tests. A BF < 3 and > 1/3 means that the data are inconclusive and cannot provide convincing evidence for either the null or the alternative.

For predictions from hypotheses 1 and 2 which relate to an impaired motor concept in PD, we examined previous studies which compared performance on an action condition vs. a non-action condition, or a high action vs. low action condition (since the sound sentences in our stimuli still feature "action" words, albeit low motion actions) between PD patients and controls. Cotelli et al. (2007) compared object and action naming performance in 32 PD patients and 15 controls. Controls named more objects and actions correctly than patients but the difference was larger in the action condition. The F value for the interaction was 41.763, which can be square rooted to get a t-value of 6.46 for the t-test on the difference of differences. Cohen's d for this t-value was 2.02 – a very large effect – which we calculated using the effect size calculator provided by Lakens (2013). Using the Summary Statistics module in JASP (JASP Team, 2017), under the alternative we entered the t-value, n for each group, selected a one-tailed test and set the Cauchy prior width to 2. The associated BF is 447544, indicating overwhelming evidence in favor of the alternative. Had there been no effect (examined by setting t to 0), the BF is .12, which is low enough to provide evidence for the null. Reducing the Cauchy prior width to the default of .707 retains these conclusions: the BF for the alternative is 297240, and BF for the null is .306, demonstrating robustness. We repeated this procedure based on parameters reported in Herrera et

Journal Pre-proof

al. (2012), in which 49 PD patients and 19 controls were compared while they named high and low motion actions. There was an interaction between group and motion content where controls named more actions correctly than patients in both conditions, but the magnitude of the difference was larger for high motion actions. The reported F value of the interaction was 62.49 which we square rooted to get a t-value of 7.905 for the t-test on the difference of differences. Cohen's d was again very large, 2.14. Setting the Cauchy prior width at 2, the BF for the alternative is 1.656e+6 which is again very strong evidence. Had there been zero effect, the BF is .106 which is sufficient evidence in favour of the null. Reducing the Cauchy prior width to the default .707 retains these conclusions; the BF for the alternative is 3.119e+8, and the BF for the null is .272.

For predictions from hypothesis 3 (no effect of syntax), to our knowledge no study has compared behavioural performance for the comprehension of nominal and predicate metaphors, so we were not able to repeat the sample size sufficiency procedure for this hypothesis.

For hypothesis 4 which predicted a general metaphor impairment in PD, we used parameters reported in Monetta and Pell (2007) in which 17 PD patients and 17 controls were tested on metaphor comprehension. After viewing a metaphorical prime sentence, controls made fewer errors when responding to a metaphor relevant target than a metaphor irrelevant target, whereas PD patients made similar numbers of errors in both conditions. The reported F value of the interaction was 6.19 and the associated t-value for the difference of differences is 2.49. Cohen's d for this t-value is .85, so we set the Cauchy prior width at this value. The BF for the alternative is 6.189, providing evidence in favour of the alternative, and under the null BF is .285 which provides evidence in favour of the null. To check robustness, reducing the Cauchy prior width to the default of .707, the BF for the alternative is 6.33. However, the BF for the null is .329 which is only just shy of 1/3 and thus provides only weak evidence in favour of the null. The sample size in this study was small, and we aimed to recruit more than this. Increasing participant numbers even just slightly to 20 in each group reduced the BF for the null to .283 which was more convincing.

Whilst these sample size sufficiency calculations are just estimates, BFs calculated for previous studies suggested that 50 subjects in each group would be sufficient to test our hypotheses. To sample efficiently, we calculated BFs for all of our effects beginning at 20 subjects per group and sequentially from there with the aim of stopping when the BFs indicated sufficient evidence for the null or the alternative in each case, or when we reached our practical recruitment maximum.

2.2. Participant Recruitment

Patients with PD were recruited from the University of Pennsylvania's Udall Center for Parkinson's Research. All recruited patients were diagnosed with idiopathic PD by a neurologist at the Udall Center, according to UK Parkinson's Disease Society Brain Bank criteria (Hughes, Daniel, Kilford, & Lees, 1992), and underwent a comprehensive neurological and cognitive neuropsychological assessment. Inclusion criteria for the study included scoring within the normal range on the MMSE/MoCA, and being a native US-English speaker. Exclusion criteria included a diagnosis of dementia, a history of stroke, a diagnosis of any other neurological

Journal Pre-proof

condition other than Parkinson's, a previous traumatic brain injury, or a sight or hearing impairment preventing the person from easily reading text on a computer screen or understanding verbal instructions. Patients were tested on their usual medication at either Pennsylvania Hospital or their own home. As described in the introduction, stronger effects might be observed in patients in the off-medication state but, nevertheless, previous studies have still found strong effects with patients in the on-state. We calculated the PD participants' daily levodopa equivalent doses (Tomlinson et al., 2010) (see Table 1). Healthy age and education-matched controls were recruited from two existing control databases maintained by the Penn Memory Center and the Penn Center for Cognitive Neuroscience. Any control participants who had not had their cognitive status confirmed as "normal" by testing within the previous 12 months were administered the MoCA at the testing session. Anyone scoring outside the normal range (< 26) was excluded from the study.

In total, 44 patients with PD and 48 age-matched controls were recruited. We exhausted the available subject pool and were not able to recruit any further eligible patients. Three PD patients and nine controls were excluded according to our pre-registered exclusion criteria. Of the three patients excluded, two were not native English speakers, and one had a history of severe traumatic brain injury and epilepsy. Of the nine excluded controls, five were excluded prior to data analysis: one had a history of stroke, and four scored below the cut-off on the MoCA. An additional four controls were excluded because of poor performance on the task (the cut-off was accuracy lower than 3 SDs below the mean according to our outlier removal procedure described in section 2.5.2. below). The final study sample included in the analyses consisted of 41 PD patients and 39 controls (see Table 1 for details). The PD group were all classified as cognitively unimpaired at the neurologists' most recent consensus meeting, and they were mostly at the mild to moderate stages of PD motor symptom severity. The groups did not differ significantly in age, education, or MoCA scores.

Table 1. Demographic and clinical features of the Parkinson's and control groups.

	PD patients	**Controls**
Age	67.93 (7.71)	69.56 (8.06)
Education	16.9 (2.04)	17.74 (3.17)
Gender	23 M, 18 F	15 M, 24 F
MOCA	28.05 (1.48)	27.95 (1.47)
	PD patients	
Disease Duration (years)	7.8 (4.13)	
UPDRS-Motor subscale	24.2 (9.48)	
Hoehn and Yahr staging	Stage 1: 1	
	Stage 2: 28	
	Stage 3: 11	
	Stage 4: 0	
	Stage 5: 1	
Levodopa Equivalent Dose	721.95 (404.51)	

2.3. Materials

Journal Pre-proof

Metaphor and literal sentence comprehension was assessed using an extension of the Metaphor Multiple Choice task previously developed in our lab and successfully used to identify patterns of metaphor impairment in patients with focal brain injury (Ianni, Cardillo, McQuire, & Chatterjee, 2014). We replicated the task and procedure of Ianni et al (2014), but used different sentences (and answer choices) that were optimized to test embodiment hypotheses about metaphor.

2.3.1. Sentences

The stimuli included 120 sentences, consisting of pairs of 60 novel metaphorical sentences, and 60 literal counterparts matched on the metaphor vehicle (see Table 2 for examples). As in Ianni et al (2014), sentences were drawn from a larger set of sentences which were created for the purpose of testing neural hypotheses about metaphor (Cardillo et al., 2010; also see Cardillo, Watson, & Chatterjee, 2016) Each metaphorical and literal sentence in this larger set was extensively normed on a large number of psycholinguistic variables. Half of the sentences in the present study are nominal metaphors, involving metaphorical extensions of nouns, and half are predicate metaphors, involving metaphorical extensions of verbs. Nominal metaphors take the form of two noun phrases connected by a copula (e.g. "The X was a Y"), where the second noun phrase is always the vehicle term of the metaphor. The predicate sentences consisted of a noun phrase, a verb (the vehicle term), and a prepositional phrase. To maximize similarity across metaphors and literal sentences, the same vehicle term was used in each metaphor-literal pair. To maximize similarity across nominal and predicate metaphors, nominal metaphors always used nominalized versions of verbs as their vehicle (see Table 2). Half of the items of each metaphor type were sentences based on verbs of motion and half were based on verbs of sound. The structure of the experiment thus involved two sentence types (metaphor and literal), two syntactic forms (nominal and predicate), and two verb types (motion and auditory), resulting in 8 total conditions with 15 sentences in each (see Table 2).

Table 2. Example sentence pairs in each condition

Condition	**Metaphorical**	**Literal**
NM	His work experience was a clumsy <u>clamber</u>.	The final ascent was an exhausting <u>clamber</u>.
NA	The man's gaze was a shameless <u>slurp</u>.	The last sip was a noisy <u>slurp</u>.
PM	The frank speaker <u>sailed</u> towards a finish.	The boat <u>sailed</u> towards the sandy shore.
PA	The sunset <u>sang</u> to the lovers.	The uncle <u>sang</u> to the baby.

Note: NM = nominal motion, NA = nominal auditory, PM = predicate motion, PA = predicate auditory

The 60 sentence-pairs for this study were selected from the larger set of 280 sentence-pairs described in Cardillo et al. (2010) with the use of the Stochastic Optimization of Stimuli (SOS) software (Armstrong, Watson, & Plaut, 2012). SOS automates the process of stimuli selection from a pool by adhering to constraints set by the user to ensure items in different conditions are closely matched or significantly different among any number of dimensions. The optimization was specified to select 15 metaphorical-literal sentence pairs from each of four populations (nominal motion, nominal auditory, predicate motion and predicate auditory) and match the four sets on a number of criteria. Because we were interested in testing PD patients' performance on novel metaphors, the stimulus selection optimization was specified to keep the familiarity of the metaphorical sentences to a minimum, resulting in overall mean metaphor familiarity of 2.83 (min: 1.58, max: 3.75, rated on a 1-7 scale). When the sentences were rated for familiarity in

Journal Pre-proof

Cardillo et al. (2010), participants were instructed to rate their frequency of experience with the sentence and its meaning (for rating instructions for the other variables reported here, please see Cardillo et al., 2010). SOS was then specified to ensure that metaphors in each condition were matched to metaphors in each of the other conditions on interpretability and figurativeness. In addition, both metaphor sentences and literal sentences in each condition were matched group-wise to the sentences in the other conditions on average frequency and concreteness of content words, familiarity, naturalness, imageability, number of words, and number of content words. Finally, the metaphorical and literal items in each pair were matched to each other in length (number of words and content words), and average frequency (Brysbaert & New, 2009) and concreteness (Brysbaert, Warriner, & Kuperman, 2014) of content words. As observed previously, the literal sentences were significantly more natural and imageable than their metaphorical counterparts (Cardillo et al., 2010; Ianni et al., 2014). Furthermore, because of the constraints placed on selecting novel (unfamiliar) metaphors, literal sentences were significantly more familiar. We also calculated average age of acquisition (AoA) for the content words in each sentence, though this variable was not explicitly controlled for during the stimulus selection procedure. AoA for metaphor sentences did not differ between conditions, and AoA for literal sentences did not differ between conditions. However, collapsing conditions across figurativeness, AoA was significantly higher for metaphorical sentences relative to literal sentences (see Table 3). Importantly, when collapsing conditions across modality, auditory and motion sentences did not differ significantly in AoA for either metaphorical ($t(58) = .81$, $p = .42$) or literal sentences ($t(58) = .95$, $p = .34$) (see Table 3 below for means and standard deviations for items in each condition).

