Cognition 200 (2020) 104157

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Teleological generics

Joanna Korman[a,*,1], Sangeet Khemlani[b]

[a] The MITRE Corporation, United States of America
[b] Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory, United States of America

ARTICLE INFO

Keywords:
Generics
Teleology
Concepts
Principled connections
Formal explanations

ABSTRACT

Certain "generic" generalizations concern functions and purposes, e.g., *cars are for driving*. Some functional properties yield unacceptable teleological generics: for instance, *cars are for parking* seems false even though people park cars as often as they drive them. No theory of teleology in philosophy or psychology can explain what makes teleological generics acceptable. However, a recent theory (Prasada, 2017; Prasada & Dillingham, 2006; Prasada, Khemlani, Leslie, & Glucksberg, 2013) argues that a certain type of mental representation – a "principled" connection between a kind and a property – licenses generic generalizations. The account predicts that people should accept teleological generics that describe kinds and properties linked by a principled connection. Under the analysis, *car* bears a principled connection to *driving* (a car's primary purpose) and a non-principled connection to *parking* (an incidental consequence of driving). We report four experiments that tested and corroborated the theory's predictions, and we describe a regression analysis that rules out alternative accounts. We conclude by showing how the theory we developed can serve as the foundation for a general theory of teleological thinking.

> So it came to me in a flash: what's a gun for? To shoot! So I shot. Simple isn't it? – screenplay for *His Girl Friday*.
>
> (Lederer, 1939)

1. Introduction

People often make general statements about the functions of objects and entities. A natural answer to the question, "what are cars for?" may be: *cars are for driving*. The answer is a "generic" statement, since it concerns cars in general instead of some specific instance of a car (Carlson, 1977; Cimpian & Markman, 2009; Gelman, 2003, 2004; Gelman & Bloom, 2007; Hampton, 2009; Pelletier, 2009; Prasada, 1999, 2000, 2010, 2012, 2017; Prasada & Dillingham, 2006, 2009). Other generics include statements such as *iguanas are green* and *cars have radios*. Generics serve to connect a representation of a kind (e.g., *iguanas*) to a specific property (e.g., *being green*). Children understand and produce generics around three years of age, long before they understand explicit quantification – and many theorists have argued that generics express the most rudimentary form of generalization in humans (Leslie, 2007, 2008, 2012; Leslie & Gelman, 2012; Gelman, 2010). Because generics refer to kinds of things in the abstract, they can tolerate exceptions (e.g., *iguanas are green* is true even if you see a purple iguana), and people draw broad inductive conclusions from them (Cimpian, Brandone, & Gelman, 2010; Khemlani, Leslie, & Glucksberg, 2012).

Few studies have focused on teleological generics that express functions and purposes, such as *cars are for driving*, even though children begin to understand the teleological functions of objects early in development (Atran, 1995; Carey, 1985; Csibra & Gergely, 1998; Keil, 1992). People often accept functional properties as applicable to particular sorts of objects, e.g., for artifacts such as *cars* or biological parts such as *eyes* (Keil, 1992). But each of these objects has many properties, and only a small subset of those properties describes functions. For instance, consider these generic statements about *cars*:

- *cars are for driving* [teleological generic]
- cars have radios
- cars have windshields
- * cars are for parking

The first statement describes the primary function cars serve, and so it is

* Corresponding author at: The MITRE Corporation, 202 Burlington Rd., Bedford, MA 01730, United States of America.
E-mail address: jkorman@mitre.org (J. Korman).

[1] Approved for Public Release, Distribution Unlimited. Public Release Case Number 20-0001. The author's affiliation with the MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author.

https://doi.org/10.1016/j.cognition.2019.104157
Received 12 November 2018; Received in revised form 8 December 2019; Accepted 11 December 2019
Available online 21 May 2020

an acceptable teleological generic. Other generalizations about the functions of cars seem unacceptable, e.g., *cars are for parking,* despite the fact that, e.g., people park cars as often as they drive them.

No existing theory explains why people accept some such generalizations but not others. In the present paper, we developed a novel theory of teleological generics that seeks to address this question. It accounts for the generalizations people make about everyday concepts. The theory distinguishes between acceptable and unacceptable generalizations for concepts in domains to which adults attribute functions, such as biological parts and artifacts (Keil, 1992; Kelemen, 1999). The theory likewise explains why teleological generics appear to reflect the normative nature of functions – i.e., why people think that cars should be for driving, and why cars that aren't for driving are somehow abnormal. Children are sensitive to such norms early on, e.g., they believe a device should be repaired when it can no longer perform its primary function (DiYanni & Kelemen, 2005). And, as Millikan (1989) observed, malfunctioning objects such as damaged cars retain their purpose despite their malfunction – the purpose of a car does not change even after a faulty engine renders it unusable.