Journal Pre-proof

Journal Pre-proof

Table 3. Psycholinguistic properties of the metaphor and literal sentence stimuli

	Metaphor						Literal						Metaphor-Literal	
	Nominal		Predicate				Nominal		Predicate					
	Auditory	Motion	Auditory	Motion	F	p	Auditory	Motion	Auditory	Motion	F	p	t	p
Interpretability (%)	.93 (.10)	.94 (.06)	.93 (.06)	.94 (.06)	0.05	0.99	n/a	n/a	n/a	n/a	n/a	n/a	0	0
Figurativeness (1-7)	5.64 (.47)	5.81 (.58)	5.80 (.70)	5.50 (1.00)	0.63	0.6	2.00 (.52)	2.11 (.59)	1.73 (.83)	1.64 (.77)	1.58	0.21	25.83	<.001
Frequency	69.73 (121.65)	69.13 (77.08)	59.73 (79.23)	51.27 (53.81)	0.15	0.93	91.73 (102.08)	46.47 (51.83)	102.80 (95.66)	79.60 (117.43)	0.99	0.41	1.19	0.24
Age of Acquisition	7.38 (1.63)	6.98 (1.15)	6.72 (1.32)	7.73 (1.55)	1.46	0.24	6.68 (1.24)	7.01 (1.13)	6.25 (1.22)	6.58 (1.63)	0.82	0.49	2.81	0.007
Concreteness (1-5)	3.61 (.35)	3.61 (.45)	3.86 (.35)	3.68 (.56)	1.12	0.35	3.76 (.37)	3.73 (.52)	3.84 (.50)	3.78 (.60)	0.14	0.94	1.34	0.19
Familiarity (1-7)	2.69 (.62)	3.04 (.54)	2.83 (.72)	2.76 (.51)	0.96	0.42	5.52 (.66)	5.44 (.62)	5.60 (1.15)	5.45 (1.15)	0.1	0.96	19.98	<.001
Naturalness (1-7)	3.15 (.76)	3.45 (.69)	3.00 (.73)	2.96 (.68)	1.47	0.23	6.05 (.49)	5.83 (.59)	5.63 (1.15)	5.59 (1.22)	0.81	0.49	18.22	<.001
Imageability (1-7)	3.55 (.84)	3.55 (.58)	3.19 (.71)	2.99 (.93)	1.95	0.13	5.64 (.55)	6.07 (.63)	5.72 (1.07)	5.88 (1.09)	0.73	0.54	18.18	<.001
No. of words	6.20 (.41)	6.27 (.46)	6.40 (.51)	6.27 (.46)	0.5	0.69	6.13 (.35)	6.20 (.41)	6.27 (.59)	6.47 (.74)	1.04	0.38	0.33	0.74
No. of content words	3.20 (.41)	3.27 (.46)	3.27 (.46)	3.33 (.49)	0.22	0.89	3.13 (.35)	3.20 (.41)	3.40 (.51)	3.40 (.51)	1.4	0.25	0.33	0.74

Journal Pre-proof

Journal Pre-proof

2.3.2. Answer Choices

As in our previous work (Ianni et al, 2014), each metaphoric and literal sentence was paired with four answer choices; one correct target and three foils (see Table 4 for examples). Each answer choice consisted of a two-word phrase, composed of an adjective and a noun. For metaphor items, the target reflected the metaphorical meaning of the sentence, Foil 1 was related to the literal interpretation of the sentence, Foil 2 was the opposite of the metaphorical meaning of the sentence and Foil 3 was unrelated. As in Ianni et al. (2014), the foils were designed in this way to be informative about the type of deficit present. Selecting Foil 1 indicates a literal bias in metaphor comprehension. A Foil 2 selection indicates an impairment in semantic integration, since the metaphorical sense of the sentence was activated but understood incorrectly in the context of the sentence. As Foil 3 is unrelated to the meaning of the sentence, selecting this answer indicates a more general sentence comprehension deficit. The answer choice pattern for the metaphor sentences was mirrored as closely as possible for the literal sentences. The target was the literal meaning of the sentence, Foil 2 was the opposite of the literal meaning of the sentence and Foil 3 was unrelated. Since it was not possible to make Foil 1 answers for the literals of the same nature as the Foil 1 answers for the metaphors, Foil 1 answers were instead designed to be related to the agent of the sentence by category membership, but not implied by the sentence. This meant that Foil 1 in both metaphor and literal conditions were as closely matched as possible, in that they were both strong competitors to the target. In the metaphors Foil 1 is strongly related to the target term of the sentence whilst in the literals it is a strong lexical associate of the agent.

Table 4. Example answer choices for each type of sentence

Condition	Sentence	Target	Foil1	Foil2	Foil3
Met-NA	The dad's decision was a balloon pop.	thwarted plans	party favor	granted permission	oily rag
Met-NM	The puzzle was a logic cartwheel.	complex riddle	gymnastics performance	obvious solution	gnarly tree
Met-PA	The inn groaned at the new guests.	crowded accommodations	audible grumble	plentiful vacancies	winding road
Met-PM	The friend mosied through the photographs.	unhurried looking	strolling companion	detailed review	oil lamp
Lit-NA	The rifle was a loud pop	gun shot	bloody knife	peaceful silence	damp earth
Lit-NM	The gymnastics stunt was a cartwheel.	athletic feat	diving event	clumsy stumble	sunny beach
Lit-PA	Their uncle groaned in the other room.	physical suffering	generous parent	comfortable rest	broken mirror
Lit-PM	The tourists mosied without a clear plan.	relaxed holiday	travel agent	rushed schedule	wood supply

Note: Met = metaphor, Lit = literal. NA = nominal auditory, NM = nominal motion, PA = predicate auditory, PM = predicate motion.

To ensure equal difficulty across conditions, the frequency and concreteness values of the answer choices were matched, using the Brysbaert databases (Brysbaert & New, 2009; Brysbaert et al., 2014). Mean frequency values of each type of answer choice (target, foil 1, etc.) in each condition were not significantly different from the frequency values of the same answer choice

Journal Pre-proof

type in all other conditions (see Supplementary Table 1 for means and SDs). In addition, collapsing across all conditions, the mean frequency of each answer choice type was not significantly different from the other three answer choices. Each answer choice type was also matched between conditions for concreteness. However, collapsing across all conditions, some expected significant differences in concreteness between the four answer choices emerged. Because metaphorical sentences generally communicate abstract ideas, the target meanings of these sentences were naturally less concrete than foil 1, the literal interpretations of the sentences. As in Ianni et al (2014), to ensure that no one answer stood out as being more or less concrete than the other three options, target and foil 2 were matched as low-concreteness competitors, and foil 1 and foil 3 were matched as high-concreteness competitors.

2.3.3. Norming

To ensure target answers were reliably identified and conditions were matched in difficulty, sentences and answer choices were normed several times and iteratively updated using Qualtrics presentation software and Amazon's Mechanical Turk subject pool (details to be reported elsewhere as part of a separate methods publication). The final stimuli set we used in this study elicited high overall accuracy in native English-speaking, healthy adults (mean age = 36.95, SD = 7.12) with a minimum high school education. The literal sentences were answered with 97.23% mean accuracy (SD = 3.05) and the metaphors with 92.46% mean accuracy (SD = 6.06), and this difference was significant ($t(43) = 6.87$, $p < .001$) consistent with the unfamiliarity of the metaphors. Nonetheless, no individual item was answered with less than 72% accuracy across the whole sample, confirming the intelligibility of the sentences and appropriateness of the answer choices. Importantly, these stimuli showed no significant difference in the accuracy of auditory versus motion items ($t(43) = .57$, $p = .57$). Thus, in non-elderly, healthy adults this task elicited a high degree of accuracy, without ceiling effects or the modality difference we predicted in PD patients. We therefore had confidence that the stimuli were well-designed to detect group differences in performance that would not be marred by floor or ceiling effects.

2.4. Procedure

Before testing, all participants gave informed consent in accordance with procedures from the University of Pennsylvania's IRB. Participants were compensated $20/hour for their time.

The Metaphor Multiple Choice task was presented to participants using E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA). Each participant viewed the items in a random order. On each trial, the sentence was displayed at the top of the screen, with the four possible answer choices arranged in a square below it. The position of the target and three foils was randomized per item and per participant. Participants responded using a custom response button box built by the Black Box Toolkit (http://www.blackboxtoolkit.com/urp.html), which has 5 buttons (see Figure 1). Four outer buttons were arranged in a square, and therefore mapped spatially onto the four possible answer choices displayed on the screen. To aid visual discrimination of the buttons, the four outer buttons were colored. Both accuracy and reaction times (RTs) were collected. We anticipated that RTs from PD patients were likely to be longer overall and a noisier measure than those from controls. The RT collection procedure was standardized to ensure that 1) any general effects of PD on RT affected all conditions equally, 2) fatiguing effects of the experiment on PD patients were minimized, and 3) the natural motor advantage of controls was minimized. The center button was used to start each new trial, which ensured that every participant began every trial from the same anchor point, and that the movement distance to each button was the same across trials. Because PD presents asymmetrically, patients are likely to have one hand more affected than the other in terms of motor symptom severity. Hypothetically, the fastest responses on a multiple-choice RT task might come from using both hands, with multiple fingers hovering over several buttons simultaneously. However, using both hands and multiple fingers to respond would be disproportionately more challenging and fatiguing for PD patients (particularly those with asymmetric symptoms), and thus would represent a large RT advantage for controls. For this reason, all patient and control participants were instructed to respond using just the pointer finger from their "best" hand only. Participants were familiarized with the task and response button method over three practice trials at the beginning of the task. They were instructed to respond as quickly and accurately as possible with their preferred hand, but there was no time limit for responses. After each item, participants pressed the start button when they were ready for the next item to be displayed. This procedure meant that participants were able to take a break after any item, although prescribed breaks were built in at 30-trial intervals.

In addition to the Metaphor Multiple Choice task to assess metaphor comprehension, participants completed the Object and Action Naming Battery (Druks, 2000), as well as an action fluency task (generate as many verbs as possible in one minute) and standard verbal fluency measures, to assess the possible presence of a lower order cognitive action-language impairment.

Finally, PD patients will underwent an extensive assessment of their motor symptoms (UPDRS, Hoehn and Yahr, grooved pegboard) and cognitive/neuropsychological status (Mattis Dementia Rating Scale-2, Hopkins Verbal Learning Test, letter-number sequencing test, trail-making A and B, clock drawing, symbol digit modalities test, Benton judgement of line orientation, Boston Naming Test, phonemic and semantic fluency, and Montreal Cognitive Assessment) as part of their participation in the UPenn PD research program. This assessment typically took place on a different day to the metaphor experiment testing session.

Journal Pre-proof

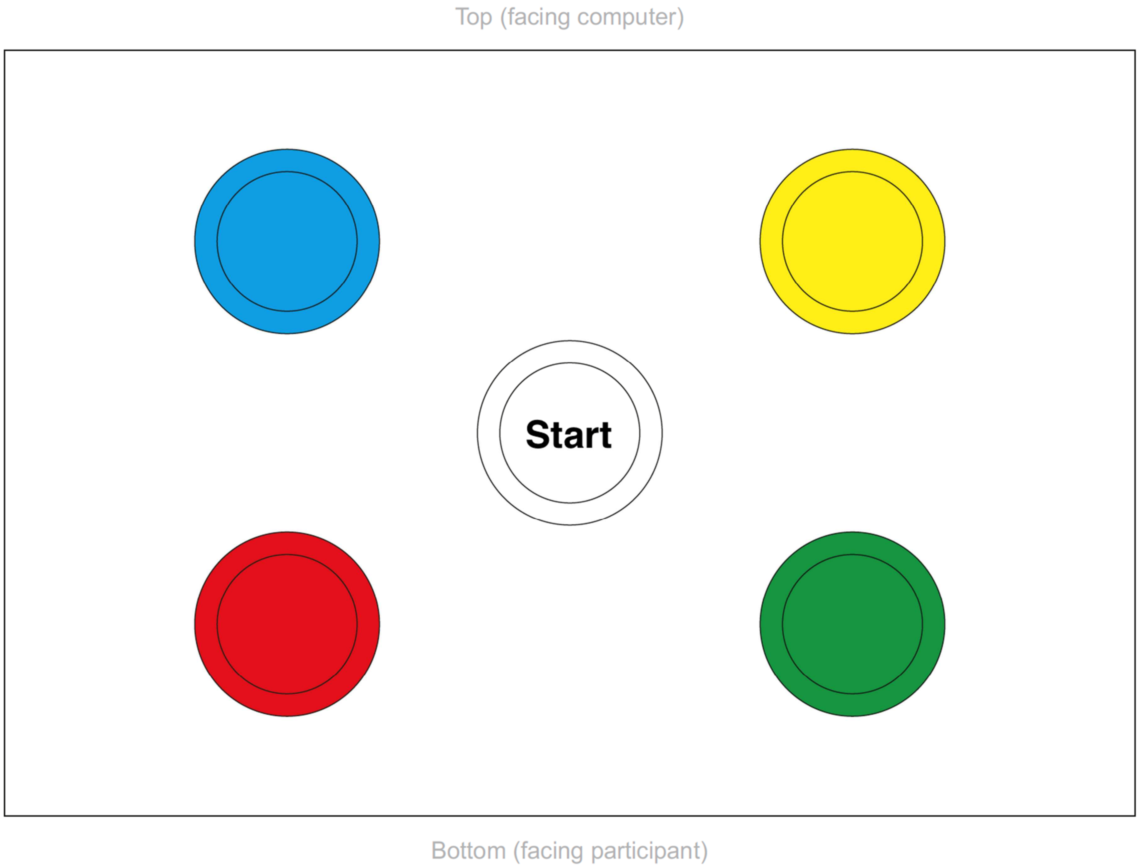


Figure 1. Depiction of the button box that participants will used to respond. The start button beganins each trial and the coloured buttons mapped spatially onto the four possible answer choices displayed on the screen.

2.5. Data Analysis

2.5.1. Data Quality Checks

To ensure that the results obtained were able to address our hypotheses, we conducted a series of data quality checks. First, we checked for the absence of floor and ceiling effects. If PD patients performed at ceiling for accuracy, then the task was too simple and we could not test our hypotheses. Likewise, if the task was so difficult that even controls were performing at floor, we again would not be able to make any inferences about the effect of an impaired motor system on action language comprehension. Ceiling performance would be 100% accuracy. Based on the results of the stimuli norming, we expected controls to perform close to ceiling on literal sentences (~97%) but significantly lower on metaphorical sentences (~92%). We also checked that PD patients' mean accuracy did not approach 100% (>97%) in any condition. Floor performance would be chance, or 25% accuracy. We checked that control performance did not approach floor (<35%) in any condition. In the case of PD patients, the most extreme version of

Journal Pre-proof

our hypotheses 1 and 2 might predict normal comprehension of sound sentences and a total failure to comprehend action sentences. While this would have been highly unlikely, we had to allow for the possibility of substantially poorer performance in action conditions. That said, if performance were at floor for sound sentences, we would not be able to address the effect of modality. Thus, we checked that PD patients' accuracy performance did not approach floor (<35%) in sound sentence conditions.

In a second data quality check we examined the error patterns produced by participants. The results of the stimuli norming show that people tend to make similar kinds of errors. On metaphor trials, most error choices are foil 1, which relate to the literal interpretation (71.7%), with some foil 2 choices which relate to the opposite of the metaphorical interpretation (25.76%), and very few foil 3 choices which are completely unrelated (2.52%). This indicates that when people make an error in metaphor comprehension, they usually have comprehended most of the meaning of the sentence but have been biased towards the literal interpretation. If a subject committed a high proportion of foil 3 errors, this would indicate either a lack of attention to the task or a substantial general language comprehension impairment, which would make it difficult for us to address our subtler questions of the effects of figurativeness and modality. We examined the error patterns at both the group and the individual level. Any participant whose error patterns appeared to be more random (25-40% of each kind of error), or where the proportion of foil 3 errors exceeded 30%, was excluded and replaced with a new participant.

2.5.2. Outlier removal

The two dependent variables we examined were accuracy (proportion of comprehension questions answered correctly) and RT (for correct responses only). We planned to exclude any experimental item answered with accuracy over 3 SDs below the mean in the control group, but no items met this criteria.. Any control participants responding with accuracy over 3 SDs below the mean in any condition were excluded (four controls were excluded according to this criteria). For each participant in each condition, boxplots were constructed to enable the detection of RT outliers (defined as those which were more than 1.5 interquartile ranges below the lower quartile [Q1] or above the upper quartile [Q3]).These RT outliers were Winsorized by replacing them with the closest non-outlying value, i.e., the values representing either Q1-1.5*IQR or Q3+1.5*IQR (Erceg-Hurn & Mirosevich, 2008).

2.5.3. Data analysis

For the prediction from hypothesis 1, the comparison between motion and auditory metaphors for PD patients relative to controls, we ran a Bayesian repeated measures ANOVA for both accuracy and RT as dependent measures. BFs were calculated using JASP (JASP Team, 2017). As reported in the sampling plan above (section 2.1), previously reported effect sizes for action language impairments in PD have been very large: approximately Cohen's d = 2. Large effect sizes are more typical in neuropsychological studies (Bezeau & Graves, 2001). However, we accepted the possibility that our effects would be smaller than this, given how well our control conditions were matched to the experimental conditions, and particularly in the metaphor conditions. Given this uncertainty, we report BFs under a range of Cauchy prior widths including 2 (based on previous effects), as well as the default (.707) to determine the robustness of the

Journal Pre-proof

effects. For metaphorical sentences only, the factors Group (PD, control), and Modality (motion, auditory) were entered. We compared the model with the interaction between Group*Modality against the null model, the model associated with each separate main effect, and the model with both main effects, to determine whether the model with the interaction was preferred (as evidenced by the BF associated with each model). A key test of our hypothesis was that PD patients should show a difference in performance between the motion and auditory metaphor trials, whilst controls should show no difference between these conditions. That is, we predicted that the evidence would favour the null hypothesis in the control group. We tested this by examining the simple effect of Modality for each group separately with Bayesian t-tests. One-sided tests were used since our theory predicted that performance on motion trials will be worse than performance on auditory trials in PD patients. Cut-offs of a BF of > 3 in the PD group and < 1/3 in the control group were used to decide if our hypotheses were supported. This analysis procedure was repeated for hypothesis 2, comparing motion and auditory literal sentences.

For the prediction from hypothesis 3, the comparison between nominal and predicate metaphors for the two groups, we again conducted a Bayesian repeated measures ANOVA for both accuracy and RT. Since we had no previous effect sizes to guide the prior, we used the default Cauchy prior width of .707 in JASP. The factors Group and Syntax were entered. We compared the model with the interaction between Group*Syntax against the null model and the models with only the main effects to determine whether there was an interaction. If an interaction was found, we planned to follow this up by testing the simple effect of Syntax for each group separately using two-sided Bayesian t-tests.

For the prediction from hypothesis 4, the comparison between metaphorical and literal sentences for the two groups, we conducted a Bayesian repeated measures ANOVA for accuracy and RT. The Cauchy prior r-scale width was set at .85 in accordance with Monetta and Pell (2007). We again tested the robustness of any effects by examining BFs at a range of different prior widths. Figurativeness (metaphor, literal) and Group were entered. We compared the model with the interaction Figurativeness*Group against the null model and the models with only the main effects to determine whether the model with the interaction was preferred. In this case, we did not predict a null effect in the control group. Controls were also expected to find metaphors more difficult than literals, but we predicted that the effect would be of a greater magnitude in PD patients. We tested the simple effect of metaphor vs. literal for each group separately using Bayesian t-tests (one-sided). We expected a BF > 3 in each group but a larger BF in the PD group.

3. Results

3.1. Overall Task Performance, Data Quality Checks, & Outliers

Overall, both groups of participants performed the tasks well (means and SDs reported in Table 5). As expected, accuracy was close to ceiling for literal sentences. When considering all literal sentences together (collapsing the various modality and syntax conditions), literal comprehension accuracy was 97.56% for controls and 95.16% for PD patients. Metaphor comprehension accuracy was lower but still high: controls answered 89.4% of the metaphors correctly and PD patients 88.17%. The data therefore satisfy our preregistered data quality

Journal Pre-proof

checks relating to the absence of floor and ceiling effects (see section 2.5.1. above). The PD group's accuracy did not exceed 97% in any condition and was not below 35% in any auditory condition. We accepted that control accuracy would be close to ceiling in literal conditions, but critically, controls did not perform at ceiling (>97%) or floor (<35%) in any metaphor condition.

Table 5. Means and SDs for accuracy and response time, separated by group and by condition.

	Metaphor				Literal			
	Nominal		Predicate		Nominal		Predicate	
	Auditory	Motion	Auditory	Motion	Auditory	Motion	Auditory	Motion
	Accuracy							
PD	85.04 (14.82)	81.95 (15.29)	92.85 (7.66)	92.85 (8.48)	96.75 (5.4)	94.96 (6.8)	94.96 (8.66)	93.98 (6.11)
NC	84.62 (11.15)	87.17 (11.59)	93.33 (7.49)	92.48 (9.45)	98.97 (2.44)	96.92 (4.0)	97.26 (4.25)	97.09 (4.27)
	Response Time							
PD	8237 (2759)	8321 (3089)	7380 (2760)	8059 (3018)	6585 (2383)	6984 (2303)	7524 (2520)	6705 (2443)
NC	7462 (1939)	7321 (1963)	6741 (1875)	6799 (1665)	5803 (1475)	6145 (1680)	6752 (1749)	5761 (1373)

When participants made an error, we examined which of the three foils they selected. As described in the materials section above, in the metaphor conditions foil 1 was a literal interpretation, foil 2 was the opposite of the metaphorical meaning, and foil 3 was unrelated. In the literal conditions, foil 1 was related to the agent of the sentence by category membership, foil 2 was the opposite of the literal meaning, and foil 3 was unrelated. A data quality check we planned was to check that participants made sensible errors (see section 2.5.1.). When a person fails to comprehend a metaphor, they usually interpret the sentence literally. A large proportion of foil 1 errors in the metaphor conditions reveals a literal bias and impaired ability to derive novel metaphoric meanings. Conversely, if a person makes a large number of errors and their error patterns appear random, or they make a large number of foil 3 errors, this pattern may point to a more general language comprehension impairment. We found that both groups of participants made sensible errors (see Table 6). In the metaphor conditions, 76% of errors made were foil 1 selections (a literal interpretation). In the literal conditions, errors were split more equally between foil 1 and foil 2. The error patterns looked similar between the two groups and we did not exclude any participants for making unusual errors. The observed error patterns were also highly similar to those found in previous patient studies from our lab using variants on the metaphor multiple choice task (Cardillo, McQuire, & Chatterjee, 2018; Ianni et al., 2014).

Only response times for correctly answered sentences were retained. We discarded the response time data for one control subject as the button box was not available at their testing appointment. Response time outliers, defined as those extending more than 1.5 interquartile ranges beyond the upper or lower quartiles (for each condition and each subject individually), were Winsorized to the boundaries according to our preregistered analysis plan. Response times from 8777 correctly answered trials were included. Of these, 464 (5.29%) were Winsorized.

Journal Pre-proof

Table 6. Types of errors made in the metaphor and literal conditions by each group.

		Metaphor			Literal		
		Foil 1 (literal)	**Foil 2 (opposite)**	**Foil 3 (unrelated)**	**Foil 1 (agent category)**	**Foil 2 (opposite)**	**Foil 3 (unrelated)**
PD	**Percent**	76.29%	19.24%	4.47%	42.02%	43.7%	14.29%
	Sum	222	56	13	50	52	17
	Mean	5.41	1.37	0.32	1.22	1.27	0.41
NC	**Percent**	76.61%	22.58%	0.81%	56.14%	35.09%	8.77%
	Sum	190	56	2	32	20	5
	Mean	4.87	1.44	0.05	0.82	0.51	0.13

3.2. Preregistered Analyses

3.2.1. Hypothesis 1: Figurative Motor Concept

To test whether an interaction was present between Group (PD/Control) and Modality (Auditory/Motion), we compared a model which included the main effects of Group and Modality against a model which included the interaction term. We ran a repeated measures JZS Bayes factor ANOVA (Morey & Rouder, 2015; Rouder, Morey, Speckman, & Province, 2012) in JASP. Previous literature indicated we should expect a large effect, so a wide prior was used (r scale for fixed effects = 1).

For accuracy, the interaction BF_{10} was .208, indicating moderate evidence for the null hypothesis. Put another way, the data were about 4.8 times more likely under the main effects model than under the model with the interaction. To allow for a potentially smaller effect size, we repeated the analysis using a narrower prior (the default prior in JASP: r scale = .5). This resulted in a BF_{10} of .4, which again provides evidence for the null hypothesis albeit slightly less compelling than under the wide prior. The accuracy data for Group and Modality are plotted in Figure 2.

Journal Pre-proof

Journal Pre-proof

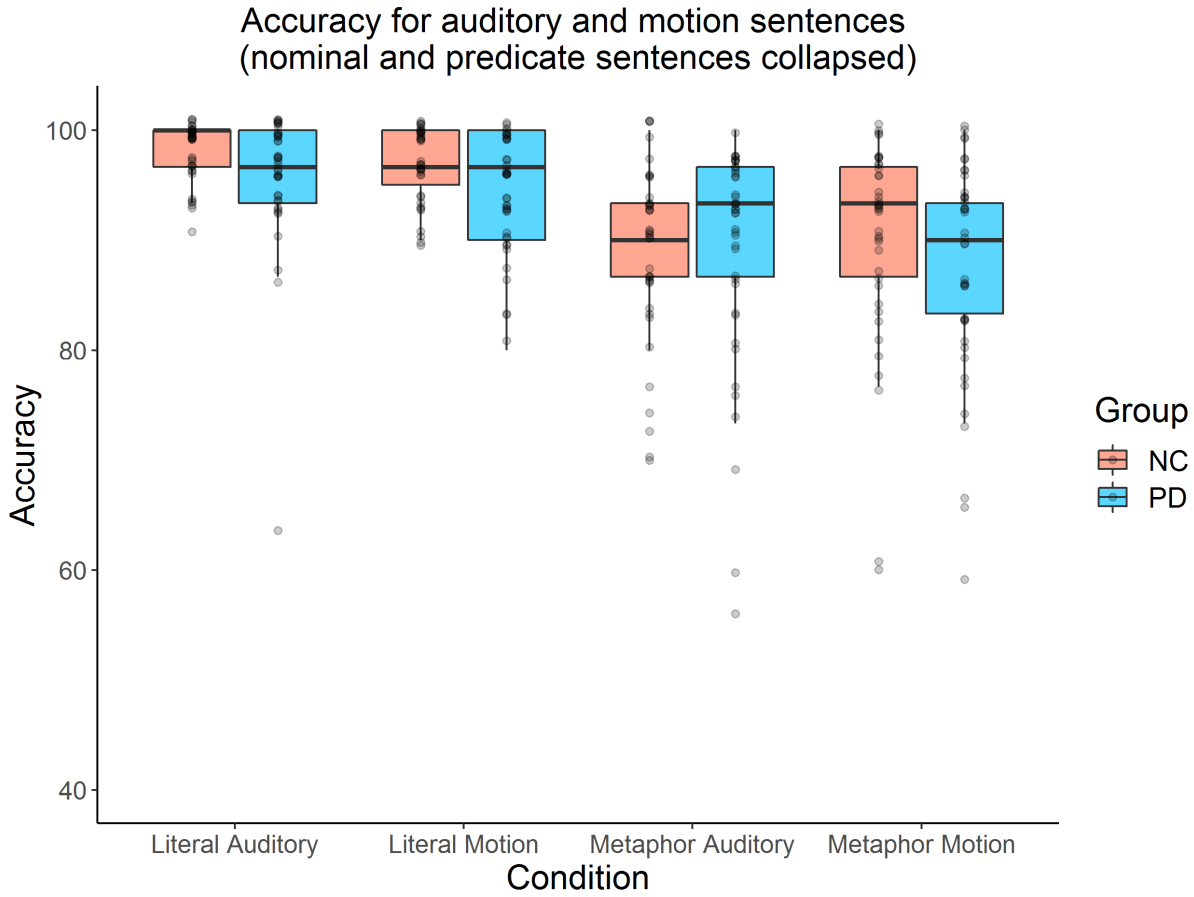


Figure 2. Boxplots of auditory and motion sentence accuracy with jittered individual data points.