In what follows, we first explain how teleological generics relate to another kind of teleological thinking explored by cognitive scientists, teleological explanations. While no existing theory explains what makes teleological generics acceptable, several proposals account for what makes teleological explanations acceptable. We consider these proposals, and we show why none of them suffices for a theory of teleological generics. We describe the new theory, which is based on a recent proposal of how people mentally represent kinds and their properties. We outline the predictions of the theory and report four studies that validate them. We also describe a regression analysis that rules out alternative accounts. We conclude by describing how the theory can be used to investigate teleological thinking more generally.

1.1. Teleological generics and teleological explanations

Many researchers have sought to explain what makes teleological explanations acceptable. Teleological explanations, like all explanations, are composed of an *explanandum* (the phenomenon to be explained) and an *explanans* (the thing that accounts for the phenomenon; see Hempel & Oppenheim, 1948). Teleological explanations are those in which the explanans of a phenomenon refers to a relevant function or purpose, e.g.,

1. Bears have fur in order to keep warm

The sentence is composed of two different generalizations: one generic describes the explanandum – the statement *bears have fur* – and another tacit generic describes the explanans: *fur is for keeping warm* (cf. Wright, 1976, p. 88). Philosophers since Aristotle have developed prominent accounts of what makes such teleological explanations valid (Allen, Bekoff, & Lauder, 1998; Mayr, 1985; Nagel, 1961; Neander, 1991; Perlman, 2004; Sober, 1984; Wright, 1976), while psychologists have focused on what features of a teleological explanation render it acceptable (e.g., Kelemen, Rottman, & Seston, 2013; Lombrozo & Carey, 2006).

Teleological generics such as *cars are for driving* do not constitute explanations, because while they can serve as part of an explanans (as in (1)), they do not refer to an explanandum or an explanans. Hence, while generics can serve as the building blocks of explanations, they are more rudimentary. Furthermore, examples of teleological explanations abound that do not appear to refer to generics at all (e.g., "This piece of paper on my desk is for taking attendance.") And as we show below, proposals designed to account for what constitutes acceptable teleological explanations have difficulty explaining what makes generics true.

1.1.1. Causal history

One prominent account of teleological explanation comes from Wright (1976), who argued that explanations such as (1) refer to the consequences of the historical processes that brought about the referenced function. Under Wright's analysis, the presence of fur in bears is best explained by the causal antecedents that drove the development of fur, i.e., the need to keep warm. He argued that a teleological statement of the form, e.g., *A function of X is F*, is true whenever the conditions in (2) are met:

2. F is a consequence of X
   X is there because it results in F

Hence, keeping warm is a consequence of fur, and fur is there because it results in animals keeping warm, so (1) is a valid explanation.

Wright's analysis was normative – it sought to explain why some ideal scientific explanations could be teleological in nature, a point that enlightenment scholars such as Francis Bacon had long rejected (*De Augmentis Scientarum, Book III*). Lombrozo and Carey (2006) re-interpreted Wright's account as a psychological proposal, and they ran studies designed to test whether humans accept teleological explanations on the basis of their causal history. They found that people endorsed explanations more often for situations in which causal antecedents were made explicit, and they argued that the pattern lends some credence to Wright's (1976) proposal. They further observed that participants preferred teleological explanations when the causal antecedents referred to a general pattern, i.e., a pattern that could be used for making predictions about the future.

Wright's causal history account may be useful in accounting for teleological explanations of natural phenomena, such as those that come about from evolutionary processes, and it influenced other theoretical accounts, such as Millikan's notion that objects have a "proper function" (Millikan, 1989, p. 292; see also Wright, 2013). But the causal history account does not – and was not intended to – explain the acceptability of teleological generics. As we show below, people reject the notion that driving is a consequence of cars, but they accept the generic, *cars are for driving*. More generally, the account cannot explain why, e.g., young children appear to understand the primary functions of common artifacts and biological parts of natural kinds (i.e., body organs such as eyes and hands) without the knowledge of any causal history. When asked to describe what a given object is for, children appeal to the object's intended function even if the object was used for some other purpose (Kelemen, 1999a, Study 3). Hence, they seem to accept teleological explanations and generalizations even when their personal history with the object suggests some alternative use.