For response time, the interaction BF_{10} was 1.33 under the wide prior, indicating that there was no evidence in either direction. When JASP's default prior was used, BF_{10} for the interaction was 2.48, providing slightly more evidence for the interaction but not strong enough to provide convincing support for H1. While the evidence in favour of the interaction was not strong, there was more evidence for H1 than for H0 so we ran follow up Bayesian paired-samples t-tests separately for each group according to our preregistered analysis plan. For PD patients, when the Cauchy prior width was set to 2 based on previous literature, BF_{10} was 1.68, indicating that there was not much evidence in either direction. Under JASP's default prior of .707, BF_{10} was 3.84, indicating that the evidence moderately favoured the alternative hypothesis. PD patients were slower to respond to Motion metaphors relative to Auditory metaphors, but the effect was not as large as suggested by previous studies. The response time data for Group and Modality are plotted in Figure 3.

Journal Pre-proof

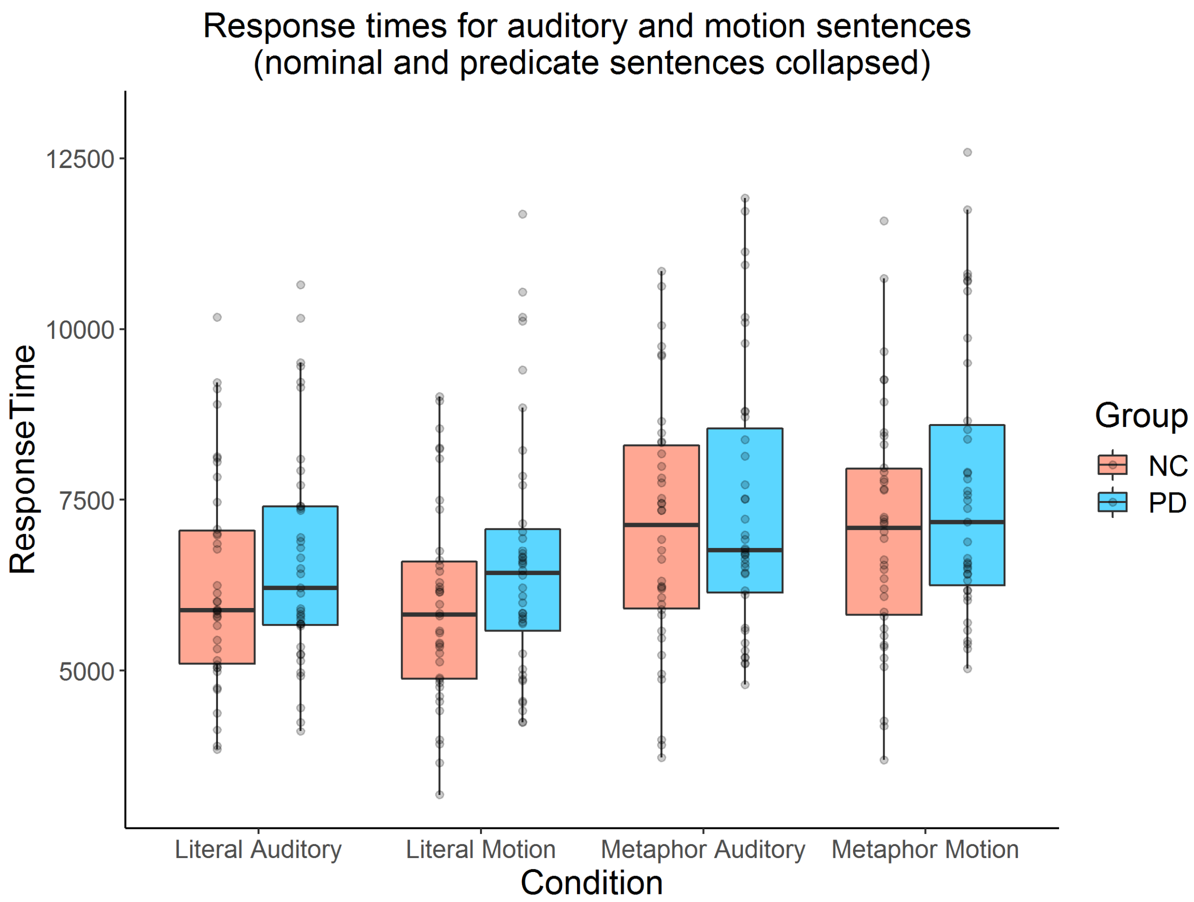
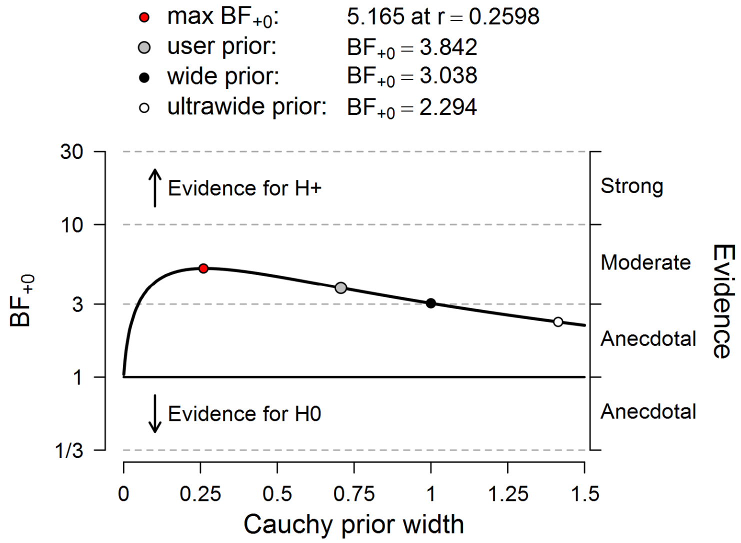


Figure 3. Boxplots of auditory and motion sentence RTs with jittered individual data points.

A Bayes factor robustness check produced in JASP (Figure 4) illustrates how evidence for H1 changed under different prior widths. Note that a narrow prior resulted in the strongest evidence for H1, suggesting that the effect size was much smaller than in previous studies.

A one-sided Bayesian paired-samples t-test was also run for the control group. When the Cauchy prior width was set to 2, BF_{10} was .047 indicating strong evidence in favour of the null hypothesis. Under JASP's default prior of .707, BF_{10} was .13, again indicating strong evidence in favour of the null hypothesis.

Parkinson's group, RT, Motion vs. Auditory metaphors.



Control group, RT, Motion vs. Auditory metaphors.

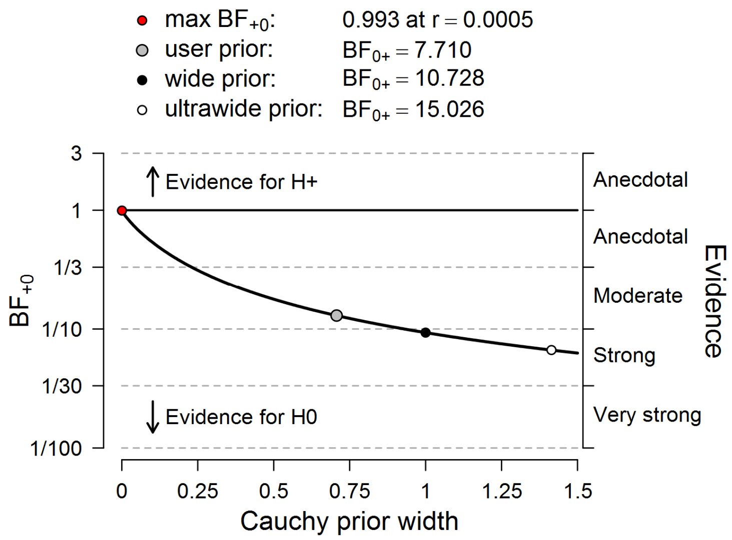


Figure 4. Bayes factor robustness checks for the paired-samples t-test comparing response times for Motion metaphors > Auditory metaphors.

3.2.2. Hypothesis 2: Literal Motor Concept

To examine the modality effect in the literal sentences, we again ran a repeated measures JZS Bayes factor ANOVA. As before, to test the interaction we compared a model which included the main effects of Group and Modality against a model which also included the interaction term.

For accuracy, under the wide prior (r scale fixed effects = 1), BF_{10} was .121, indicating strong evidence against the interaction. The data were 8.26 times more likely under the main effects model. Under the default prior, BF_{10} was .234, again indicating moderate evidence against the interaction (see Figure 2).

For response time, under the wide prior, the interaction BF_{10} was .161, indicating strong evidence against the interaction, and under the default prior BF_{10} was .32, indicating moderate evidence against the interaction (see Figure 3).

3.2.3. Hypothesis 3: Syntactic Structure

We tested whether there was an interaction between Group and the Syntactic structure of the metaphors (nominal and predicate). As we had matched the modality semantics of the sentences by using nominalized verbs as event nouns in the nominals, we predicted that there would not be a difference in performance on the two sentence types. As planned, default priors were used for these analyses.

For metaphor accuracy, BF_{10} for the interaction was .36 indicating moderate evidence against the interaction. We did not plan to test a main effect of syntactic structure; however, the model comparison revealed that the BF_{10} for this main effect was 222500000 – robust evidence for this effect (see Figure 5). Both groups were less accurate when responding to nominal metaphors compared to predicates. This finding was unexpected; we followed it up with additional exploratory analysis in section 3.3. below.

Similar results were obtained when we examined response times. BF_{10} for the interaction was .243, indicating moderate evidence against the interaction. An unanticipated main effect of Syntactic structure was again observed (BF_{10} = 1190000). Both groups were slower to respond to nominal metaphors than to predicates (see Figure 6).