A more recent psychological proposal describes a cue – the fit between an object's physical structure and its primary function – for when to accept or reject teleological explanations (Liquin & Lombrozo, 2018). But it, too, cannot explain the acceptability of teleological generics.

1.1.2. Structure-function fit

Liquin and Lombrozo (2018) argued that the acceptability of a teleological explanation as in (1) is supported by the apparent "fit" between the physical structures involved and the function described. Hence, fur has a physical structure that consists of thin, densely-packed, soft hairs that support the ability to keep an animal warm, and so the structure of fur and the function described in (1) are compatible. As Liquin and Lombrozo (2018) argued, structure-function fit serves as a cue that provides a rapid way to assess the quality of an explanation. Unlike Wright's (1976) account, structure-function fit can explain why children can understand and accept teleological explanations without knowing anything about the causal history of a phenomenon.

Structure-function fit cannot explain the acceptability of teleological generics, because many unacceptable generics, e.g., *eyes are for blinking*, nevertheless reflect a strong compatibility between structure and a particular function. Furthermore, generics often express normative expectations, e.g., the generic *noses are for smelling* suggests that a nose that isn't for smelling is an abnormal nose, or possibly not a nose at all. However, people may accept some teleological explanations that describe a fit between the structure of a nose and its function without licensing a normative expectation (Liquin & Lombrozo, 2018, p. 23), e.g.,

3. People have noses because noses hold up glasses

In spite of the structural compatibility of noses with the function of holding up glasses, people are unlikely to regard a nose that fails to hold up glasses as abnormal. Hence, structure-function fit does not capture the normative expectations derived from teleological generics.

To explain people's lay conceptions of everyday teleological generics, we propose a new account of teleological generics based on a recent theory of how people mentally represent abstract kinds (Prasada, Khemlani, Leslie, & Glucksberg, 2013). The theory posits that people represent a privileged connection between kind representations (e.g., *cars*) and some of their distinctive properties (Prasada & Dillingham, 2006). These privileged connections, known as "principled connections," permit certain inferences that other sorts of non-privileged connections do not. We turn to describe the theory and the empirical predictions it makes.

1.2. Generics and the theory of principled connections

Prasada and colleagues argue that generic language is a byproduct of how people mentally represent kinds and their properties (Prasada, 2017; Prasada et al., 2013; Prasada & Dillingham, 2006, 2009). Their theory identifies three separate mental representations invoked when people use generic language: the representation of a kind (e.g., *cars*), the representation of a property (e.g., *having four wheels, having rearview mirrors*, *having radios*), and the representation of a connection between the kind and the property. Fig. 1 outlines these three distinct representations.

The theory posits that people mentally represent distinct types of connections between kinds and properties, and that these connections underlie generic statements of form *Xs have property Y*. For the purposes of the present paper, the most important kind of connection between kinds and properties is a "principled" connection, which connects kinds to their most characteristic and defining properties, i.e., it connects the kind *cars* to those properties that are part of what it means to be a car. Properties that bear a principled connection to a kind do not depend on

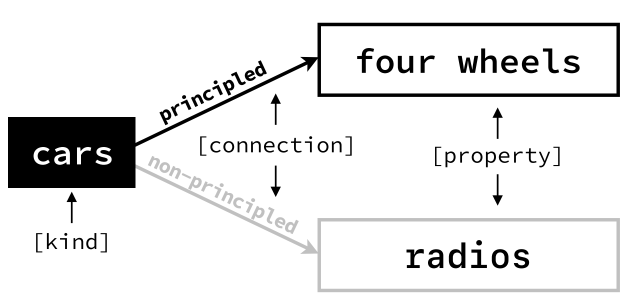

Fig. 1. The structure of kind representations posited by Prasada et al.'s (2013) theory of generic language. The theory separates representations of kinds (denoted by the black box labeled "cars") from representations of relevant properties (denoted by the two white boxes) as well as the connections between kinds and properties (denoted by the arrows). Principled connections can be distinguished from non-principled connections: for instance, they yield normative expectations (people should accept the statement "cars are supposed to have four wheels" but reject the statement "cars are supposed to have radios").