Journal Pre-proof

Journal Pre-proof

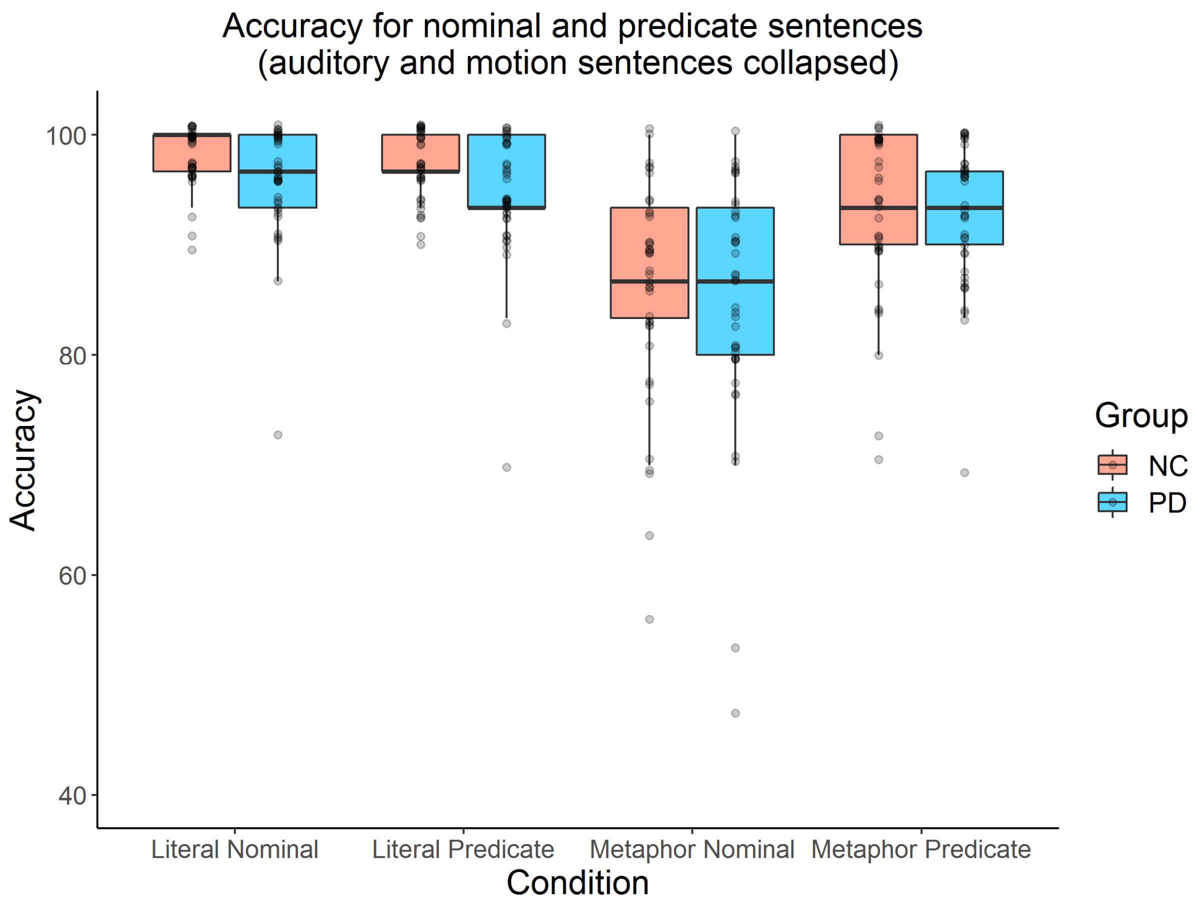


Figure 5. Boxplots of nominal and predicate sentence accuracy with jittered individual data points.

Journal

Journal Pre-proof

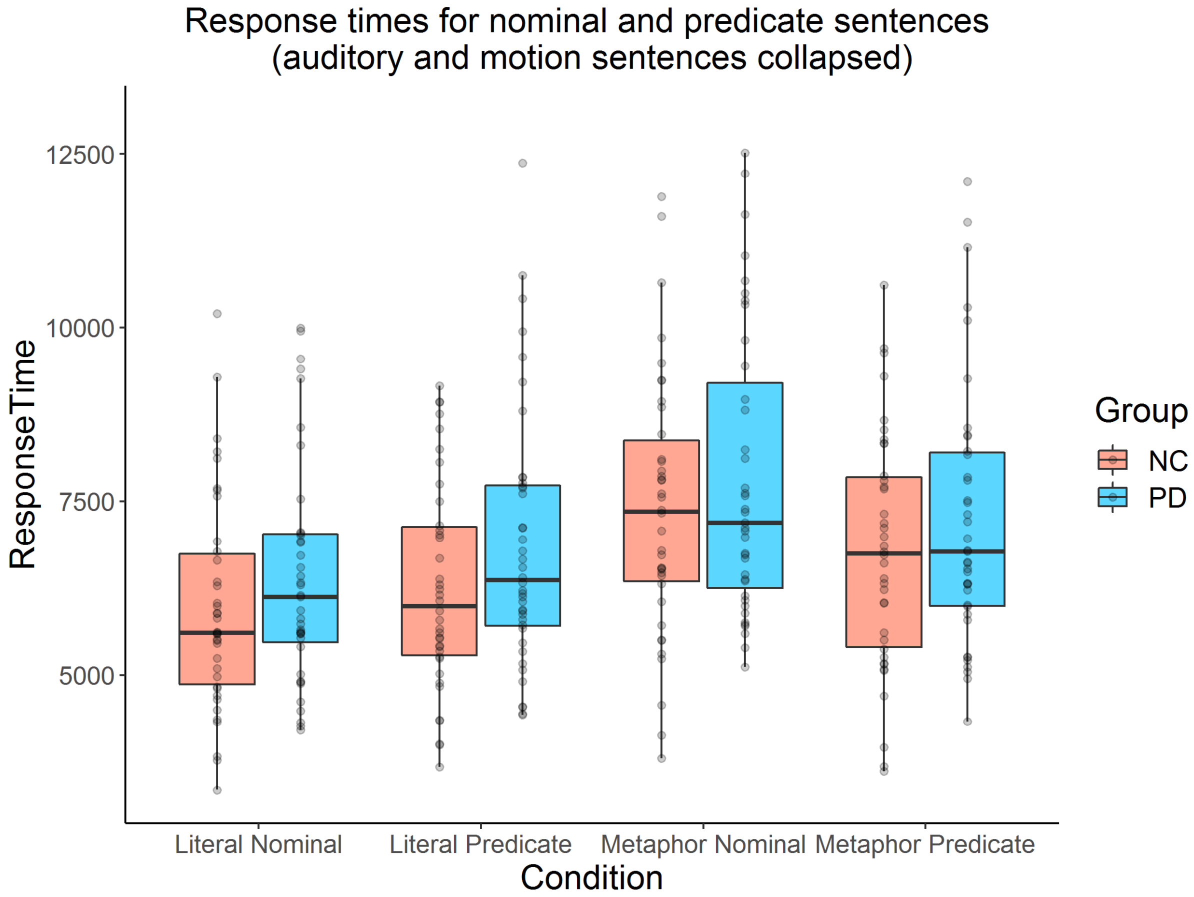


Figure 6. Boxplots of nominal and predicate sentence RTs with jittered individual data points.

3.2.4. Hypothesis 4: Figurativeness

A JZS Bayes factor repeated-measures ANOVA was used to test the interaction between Group and Figurativeness. As before, the model with the interaction term was compared against the model with only the main effects. For accuracy, BF_{10} for the interaction was .315, providing moderate evidence against the interaction. For response time, BF_{10} was .286, again indicating moderate evidence against the interaction. As illustrated in Figures 2, 3, 5, and 6, both groups responded less accurately to metaphors compared to literal sentences, and response times in both groups were longer for metaphors. The Bayes factors for the interaction terms indicate that the PD group did not demonstrate impaired metaphor comprehension relative to controls.

3.3. Exploratory Analyses

As described in section 3.2.3., we found an unexpected effect of the syntactic structure of the metaphor such that both groups found the nominal metaphors more difficult than predicate metaphors. We had planned only to examine the interaction between Group and Syntax in the

Journal Pre-proof

preregistered aims of this study, but the sheer largeness of the Bayes factors for the main effect of Syntax warranted closer inspection. We report here further exploration of performance in the different Syntax conditions but note that these analyses should be considered as hypothesis generating rather than hypothesis testing or confirming, as they were not planned at the outset.

Both groups found the nominal metaphors substantially more difficult to understand than the predicates, and this result was evident in both accuracy and response times. As reported above, we combined the nominal and predicate metaphors together in our preregistered tests of the interaction between group and modality. But if performance on the two metaphor types is so different, this unexpected difference may warrant examining the modality effect within each metaphor type separately. We report these exploratory analyses here.

3.3.1 Nominal Metaphors

For nominal metaphor accuracy, we ran a JZS Bayesian repeated-measures ANOVA, using default priors in JASP, to test the interaction between Group and Modality. BF_{10} was 1.3, suggesting that there was not much evidence in either direction (the data are equally likely under H1 and H0) (see Figure 7). For response time, BF_{10} for the interaction was .294, indicating moderate evidence against the interaction (see Figure 8).

3.3.2. Predicate Metaphors

For predicate metaphor accuracy, BF_{10} for the interaction between Group and Modality was .242, providing moderate evidence against the interaction. For response time, BF_{10} for the interaction was 4.46, providing moderate evidence for the presence of an interaction effect (the data were 4.46 times more likely under the model with the interaction term compared to the model with only the main effects) (see Figures 7 and 8). Follow-up paired-samples t-tests (one-sided) indicated that there was extreme evidence for H1 in the PD group (BF_{10} = 133.6), and moderate evidence for the null hypothesis in the control group (BF_{10} = .241). This result indicates that PD patients were significantly slower to respond to predicate motion metaphors than predicate sound metaphors, while response times in the control group were equal in the two conditions.

3.3.3. Standard Measures of Action Language Impairment in PD

Most previous studies reporting impaired action language processing in PD have used either verbal fluency measures (generating lists of verbs compared to other standard fluency tests) or picture naming tests (object vs action naming). To help situate the results of the current manuscript within the broader literature, we also collected these measures. Participants completed a phonetic fluency (F words) and verb fluency task, in which they produced as many words as they could think of in the given category for one minute. The Object and Action Naming Test was also administered. In this task, participants named line drawings of common objects (80) and actions (50) (Druks & Masterson, 2000). In cases where the testing session had exceeded two hours, it was not always possible to collect these measures in every subject. We collected phonetic fluency in 41 patients and 36 controls, verb fluency in 25 patients and 26 controls, and object and action naming in 24 patients and 24 controls. Note that these numbers are in line with or exceed sample sizes in most previous studies, with the norm being 15-25 participants in each group.

For verbal fluency, we examined the interaction between Group and Task (F fluency or verb fluency). Using default priors in JASP, we found that both groups generated more words on the verb fluency task than the phonetic fluency task (PD group BF_{10} = 3.5, Control group BF_{10} = 9.03), but BF_{10} for the interaction was .322, providing moderate evidence that there was no interaction between Group and Task (see Table 7). Independent samples Bayesian t-tests also showed that there was no difference between the groups for phonetic fluency (BF_{10} = .406) or verb fluency (BF_{10} = .304).

For naming, the BF_{10} for the interaction between Group and Task (object or action) was .945, indicating that there was no evidence in either direction. We cannot conclude that there is evidence either for or against the interaction. Independent samples Bayesian t-tests showed that there was moderate evidence for no group difference in object naming (BF_{10} = .33) and weak evidence for no group difference in action naming (BF_{10} = .65). Looking within each group, we found evidence for no difference in object and action naming performance in controls (BF_{10} = .217), while in PD patients there was no evidence in either direction (BF_{10} = 1.13)

Table 7. Means and standard deviations for verbal fluency and naming scores.

	Phonetic fluency	Verb fluency	Object naming %	Action naming %
PD	17.71 (4.92)	21.84 (5.51)	97.76 (2.05)	96.17 (4.08)
NC	18.97 (4.72)	22.65 (7.46)	97.5 (2.39)	97.5 (2.15)

Journal Pre-proof

Journal Pre-proof

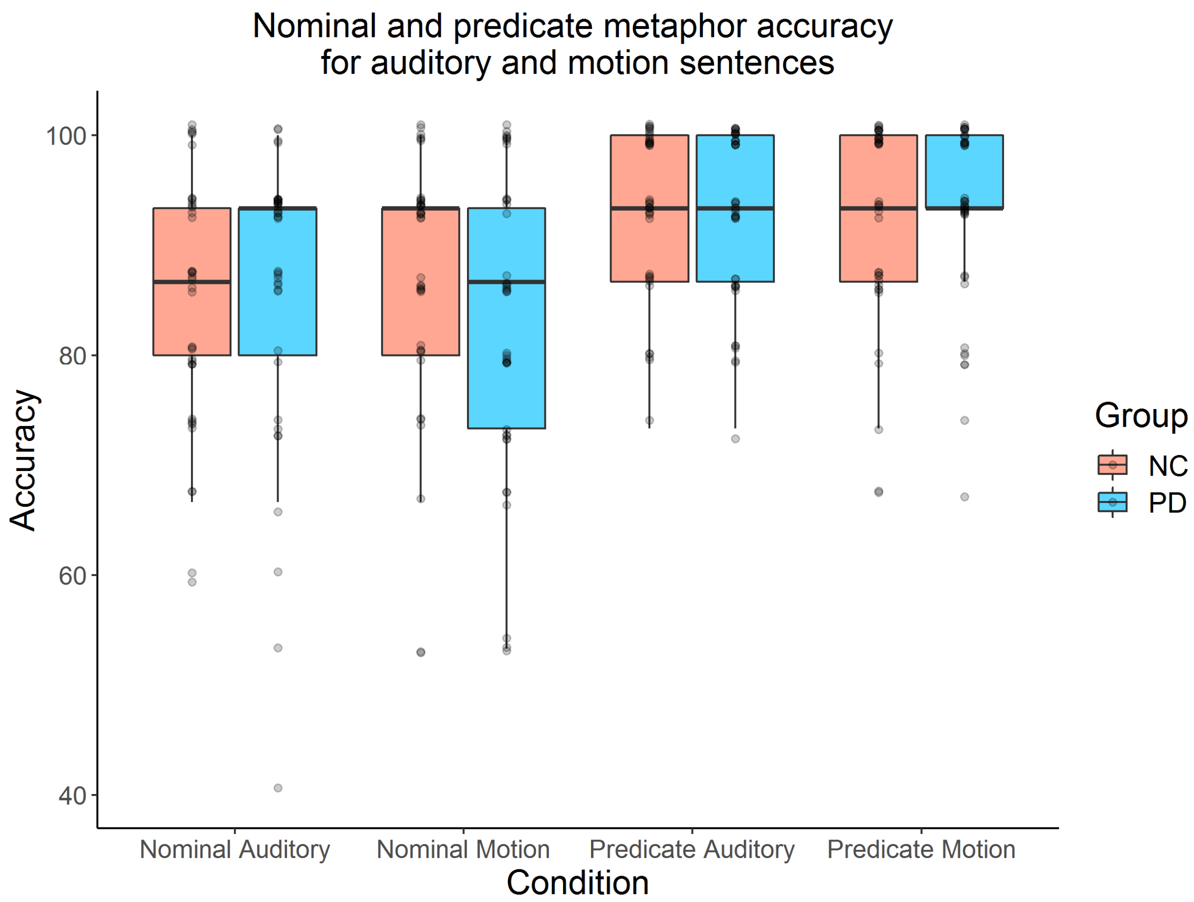


Figure 7. Boxplots of nominal and predicate & auditory and motion sentence accuracy with jittered individual data points.