considerations of prevalence – they are not true merely because many, most, or all instances of that kind possess the property (Haward, Wagner, Carey, & Prasada, 2018). Indeed, certain properties are characteristic of kinds even though they are held by a minority of instances. For instance, people treat *laying eggs* as part of what it means to be a *duck*, and so they accept the generic "ducks lay eggs" even though a minority of ducks lay eggs (Leslie, Khemlani, & Glucksberg, 2011). As recent evidence shows, people privilege properties that bear principled connections to kinds in systematic ways: such properties come to mind more readily (Hussak & Cimpian, 2018), they yield the default inference that an arbitrary instance of the kind possesses the property (Khemlani et al., 2012), and their absence can cause people to question whether a specific instance is a member of the kind (Prasada et al., 2013). Thus, they can be distinguished from properties that bear a "statistical" connection to a kind. Statistical connections link a kind to a property only in virtue of being true of many or most instances of the kind. For example, *having radios* bears a statistical connection to the kind *cars:* most cars happen to have radios. Nevertheless, since there is no principled connection between being a car and having a radio, people should not question whether a particular car that lacks a radio is indeed a car.

The theory of principled connections helps explain how people rapidly map mental representations of properties to representations of kinds. For example, consider the generic, *daxes have stripes*. The theory posits that listeners who know nothing else about *daxes* should interpret *having stripes* as a characteristic property (Gelman & Brandone, 2010), and afford it all of the privileges described above (e.g., they should permit the default inference that an arbitrary dax has stripes). Recent data from children and adults show that people accept generics even when information about the prevalence of such properties should prevent them from creating strong mappings (Cimpian et al., 2010).

Principled connections are conceptually distinct from non-principled connections (Prasada et al., 2013, p. 408 et seq.), and four separate linguistic tests can be used to distinguish them:

i. The self-referential generalization test: when a kind and a property bear a principled connection, that connection licenses "self-referential" generalizations of the property, e.g., *Xs, by virtue of being Xs, are Y* (Prasada & Dillingham, 2006). Hence, people should accept the statement *cars, by virtue of being cars, have four wheels* but reject the statement *cars, by virtue of being cars, have radios.*
ii. The norm test: principled connections set up norms (Prasada & Dillingham, 2006). Hence, people should consider it normal for a car to have four wheels; they should consider it abnormal when a car doesn't have four wheels; and they should think that cars are supposed to have four wheels. In contrast, cars that don't have radios shouldn't be considered abnormal, and no expectation should exist that cars are supposed to have radios.
iii. The aspect test: properties that are linked to a kind by a principled connection should be considered an aspect of the kind (Prasada & Dillingham, 2009). Hence, people should agree with the statement *having four wheels is one aspect of being a car* but disagree with the statement *having a radio is one aspect of being a car.*
iv. The formal explanations test: principled connections should license "formal" explanations, i.e., shallow explanations that explain properties by reference to the kind (Prasada, 2017). Hence, people should agree with statements such as *cars have four wheels because they're cars* but disagree with statements such as *cars have radios because they're cars.*

Prasada et al. (2013) corroborated the predictions outlined above. They further showed that when the four tests are satisfied between a kind (*car*) and a property (*having four wheels*), people accept the corresponding generic (e.g., *cars have four wheels*), and that other sorts of conceptual connections can cause people to accept generics as well.

We developed a novel theory of the acceptability of teleological generics by positing that people represent a principled connection between a kind and its primary function. Teleological generics of the form *Xs are for Y* should be deemed acceptable when the kind (*Xs*) and the property (*being for Y*) are linked by a principled connection. The proposal is consistent with Millikan's (1989) and Neander's (1991) philosophical proposals that teleology concerns norms of proper function. It also comports with developmental data that show that children regard particular functions as normatively tied to certain kinds of objects (Kelemen, 1999a; DiYanni & Kelemen, 2005; Casler, Terziyan, & Greene, 2009). And since the theory applies to all teleological generics of the form *Xs are for Y*, it provides a uniform treatment for parts of biological kinds and for artifacts. But it goes beyond any other account by assuming that teleological generics are acceptable based on the way in which people mentally represent kinds and properties, and it specifies the expectations licensed by those representations. Accordingly, the theory makes the following prediction:

> **Prediction 1:** People should accept teleological generics (e.g., *cars are for driving*) whenever they accept assertions corresponding to the four linguistic tests of principled connections, i.e., self-referential generalizations, expectations regarding norms, aspectual descriptions, and formal explanations (see Fig. 1). When people reject any linguistic tests of principled connections, they should also reject the corresponding teleological generalizations.

We describe four separate experiments that tested prediction 1.