Journal Pre-proof

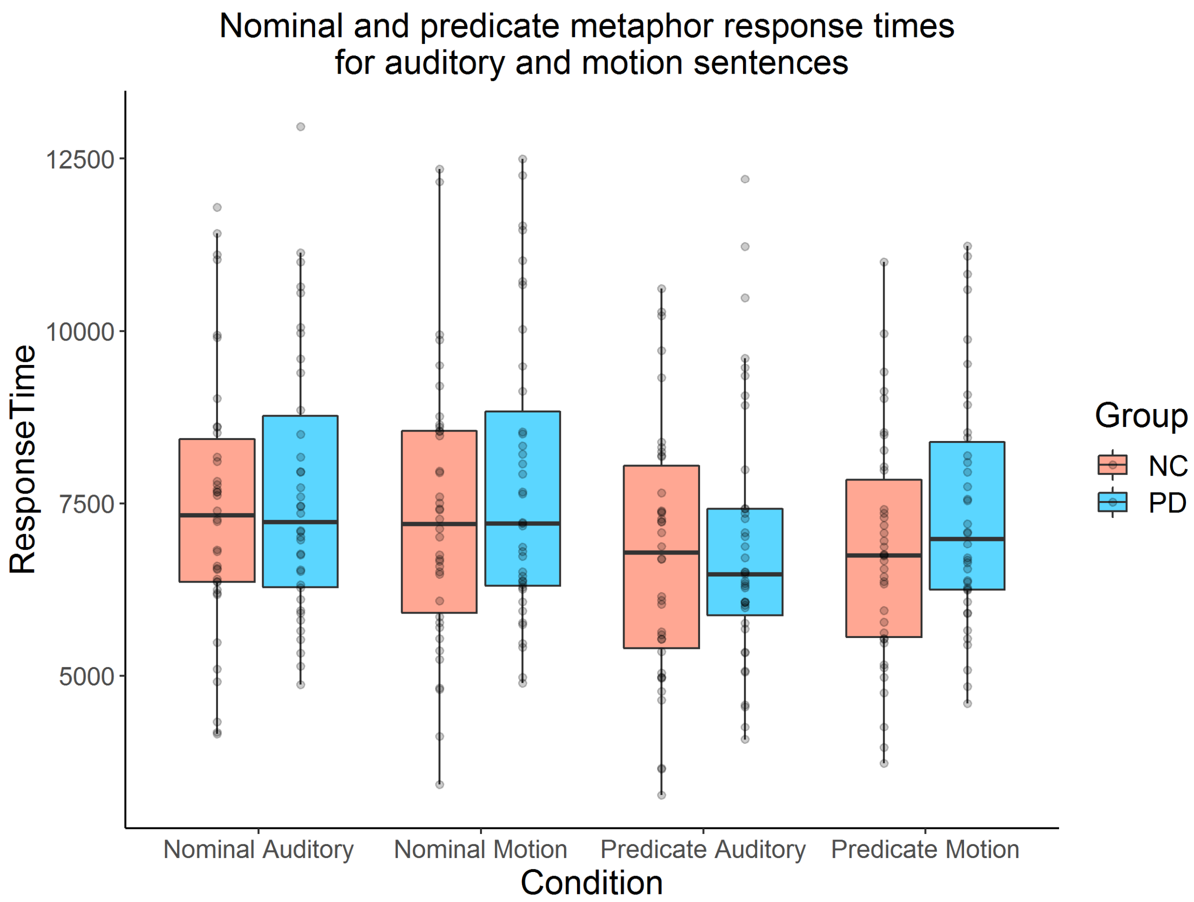


Figure 8. Boxplots of nominal and predicate & auditory and motion sentence RTs with jittered individual data points.

4. Discussion

This Registered Report tested the hypothesis that modality specific sensorimotor systems in the brain functionally contribute to the comprehension of language that refers to those modalities – even when words are used figuratively. Specifically, we examined whether an impaired motor system (in Parkinson's disease) results in a specific impairment in comprehending action metaphors relative to sound metaphors. We found that accuracy for action sentences compared to sound sentences was not impaired in PD; this observation was true for both literal and metaphorical sentences. For response times (RT), we found relatively weak evidence that PD patients were slower to respond to action metaphors than sound metaphors. Follow-up exploratory analyses suggest that this RT effect was driven by a strong effect in the predicate metaphors, while there appeared to be no effect on RTs for the nominal metaphors. There was no modality effect in PD patients' RTs to literal sentences. Contrary to our predictions, we found that the type of metaphor construction affected comprehension. Both groups were impaired in their comprehension of nominal metaphors, which involve a direct comparison between two nouns (though the metaphorical noun was a nominalized verb), compared to predicate metaphors

Journal Pre-proof

where verbs are used figuratively. In what follows, we situate these results within the broader literature and discuss the contribution this Registered Report makes to the embodiment debate.

4.1. Literal Sentences

PD patients were slower and less accurate than controls to respond to all literal sentences, but there was no interaction between Group and Modality. That is, PD patients were not specifically impaired in responding to action sentences compared to sound sentences. This finding contrasts with results from several studies reporting that PD patients show deficits in processing, comprehending, and producing action words or verbs. Two key differences between this study and previous studies account for this apparent discrepancy. First, the present study compared performance on sentences that use motion verbs to sentences that use sound verbs (nominalized to event nouns in the nominal sentences). In contrast, many previous studies have compared performance on verbs relative to nouns, or actions relative to objects (Bertella et al., 2001; Bocanegra et al., 2015; Boulenger et al., 2008; Cotelli et al., 2007; Herrera, Cuetos, et al., 2012; Péran et al., 2003; Piatt et al., 1999; Signorini & Volpato, 2006). However, impaired performance on verbs relative to nouns in PD does not necessarily indicate an embodiment effect. Verbs (or actions) can be more difficult to process than nouns (or objects) for reasons other than action semantics. For example, verbs are more polysemous than nouns. Of the 117097 nouns listed in WordNet (wordnet.princeton.edu), 13.4% are polysemous, and they have 1.23 senses on average. Of the 11488 verbs listed, 45.5% are polysemous, and they have 2.16 senses on average. The greater polysemy of verbs reflects that verbs are more flexible, abstract, and relational, and their meanings are more mutable depending on context (Gentner & France, 1988). The cognitive demands involved in processing nouns and verbs are therefore likely to be different. There is a closer one-to-one mapping between nouns and concepts, whereas verbs can describe many more ideas. In previous studies showing PD patients are more impaired on verbs than nouns, it is not possible to tell whether this disparity is caused by the patients' impaired motor systems, or simply because verbs are cognitively more complex.

In the present study we did not replicate previously reported action language processing impairments in PD using standard action naming and fluency tasks. These analyses were not part of our pre-registered analysis plan, but we ran them to help contextualize the findings of the metaphor task. For verb fluency, we found evidence that there was no effect, while for naming, the evidence did not convincingly favour either the null or the alternative hypothesis.

The confound between semantics and grammatical class has been noted in some previous studies. One attempt to resolve this confound has been to compare the performance of PD patients and controls on the processing of action verbs (e.g. to climb, to swim) relative to abstract verbs (e.g. to justify, to believe) (Fernandino et al., 2013b, 2013a). However, this comparison is also problematic because action and abstract verbs differ on other dimensions, such as concreteness, imageability, frequency, and age of acquisition. A general or nonspecific language impairment in PD could be interpreted as a specific action deficit simply because controls perform at ceiling for action conditions and worse or more variably on abstract conditions.

In the present study, the comparison between action verbs and sound verbs overcomes many shortcomings associated with previous noun-verb or action-abstract comparisons. The conditions

Journal Pre-proof

were matched in grammatical class, on psycholinguistic variables such as concreteness and imageability, and in their strong associations with sensory modalities. All of the motion verbs and event nouns were associated with visual motion (such as glide and roll) and most described bodily actions (such as jog, swim, cartwheel and chop). Some of the sound verbs referred to biological sounds (sneeze, slurp), and others to non-biological sounds (such as pop, sizzle). Neuroimaging studies have found that action and sound concepts are associated with functional activity in different regions of the brain. A somatotopic representation of action verbs has been found in primary sensorimotor areas (Carota, Moseley, & Pulvermüller, 2012; O Hauk et al., 2004; Kemmerer, Castillo, Talavage, Patterson, & Wiley, 2008; Tettamanti et al., 2005), but a higher level action association network involving lateral temporal areas such as posterior middle temporal gyrus is more commonly associated with action semantics (Binder, Desai, Graves, & Conant, 2009; Desai et al., 2009; Kable, Kan, Wilson, Thompson-Schill, & Chatterjee, 2005; Kable, Lease-Spellmeyer, & Chatterjee, 2002). Processing sound words is linked to activity in an auditory association region, the posterior superior temporal sulcus (Goldberg, Perfetti, & Schneider, 2006; Kellenbach, Brett, & Patterson, 2001; Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008; Kiefer et al., 2012), and a patient with a circumscribed lesion in this area had a specific deficit in processing sound words (Trumpp, Kliese, Hoenig, Haarmeier, & Kiefer, 2013). In addition, patients with the logopenic variant of primary progressive aphasia demonstrated grey matter atrophy in this auditory association area, which was directly correlated with their impaired comprehension of sound words (Bonner & Grossman, 2012). The fact that action and sound verbs are closely matched in many ways (grammatical class, concreteness, imageability, strong sensorimotor associations) and yet are processed in different brain regions make them an ideal test case for questions about embodiment. It is thus noteworthy that PD patients in this study were not impaired in processing literal sentence constructions using action verbs relative to sound verbs.

A second major difference between this study and most previous studies of action language in PD is that the present study used sentence stimuli where other studies have mostly used single word stimuli. By some accounts of semantics, a concept is never "static" or context-free (e.g., Yee & Thompson-Schill, 2016). In a sentence, the context constrains and makes salient the features and meaning(s) of individual words. But when single words are processed, their "meanings" may be more variable because contexts vary across subjects. Regardless, action sentences have been found to activate motor regions in the same way as individual action words (Raposo, Moss, Stamatakis, & Tyler, 2009b; Tettamanti et al., 2005). So, action semantics are not inherently diluted by sentences. If motor systems functionally contribute to the comprehension of action concepts, one might expect even greater involvement of sensorimotor areas when a sentence context constrains the word's meaning to its motor features. Why then should PD patients be impaired in processing single action words but not action sentences? As outlined above, some previous studies of PD action-language processing may have suffered from compromising confounds.

4.2. Metaphorical sentences

Whilst there was no Group by Modality interaction in literal sentences, we did find evidence of an interaction effect in metaphorical sentences. This effect provides partial support for our hypotheses but must be interpreted cautiously. First, the effect was present in RTs but not accuracy. The lack of difference in accuracy was not because of ceiling effects, as both groups responded to metaphorical sentences less accurately. Secondly, the evidence for a Group by Modality RT interaction was weak when we considered both types of sentence construction together (nominal and predicate sentences). A preregistered analysis examined whether the processing of nominal and predicate sentences differed, but we expected any modality interaction effects to be independent of sentence construction. Since both sentence types used action and sound words either as verbs or as event-nouns, the sensorimotor semantics were similar in each type. While we did not find an interaction between group and sentence type, we did find a large main effect of sentence type, such that both groups performed worse on the nominal metaphors relative to the predicates. Given that the cognitive demands in processing these two sentence types seemed to differ, we conducted additional exploratory analyses where we tested the interaction between group and modality separately in each type of metaphor. The results suggest that PD patients' responses to predicate-action metaphors may be slowed (but not less accurate) relative to predicate-sound metaphors, while no analogous effects were found for the nominal metaphors. However, these exploratory analyses were not preregistered and thus should be considered hypothesis generating rather than confirming.

These exploratory results generate two questions: First, why is there a group by modality interaction in RTs to metaphors but not literal sentences? And second, why should this effect be present in predicate but not nominal metaphors? Below, we speculate about possible explanations that could be directly tested in future confirmatory studies.

When words are used figuratively they take on a more abstracted sense, shedding some of their (typically concrete) features in the process. For this reason, theories of embodied cognition may predict that sensorimotor systems are less involved in comprehension when figurative extensions of words result in a loss of their sensorimotor features. As outlined in the Introduction, some studies report that sensorimotor metaphors activated sensorimotor brain regions in the same way as literal sentences (Boulenger et al., 2009; Boulenger et al., 2012; Citron & Goldberg, 2014, Lacey et al., 2017), while other studies observed sensorimotor activity in response to literal sentences but not metaphors (Aziz-Zadeh et al., 2006; Chen et al., 2008; Raposo et al., 2009; Ruschemeyer et al., 2007). We have argued that this apparent discrepancy is because of differences in the familiarity of the metaphor stimuli used in each study. When metaphors are highly familiar, their meanings become lexicalized and can be understood without reference to the literal features of the word. However, when metaphors are highly novel, as in the present study, we have argued that people may need to engage more explicitly with the figurative word in order to resolve the metaphor. Doing so may involve activating sensorimotor features of the word, even if only to later inhibit those features as irrelevant to the new figurative sense (Jamrozik et al., 2016).

Some theories of language and conceptual processing propose that two systems interact in the representation of knowledge: a statistical, distributional, linguistic information system, and a situated, modal, simulation system (Barsalou et al., 2008; Lynott & Connell, 2010). The linguistic system provides a shortcut to comprehension by using knowledge of the statistical patterns of language and natural co-occurrences of words. The literal sentences in this study described familiar situations and used words that are highly likely to co-occur together in natural language (e.g. “The uncle sang to the baby”, “The fisherman reeled in a bass”). When

Journal Pre-proof

Journal Pre-proof

participants comprehended these literal sentences, they may have used statistical linguistic knowledge to arrive at meaning, without needing to activate or simulate motor features of the verbs. In contrast, novel metaphors use combinations of agents and actions that are unlikely to co-occur frequently in natural language (e.g. “The sunset sang to the lovers”, “The colonel reeled in the officers”), and statistical knowledge might not provide enough information for people to resolve the meanings of these sentences. Novel metaphors may therefore *require* deeper engagement and simulation of word features, even if some of those features are later deemed irrelevant to meaning. These two routes to comprehension may explain why we observed slowed responses to action metaphors but not literal action sentences in PD.

Yet this explanation does not account for why PD patients responded more slowly to predicate action metaphors and not to nominal action metaphors. Both types of sentences used similar action and sound semantics, so embodied theories would predict that PD patients would be impaired in processing action sentences in both conditions. However, both groups of participants found the nominal metaphors more difficult to comprehend, with greater variability, lower accuracy, and slower responses to nominal metaphors in all cases. This finding is corroborated by a recent study of focal lesion patients that, despite careful matching between metaphor conditions, also found lower accuracy for nominal metaphors relative to predicate metaphors (Cardillo et al., 2018). Consequently, even if a modality effect did exist it would be harder to detect because of the increased noise in the nominal conditions. Nominal metaphors may be harder to comprehend because they entail a categorical assertion, an inherently more abstract construction. Solving the puzzle of how two apparently dissimilar concepts are in fact alike requires conscious thought. On the other hand, figurative uses of verbs in predicate metaphors may be easier to comprehend because verbs are already more polysemous and flexible than nouns. For example, Gentner & France (1988) demonstrated what they call the “verb mutability effect”: when verbs are combined with nouns that strain the semantics of each (e.g. “The lizard worshipped the sun.”), people naturally adjust the meaning of the verb more than the noun. Novel predicate metaphors may therefore be a good candidate for tests of embodied cognition, in that 1) their novelty means that people cannot rely on statistical language knowledge to resolve them, and 2) their relative ease of comprehension means that responses to them are not so noisy that it becomes difficult to detect group differences.

4.3. Parkinson’s Disease, Statistics, and Embodied Cognition

While we offer some plausible explanations for why the comprehension of motor language in PD could be impaired in some cases but not others, it is worth considering the implications of these results for embodied cognition theories. Others have already proposed a “weak” version of the embodied cognition hypothesis, where sensorimotor systems are not necessarily always involved in sensorimotor representations. Under this view, metaphors are one proposed route by which we develop abstract representations that are initially derived from embodied experiences (Jamrozik et al., 2016). However, in the context of PD, the circumstances under which an embodiment effect is observed appear to be limited. In this study we observed an absence of any impairment in standard measures of verb fluency and action naming in PD (and argue that studies that did observe these effects cannot rule out alternative explanations). We also observed an absence of a modality effect in PD in response to literal sentences, and in accuracy to metaphorical sentences. Only one of the many tests we conducted yielded the hypothesized interaction effect, and even

Journal Pre-proof

Journal Pre-proof

that effect is qualified by the fact that it appears only observable in certain kinds of metaphor. Patient studies should provide the strongest tests of embodied cognition theories, but perhaps Parkinson's disease is not the ideal patient group for these tests. While PD indeed has devastating effects on motor function, it also has other effects on the brain, including widespread atrophy and frontostriatal dopamine depletion resulting in general cognitive impairment. It is therefore difficult to separate out the effect of an impaired motor system on conceptual representations in PD from other structural and functional brain changes taking place at the same time.

Previous studies of action language in PD generally use two types of statistical test to conclude that embodiment effects occur in PD. Some studies specifically test the interaction between group and condition, to demonstrate that the magnitude of the difference between patients and controls on some action condition is larger than in some non-action condition. Other studies have conducted two between-groups t-tests to show that PD patients and controls differ significantly in an action condition, and not in a non-action condition, or two within-groups tests to show that the PD group differs in their performance on the two conditions, while the control group does not. The conclusions derived from these results often depend on demonstrating “evidence of absence” in control conditions, e.g.:

1. There is a significant difference between PD patients and controls in processing action words, but there is *no difference between the groups* in processing abstract words.
    a. PD action vs. Control action = different.
    b. PD abstract vs. Control abstract = not different.
2. PD patients performed significantly worse in the action condition relative to the object condition, while controls showed *no difference* in performance between the two conditions.
    a. PD action vs. PD object = different.
    b. Control action vs. Control object = not different.

In null hypothesis significant testing (NHST), statistical non-significance indicates only that there is an “absence of evidence” and cannot provide “evidence of absence”. Yet the ability to demonstrate evidence of *no effect* in control subjects or control conditions is often critical to whether or not one can make claims about embodiment. Consider an example where PD patients and controls completed some language task on object words and action words. Both groups had lower scores in the action condition compared to the object conditions, but the difference between the object and action scores was significant only for the patients and not the controls. Under a NHST framework it is not possible to conclude that there is truly no difference between the conditions in the controls. The action condition may have been harder than the object condition for both groups, and the difference was magnified in the patients because their cognition is generally worse. Tests of interactions between groups and conditions are common in PD embodiment studies, but the strongest test arguably hinges on demonstrating both a) that a difference exists between conditions in the patients, and b) that no difference exists between conditions in controls. Failing to demonstrate b) leaves open the possibility that a difference exists in both groups and does not provide convincing evidence for embodied cognition.

Journal Pre-proof

The present study used Bayesian hypothesis testing, where Bayes factors provide a measure of the relative evidence in the data for either the null or the alternative hypothesis. Bayes factors allow us to demonstrate "evidence of absence" where frequentist statistics are unable to do so. Future studies in embodied cognition should ensure that they are able to provide evidence for no effect in control conditions where this observation is critical to their conclusions about the presence of embodiment effects.

A final point on statistics concerns the sample size used in this and previous studies. We report data collected from 41 PD patients and 39 age-matched controls. This number is by no means a large sample when compared to other studies in psychology, and some embodiment effects may exist that are smaller than what we could detect with this number of subjects. Given the difficulty in recruiting PD patients with motor disturbance yet relatively normal cognition, the fact that we did not detect significant effects in most of the reported tests here nevertheless makes an important point for the field. Our sample contains at least twice as many subjects as those reported in previous PD embodiment studies and would be considered large for a patient study. It is likely infeasible for researchers to collect samples of 80 or 100+ cognitively normal PD patients to test for the presence of smaller effects, unless multiple institutes pool resources.

4.4. Conclusions

This Registered Report study represents one of the most tightly controlled tests of literal and figurative action language embodiment in Parkinson's disease to date. The conditions tested were matched extensively in ways that overcome shortcomings of previous studies, the methods and analyses were preregistered, and Bayesian analytic methods were used to provide meaningful evidence for the *absence* of effects, which was not possible in previous studies. We found evidence *against* an embodiment effect in PD patients' comprehension of literal language about action concepts. At the same time, we found evidence that responses to predicate action metaphors were slowed in these patients, which may be because comprehending novel metaphors is more likely to require deeper engagement with the sensorimotor features of words. We suggest limits in the use of PD as a population from which to test embodiment hypotheses and offer some suggestions for how future research in this area could provide more convincing evidence for and against embodied cognition.

Acknowledgements:

This research was supported by NIH grant 5R01DC012511 awarded to AC.
We thank Zoltan Dienes for his invaluable guidance on Bayesian inferential statistics.

Journal Pre-proof

Journal Pre-proof

References

Abrevaya, S., Sedeño, L., Fitipaldi, S., Pineda, D., Lopera, F., Buritica, O., … García, A. M. (2017). The Road Less Traveled: Alternative Pathways for Action-Verb Processing in Parkinson's Disease. *Journal of Alzheimer's Disease*, *xx*, 1429–1435. https://doi.org/10.3233/JAD-160737

Armstrong, B. C., Watson, C. E., & Plaut, D. C. (2012). SOS! An algorithm and software for the stochastic optimization of stimuli. *Behavior Research Methods*, *44*(3), 675–705. https://doi.org/10.3758/s13428-011-0182-9

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent Embodied Representations for Visually Presented Actions and Linguistic Phrases Describing Actions. *Current Biology*, *16*(18), 1818–1823. https://doi.org/10.1016/j.cub.2006.07.060

Bak, T. H. (2013). The neuroscience of action semantics in neurodegenerative brain diseases. *Current Opinion in Neurology*, *26*(6), 671–677. https://doi.org/10.1097/WCO.0000000000000039

Bak, T. H., & Hodges, J. R. (2004). The effects of motor neurone disease on language: Further evidence. *Brain and Language*, *89*(2), 354–361. https://doi.org/10.1016/S0093-934X(03)00357-2

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). *Handbook of Psycholinguistics*, *2*, 285–375.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *25*, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(August), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. *Symbols, Embodiment, and Meaning*, 245–283.

Berg, E., Björnram, C., Hartelius, L., Laakso, K., & Johnels, B. (2003). High-level language difficulties in Parkinson's disease. *Clinical Linguistics & Phonetics*, *17*(1), 63–80. https://doi.org/10.1080/0269920021000055540

Bertella, L., Albani, G., Greco, E., Priano, L., Mauro, A., Marchi, S., … Semenza, C. (2001). Noun verb dissociation in Parkinson's disease. *Brain and Cognition*, *48*(2–3), 277–280.

Bezeau, S., & Graves, R. (2001). Statistical Power and Effect Sizes of Clinical Neuropsychology Research. *Journal of Clinical and Experimental Neuropsychology*, *23*(3), 399–406. https://doi.org/10.1076/jcen.23.3.399.1181

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.

Bocanegra, Y., García, A. M., Lopera, F., Pineda, D., Baena, A., Ospina, P., … Cuetos, F. (2017). Unspeakable motion: Selective action-verb impairments in Parkinson's disease patients without mild cognitive impairment. *Brain and Language*, *168*, 37–46. https://doi.org/10.1016/j.bandl.2017.01.005

Bocanegra, Y., García, A. M., Pineda, D., Buriticá, O., Villegas, A., Lopera, F., … Ibáñez, A.

Journal Pre-proof

(2015). Syntax, action verbs, action semantics, and object semantics in Parkinson's disease: Dissociability, progression, and executive influences. *Cortex*, *69*, 237–254. https://doi.org/http://dx.doi.org/10.1016/j.cortex.2015.05.022

Bonner, M. F., & Grossman, M. (2012). Gray matter density of auditory association cortex relates to knowledge of sound concepts in primary progressive aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991.

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*(8), 1905–1914. https://doi.org/10.1093/cercor/bhn217

Boulenger, V., Mechtouff, L., Thobois, S., Broussolle, E., Jeannerod, M., & Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, *46*(2), 743–756. https://doi.org/10.1016/j.neuropsychologia.2007.10.007

Boulenger, V., Shtyrov, Y., & Pulvermüller, F. (2012). When do you grasp the idea? MEG evidence for instantaneous idiom understanding. *Neuroimage*, *59*(4), 3502–3513.