We further show how the theory diverges from alternative accounts. For instance, a recent proposal suggests that people may distinguish between acceptable and unacceptable generics on the basis of probabilistic expectations (see, e.g., Tessler & Goodman, 2019). Hence, people may think that *chairs are for sitting* is more acceptable than *chairs are for dusting* because the probability of sitting on a chair is higher than the probability of dusting it. Our theory posits, however, that principled connections can yield stronger generalizations than probabilistic knowledge should permit, and so people's endorsement of principled connection tests should better predict their endorsement of teleological generics. Likewise, people's acceptance of generics should be based neither on their consideration of causal history (Wright, 1976), nor on their assessment of the "fit" between the physical structure of a kind and its function (Liquin & Lombrozo, 2018). Hence, the theory makes the following additional prediction:

> **Prediction 2:** Participants' assessment of the four linguistic tests of principled connections should be more strongly associated with the acceptability of teleological generics than alternative metrics, including those derived from theories of teleological explanation (i.e., causal history and structure-function fit) and theories of generic endorsement (i.e., probabilistic expectations).

We conducted a hierarchical regression analysis to test prediction 2.

2. Experiment 1

Experiment 1 served two purposes: first, it tested whether participants distinguish teleological generics that potentially represent principled connections from those that do not (prediction 1). Half the participants in the study had to assess the truth of teleological generics such as *eyes are for seeing* and *eyes are for blinking*. Materials consisted of pairs of items in which a noun (e.g., "eyes") appeared with a potentially acceptable teleological function verb ("seeing") or a potentially unacceptable one ("blinking"). Second, the experiment examined the first of four separate linguistic tests that distinguish principled connections from non-principled connections, i.e., the other half of the participants rated corresponding self-referential generalizations (e.g., *eyes, by virtue of being eyes, are for seeing* vs. *eyes, by virtue of being eyes, are for blinking*). If prediction 1 is true, then people's evaluations of self-referential generalizations should be highly correlated with their evaluation of teleological generics.

2.1. Method

2.1.1. Participants

The optimal sample size for each Experiment 1 and all subsequent studies was projected by doubling the sample size from a reference study (Prasada et al., 2013, Experiment 2, which used $n = 25$). To verify that our sample sizes were reasonable, we conducted a post-hoc power analysis using the *pwr* package in R (Champely, 2018), which estimated an ideal sample size of 45 participants. 51 participants (21 female; mean age = 37.8 years) completed Experiment 1 on Amazon Mechanical Turk. All participants came from a pool of IP addresses from the United States of America, and all participants self-reported as native English speakers but one; excluding that participant's responses had no effect on the results below. Participants tended to take less than 20 min to complete the study and were compensated $2. They were further notified that they would be entered into a drawing to win a $10 bonus if they followed the task instructions. The bonus was subsequently paid to 10% of participants, chosen randomly. All subsequent experiments and norming studies below adopted the same protocol.

2.1.2. Materials

The materials consisted of a concept (noun) that was either an artifact or biological part of a natural kind paired with either an experimental or a control verb. Verbs described common actions performed with or by the noun, e.g., the noun "eyes" was paired with the verb "seeing" or else "blinking." Experimental and control verbs differed in that experimental verbs were chosen so as to yield acceptable teleological generics (e.g., *eyes are for seeing*). Control verbs were chosen based on two criteria: first, control verbs should all yield false generalizations (e.g., *eyes are for blinking*). Second, they should span the range of various probability estimates, e.g., how likely something is an eye given that it's for blinking, P(*eye* | *for-blinking*). Those criteria were validated in a norming study (reported in Section 6.2 below). Half of the 22 objects were artifacts (e.g., *cars*), while the other half were biological parts (i.e., bodily organs such as *eyes*). Table 1 provides a list

Table 1
The nouns and their corresponding verbs in Experiments 1–4. Each concept appeared as a plural noun (e.g., *eyes…*) and was paired with either an experimental (e.g., *seeing*) or a control (e.g., *blinking*) verb.

	Noun	Verb	
		Experimental	Control
Artifacts	bags	carrying	storing
	books	reading	packing
	cars	driving	painting
	chairs	sitting	dusting
	cups	drinking	stacking
	forks	eating	washing
	knives	cutting	sharpening
	ladders	climbing	leaning
	pencils	writing	carrying
	pianos	playing	tuning
	shoes	wearing	polishing
Biological parts	brains	thinking	sleeping
	ears	hearing	plugging
	eyes	seeing	blinking
	hands	grasping	clapping
	muscles	moving	flexing
	noses	smelling	sniffing
	stomachs	digesting	grumbling
	teeth	chewing	brushing
	throats	swallowing	coughing
	tongues	tasting	licking
	wings	flying	flapping

of the 22 objects and their corresponding verbs.

2.1.3. Design and procedure

Participants were instructed to evaluate the truth of statements about common objects and entities. Half of the participants evaluated teleological generics and the other half evaluated self-referential generalizations. Teleological generics were assertions of the form NP_{plural} + $VP_{purposive}$, e.g., "eyes" + "are for seeing". Self-referential generalizations described the purpose of a given object by virtue of being that object, e.g., *eyes, by virtue of being eyes, are for seeing*. In both conditions, participants registered their responses by moving a slider displayed on the screen that implemented a Likert scale that ranged from −3 (definitely false) to 3 (definitely true). They served as their own controls, i.e., each received 22 distinct nouns × 2 types of verb (control vs. experimental). Hence, participants assessed 44 assertions in total. The study presented the items in a randomized order.

2.1.4. Open science

Data, code, and complete materials for Experiment 1 and subsequent experiments can be found at: https://osf.io/92qp3/. Predictions were preregistered prior to data collection for this and subsequent experiments (https://osf.io/vzmdj).

2.2. Results and discussion

Fig. 2 (left two columns) shows the results of participants' ratings for the items in Experiment 1. Their evaluations corroborated prediction 1: they rated experimental items as more truthful than control items ($M_{experimental} = 2.49$ vs. $M_{control} = -0.80$, Wilcoxon test, $z = 6.22$, $p < .0001$, Cliff's $\delta = 0.99$). Their ratings did not differ as a function of whether the statement was a teleological generic or self-referential generalization ($M_{generic} = 0.92$ vs. $M_{self\text{-}referential} = 0.78$, Mann-Whitney test, $z = 0.70$, $p = .48$, Cliff's $\delta = 0.08$). And, their ratings did not yield a reliable interaction between the type of verb (experimental vs. control) and the type of generalization (generic vs. self-referential; Mann-Whitney test, $z = 1.44$, $p = .15$, Cliff's $\delta = 0.24$).

Planned comparisons examined experimental vs. control items for teleological generics and self-referential generalizations in isolation. The participants who evaluated teleological generics rated assertions describing experimental items as truer ($M = 2.50$) than those describing control items ($M = -0.67$; Wilcoxon test, $z = 4.37$, $p < .0001$, Cliff's $\delta = 0.99$). The result served as a manipulation check: it confirmed that the participants construed experimental items as acceptable generics. A similar pattern held for the group of participants that evaluated self-referential generalizations ($M_{experimental} = 2.48$ vs. $M_{control} = -0.91$; Wilcoxon test, $z = 4.46$, $p < .0001$, Cliff's $\delta = 0.99$). Ratings of teleological generics were strongly correlated with ratings of self-referential generalizations, which confirmed prediction 1 ($r = 0.99$, $p < .0001$). Hence, one group's evaluations of the truth of self-referential generalizations strongly predicted another group's evaluations of teleological generics. Prasada et al. (2013) posit that self-referential generalizations are diagnostic of principled connections, and so the results of Experiment 1 suggest that participants represented principled connections for experimental items. Experiments 2a and 2b sought to extend the finding by exploring the second of four linguistic tests of principled connections.

3. Experiments 2a and 2b

Principled connections set up norms, and so the second test of principled connections concerns how people construe what's normal for a particular object and the functions that object should have. As Prasada et al. (2013) point out, there are two ways of examining norm-based expectations. One way focuses on normality, i.e., the assessment of whether it's normal for a particular kind to possess a property. Another way focuses on normativity, i.e., whether it's ideal and desired for a particular kind to possess the property. Hence, Experiment 2a provided participants with statements about normality (e.g., *all normal eyes are for seeing*) and normativity (e.g., *eyes are supposed to be for seeing*) for the experimental and control verbs provided in Table 1. Experiment 2b provided participants with an alternative statement about normativity (e.g., *an eye should be for seeing*). If prediction 1 is true, then participants should distinguish experimental vs. control verbs for both statements of normality and statements of normativity. Likewise, people's evaluations of statements based on normality and normativity should be correlated with one another, and they should not differ reliably from one another, since both should be equivalently predictive of participants' tendency to accept corresponding teleological generics. Prediction 1 is false if participants' evaluations of the two sorts of statements do not correlate or inversely correlate with one another.

3.1. Method

3.1.1. Participants

43 participants (28 female; mean age = 36.9 years) in Experiment 2a and 50 participants (20 female; mean age = 37.0) in Experiment 2b completed the tasks on Amazon Mechanical Turk. All participants were native English speakers and adhered to the constraints described in Section 2.1.1.