Bowdle, B., & Gentner, D. (1999). Metaphor comprehension: From comparison to categorization. *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*, *21*, 90. Lawrence Erlbaum Associates.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Cardillo, E. R., McQuire, M., & Chatterjee, A. (2018). Selective Metaphor Impairments After Left, Not Right, Hemisphere Injury. *Frontiers in Psychology*, *9*, 2308. https://doi.org/10.3389/fpsyg.2018.02308

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*(3), 651–664. https://doi.org/10.3758/BRM.42.3.651

Cardillo, E. R., Watson, C., & Chatterjee, A. (2016). Stimulus needs are a moving target: 240 additional matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*. https://doi.org/10.3758/s13428-016-0717-1

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage*, *59*(4), 3212–3221. https://doi.org/10.1016/j.neuroimage.2011.11.079

Cardona, J. F., Gershanik, O., Gelormini-Lezama, C., Houck, A. L., Cardona, S., Kargieman, L., … Ibáñez, A. (2013). Action-verb processing in Parkinson's disease: new pathways for motor–language coupling. *Brain Structure and Function*, *218*(6), 1355–1373. https://doi.org/10.1007/s00429-013-0510-1

Carota, F., Moseley, R., & Pulvermüller, F. (2012). Body-part-specific representations of semantic noun categories. *Journal of Cognitive Neuroscience*, *24*(6), 1492–1509.

Chatterjee, A. (2010). Disembodying cognition. *Language and Cognition*, *2*(1), 79–116. https://doi.org/10.1515/LANGCOG.2010.004

Journal Pre-proof

Chen, E., Widick, P., & Chatterjee, A. (2008). Functional-anatomical organization of predicate metaphor processing. *Brain and Language*, *107*(3), 194–202. https://doi.org/10.1016/j.bandl.2008.06.007

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience*.

Cleary, R. A., Poliakoff, E., Galpin, A., Dick, J. P. R., & Holler, J. (2011). An investigation of co-speech gesture production during action description in Parkinson's disease. *Parkinsonism & Related Disorders*, *17*(10), 753–756. https://doi.org/10.1016/j.parkreldis.2011.08.001

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*(3), 542–550.

Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., & Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage*, *22*(1), 11–21. https://doi.org/10.1016/j.neuroimage.2004.01.001

Cotelli, M., Borroni, B., Manenti, R., Zanetti, M., Arévalo, A., Cappa, S. F., & Padovani, A. (2007). Action and object naming in Parkinson's disease without dementia. *European Journal of Neurology*, *14*(6), 632–637.

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., & Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. *Journal of Cognitive Neuroscience*, *23*(9), 2376–2386. https://doi.org/10.1162/jocn.2010.21596

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M. S. (2009). Activation of sensory–motor areas in sentence comprehension. *Cerebral Cortex*, bhp115.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412–431. https://doi.org/10.1016/j.cognition.2008.11.016

Druks, J., & Masterson, J. (2000). *Object and action naming battery*. Psychology Press.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, Vol. 63, pp. 591–601. https://doi.org/10.1037/0003-066X.63.7.591

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013a). Parkinson's disease disrupts both automatic and controlled processing of action verbs. *Brain and Language*, *127*(1), 65–74. https://doi.org/10.1016/j.bandl.2012.07.008

Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., & Desai, R. H. (2013b). Where is the action? Action sentence processing in Parkinson's disease. *Neuropsychologia*, *51*(8), 1510–1517. https://doi.org/10.1016/j.neuropsychologia.2013.04.008

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & Von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, *16*(12), 1709–1717.

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3–4), 455–479. https://doi.org/10.1080/02643290442000310

García, A. M., Bocanegra, Y., Herrera, E., Pino, M., Muñoz, E., Sedeño, L., & Ibáñez, A. (2017). Action-semantic and syntactic deficits in subjects at risk for Huntington's disease. *Journal of Neuropsychology*. https://doi.org/10.1111/jnp.12120

Garcia, A. M., Carrillo, F., Orozco-Arroyave, J. R., Trujillo, N., Vargas Bonilla, J. F., Fittipaldi,

Journal Pre-proof

S., … Cecchi, G. A. (2016). How language flows when movements don’t: An automated analysis of spontaneous discourse in Parkinson’s disease. *Brain and Language*, *162*, 19–28. https://doi.org/10.1016/j.bandl.2016.07.008

Gentner, D., Bowdle, B., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. *The Analogical Mind: Perspectives from Cognitive Science*, 199–253.

Gentner, D., & France, I. M. (1988). The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In *Lexical ambiguity resolution* (pp. 343–382). Elsevier.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.

Gibbs, R. W. (2005). *Embodiment and cognitive science*. Cambridge University Press.

Gibbs, R. W. (2006). Metaphor interpretation as embodied simulation. *Mind and Language*, *21*(3), 434–458. https://doi.org/10.1111/j.1468-0017.2006.00285.x

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *Journal of Neuroscience*, *26*(18), 4917–4921.

Hauk, O, Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representations of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Hauk, Olaf, Davis, M. H., & Pulvermüller, F. (2008). Modulation of brain activity by multiple lexical and word form variables in visual word recognition: A parametric fMRI study. *NeuroImage*, *42*(3), 1185–1195. https://doi.org/10.1016/j.neuroimage.2008.05.054

Herrera, E., & Cuetos, F. (2012). Action naming in Parkinson’s disease patients on/off dopamine. *Neuroscience Letters*, *513*(2), 219–222. https://doi.org/10.1016/j.neulet.2012.02.045

Herrera, E., Cuetos, F., & Ribacoba, R. (2012). Verbal fluency in Parkinson’s disease patients on/off dopamine medication. *Neuropsychologia*, *50*(14), 3636–3640. https://doi.org/10.1016/j.neuropsychologia.2012.09.016

Herrera, E., Rodríguez-Ferreiro, J., & Cuetos, F. (2012). The effect of motion content in action naming by Parkinson’s disease patients. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(7), 900–904. https://doi.org/10.1016/j.cortex.2010.12.007

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson’s disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*(3), 181–184. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1014720/

Humphries, S., Holler, J., Crawford, T. J., Herrera, E., & Poliakoff, E. (2016). A third-person perspective on co-speech action gestures in Parkinson’s disease. *Cortex, 78*, 44–54. https://doi.org/10.1016/j.cortex.2016.02.009

Ianni, G. R., Cardillo, E. R., McQuire, M., & Chatterjee, A. (2014). Flying under the radar: figurative language impairments in focal lesion patients. *Frontiers in Human Neuroscience*, *8*(November), 871. https://doi.org/10.3389/fnhum.2014.00871

Ibáñez, A., Cardona, J. F., Dos Santos, Y. V., Blenkmann, A., Aravena, P., Roca, M., … Bekinschtein, T. (2013). Motor-language coupling: Direct evidence from early Parkinson’s disease and intracranial cortical recordings. *Cortex*, *49*(4), 968–984. https://doi.org/10.1016/j.cortex.2012.02.014

Jamrozik, A., Mcquire, M., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor : Bridging embodiment to abstraction. *Psychonomic Bulletin & Review*, *23*(4), 1080–1089.

https://doi.org/10.3758/s13423-015-0861-0
Just, M. A., Newman, S. D., Keller, T. A., McEleney, A., & Carpenter, P. A. (2004). Imagery in sentence comprehension: an fMRI study. *NeuroImage*, *21*(1), 112–124. https://doi.org/10.1016/j.neuroimage.2003.08.042
Kable, J. W., Kan, I. P., Wilson, A., Thompson-Schill, S. L., & Chatterjee, A. (2005). Conceptual representations of action in the lateral temporal cortex. *Journal of Cognitive Neuroscience*, *17*, 1855–1870. https://doi.org/10.1162/089892905775008625
Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience*, *14*, 795–805. https://doi.org/10.1162/08989290260138681
Kargieman, L., Herrera, E., Baez, S., García, A. M., Dottori, M., Gelormini, C., … Ibáñez, A. (2014). Motor–Language Coupling in Huntington's Disease Families. *Frontiers in Aging Neuroscience*, *6*, 122. https://doi.org/10.3389/fnagi.2014.00122
Kellenbach, M. L., Brett, M., & Patterson, K. (2001). Large, colorful, or noisy? Attribute-and modality-specific activations during retrieval of perceptual attribute knowledge. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(3), 207–221.
Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: evidence from fMRI. *Brain and Language*, *107*(1), 16–43.
Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, *28*(47), 12224–12230.
Kiefer, M., Trumpp, N., Herrnberger, B., Sim, E.-J., Hoenig, K., & Pulvermüller, F. (2012). Dissociating the representation of action-and sound-related concepts in middle temporal cortex. *Brain and Language*, *122*(2), 120–125.
Lacey, S., Stilla, R., Deshpande, G., Zhao, S., Stephens, C., McCormick, K., … Sathian, K. (2017). Engagement of the left extrastriate body area during body-part metaphor comprehension. *Brain and Language*, *166*, 1–18. https://doi.org/10.1016/j.bandl.2016.11.004
Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic books.
Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.
Lewis, F. M., Lapointe, L. L., Murdoch, B. E., & Chenery, H. J. (1998). Language impairment in Parkinson's disease. *Aphasiology*, *12*(3), 193–206. https://doi.org/10.1080/02687039808249446
Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, *1*.
Monetta, L., & Pell, M. D. (2007). Effects of verbal working memory deficits on metaphor comprehension in patients with Parkinson's disease. *Brain and Language*, *101*(1), 80–89. https://doi.org/http://doi.org/10.1016/j.bandl.2006.06.007
Morey, R. D., & Rouder, J. N. (2015). *BayesFactor 0.9. 12-2. Comprehensive R Archive Network*.
Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126.
Péran, P., Cardebat, D., Cherubini, A., Piras, F., Luccichenti, G., Peppe, A., … Sabatini, U. (2009). Object naming and action-verb generation in Parkinson's disease: A fMRI study. *Cortex*, *45*(8), 960–971. https://doi.org/10.1016/j.cortex.2009.02.019

Journal Pre-proof

Péran, P., Rascol, O., Démonet, J., Celsis, P., Nespoulous, J., Dubois, B., & Cardebat, D. (2003). Deficit of verb generation in nondemented patients with Parkinson's disease. *Movement Disorders*, *18*(2), 150–156.

Piatt, A. L., Fields, J. A., Paolo, A. M., Koller, W. C., & Tröster, A. I. (1999). Lexical, semantic, and action verbal fluency in Parkinson's disease with and without dementia. *Journal of Clinical and Experimental Neuropsychology*, *21*(4), 435–443.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009a). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396. https://doi.org/10.1016/j.neuropsychologia.2008.09.017

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009b). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396. https://doi.org/10.1016/j.neuropsychologia.2008.09.017

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/J.JMP.2012.08.001

Rüschemeyer, S.-A., Brass, M., & Friederici, A. D. (2007). Comprehending Prehending: Neural Correlates of Processing Verbs with Motor Stems. *Journal of Cognitive Neuroscience*, *19*(5), 855–865. https://doi.org/10.1162/jocn.2007.19.5.855

Schmidt, G. L., Kranjec, A., Cardillo, E. R., & Chatterjee, A. (2010). Beyond laterality: a critical assessment of research on the neural basis of metaphor. *Journal of the International Neuropsychological Society*, *16*(01), 1–5. https://doi.org/10.1017/S1355617709990543

Scorolli, C., & Borghi, A. M. (2007). Sentence comprehension and action: Effector specific modulation of the motor system. *Brain Research*, *1130*, 119–124. https://doi.org/10.1016/j.brainres.2006.10.033

Signorini, M., & Volpato, C. (2006). Action fluency in Parkinson's disease: a follow-up study. *Movement Disorders : Official Journal of the Movement Disorder Society*, *21*(4), 467–472. https://doi.org/10.1002/mds.20718

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., … Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281.

Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., & Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders*, *25*(15), 2649–2653. https://doi.org/10.1002/mds.23429

Trumpp, N. M., Kliese, D., Hoenig, K., Haarmeier, T., & Kiefer, M. (2013). Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex*, *49*(2), 474–486.

Watson, C. E., Cardillo, E. R., Bromberger, B., & Chatterjee, A. (2014). The specificity of action knowledge in sensory and motor systems. *Frontiers in Psychology*, *5*(MAY), 494. https://doi.org/10.3389/fpsyg.2014.00494

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action concepts in the brain: an activation likelihood estimation meta-analysis. *Journal of Cognitive Neuroscience*, *25*(8), 1191–1205.

Yang, F. G., Edens, J., Simpson, C., & Krawczyk, D. C. (2009). Differences in task demands influence the hemispheric lateralization and neural correlates of metaphor. *Brain and Language*, *111*(2), 114–124. https://doi.org/10.1016/j.bandl.2009.08.006

Yee, E., & Thompson-Schill, S. L. (2016). Putting Concepts into Context. *Psychonomic Bulletin*

Journal Pre-proof

Journal Pre-proof

& Review, (June), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, Acting, Understanding: Motor Resonance in Language Comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Journal Pre-proof