3.1.2. Materials, design, and procedure

As in the previous study, participants were instructed to evaluate the truth of statements about the objects and entities provided in Table 1 in Experiments 2a and 2b. Half of the statements concerned experimental verbs, i.e., verbs that were likely to bear principled connections, and the remaining half concerned control verbs, i.e., verbs that were likely to produce false generalizations. But participants in Experiments 2a and 2b did not consider teleological generics. Rather, in Experiment 2a, half the statements referenced expectations of what was normal of the objects in the study (e.g., they appeared as *all normal eyes are for seeing*) and the other half referenced a normative property of those objects (e.g., they appeared as *eyes are supposed to be for seeing*) for all 44 control and experimental items, and so participants in Experiment 2a rated 88 items in total. Experiment 2b addressed a concern raised by reviewers that participants may have been confused by the "supposed to" formulation. Specifically, they may have failed to recognize it as referring to a norm and mistakenly assumed that the formulation was simply another way of expressing the generic (e.g., *eyes are for seeing*). Experiment 2b addressed this issue by inviting a separate group of participants to rate 44 items using a different way of describing normative properties (e.g., *an eye should be for seeing*). Participants in both studies rated the extent to which each statement struck them as true on a scale that ranged from 3 to −3.

3.2. Results and discussion

Fig. 2 (middle three columns) shows the results of participants' ratings for the items in Experiments 2a and 2b. Participants rated experimental items higher than control items in Experiment 2a ($M_{experimental} = 2.30$ vs. $M_{control} = -0.61$, Wilcoxon test, $z = 7.63$, $p < .0001$, Cliff's $\delta = 0.97$). In Experiment 2a, normativity was measured with the "supposed to" formulation, e.g., *eyes are supposed to be for seeing.* Participants' ratings did not differ for generalizations concerning normality vs. normativity ($M_{normality} = 0.86$ vs. $M_{normativity} = 0.83$, Wilcoxon test, $z = 0.55$, $p = .58$, Cliff's $\delta = 0.004$). The results did, however, yield a reliable and unpredicted interaction between the type of verb (experimental vs. control) and the type of assertion (normality vs. normativity; Wilcoxon test, $z = 2.62$, $p = .009$,

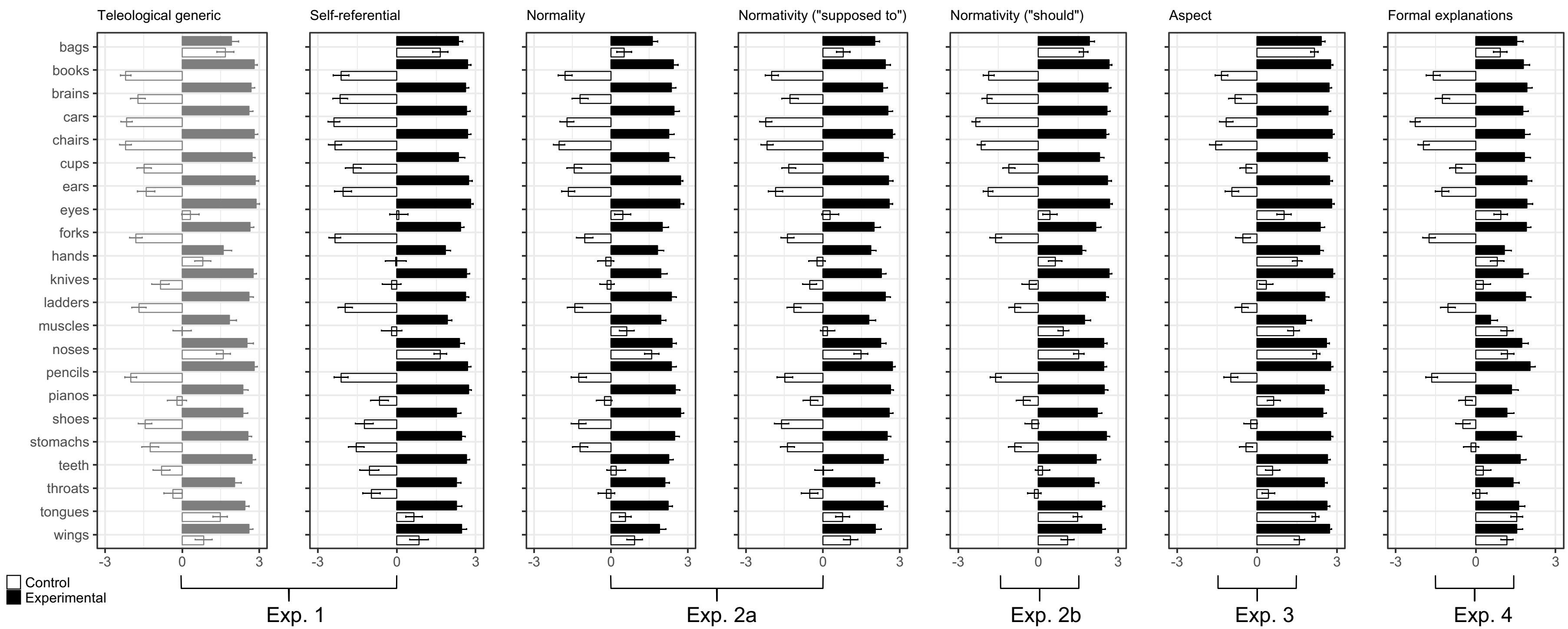

Fig. 2. Participants' truth ratings on teleological generics and self-referential generalizations (Experiment 1); on assertions concerning normality and normativity (Experiments 2a and 2b); on assertions concerning aspect (Experiment 3); and on formal explanations (Experiment 4). All panels show ratings as a function of the 22 nouns in alphabetical order and the two types of teleological property (control vs. experimental). Error bars denote 95% confidence intervals. Teleological generics are shown in gray and the assessments that correspond to principled connection tests are shown in black.

Cliff's $\delta = 0.33$), because participants gave more extreme ratings for normativity assertions than for normality assertions. A planned comparison on statements that concerned normality revealed that participants provided higher ratings for experimental items than control items ($M_{experimental} = 2.26$ vs. $M_{control} = -0.54$, Wilcoxon test, $z = 5.43$, $p < .0001$, Cliff's $\delta = 0.96$). A more extreme pattern of the same effect held for a planned comparison on statements that concerned normativity ($M_{experimental} = 2.34$ vs. $M_{control} = -0.68$, Wilcoxon test, $z = 5.40$, $p < .0001$, Cliff's $\delta = 0.97$), which explains the interaction observed.

The results from Experiment 2b, which measured normativity using a different formulation, e.g., *eyes should be for seeing*, mirrored those from 2a: participants rated experimental items as higher than control items ($M_{experimental} = 2.36$ vs. $M_{control} = -0.43$, Wilcoxon test, $z = 6.15$, $p < .0001$, Cliff's $\delta = 0.96$). Both studies corroborated prediction 1, which states that principled connections should set up norms. People accepted norm-based statements that the theory predicts should bear principled connections (the experimental items), and they rejected norm-based statements that produce false generalizations.

Experiment 3 examined the third of four linguistic tests of principled connections.

4. Experiment 3

When a principled connection is present between a kind and a property, people construe the property as an aspect of the kind. Hence, people readily agree that having four wheels is one aspect of what it means to be a car (Prasada et al., 2013). Experiment 3 tested participants' evaluations of aspectual descriptions, that is, statements that describe a given functional property as being "an aspect" of the kind in question (e.g., *one aspect of eyes is that they're for seeing*). If prediction 1 is true, people's evaluations of experimental and control items should reliably differ.

4.1. Method

4.1.1. Participants

51 participants (25 female; mean age = 37.86) completed the task on Amazon Mechanical Turk. All but one participant was a native English speaker; excluding that participant's data had no effects on the results below. And all adhered to the constraints described in Section 2.1.1.

4.1.2. Materials, design, and procedure

Participants considered 44 statements in Experiment 3, which were composed of the 22 experimental verbs and the 22 control verbs provided in Table 1. They considered statements that referenced expectations of what properties could be considered an aspect of objects in the study, e.g., *one aspect of eyes is that they're for seeing.* As in the previous studies, participants rated the extent to which each statement struck them as true on a scale that ranged from 3 to −3.

4.2. Results and discussion

Fig. 2 (second-to-last column) shows the results of participants' ratings for the items in Experiment 3. They provided higher ratings for experimental verbs than control verbs ($M_{experimental} = 2.60$ vs. $M_{control} = 0.23$, Wilcoxon test, $z = 6.19$, $p < .0001$, Cliff's $\delta = 0.96$). The experimental verbs therefore passed the aspect test described in Section 1.2, the third of four linguistic tests of principled connections. And the control verbs failed the test. The results further corroborated prediction 1, which states that principled connections permit people to consider relevant properties as being an aspect of the kind.

Experiment 4 investigated the fourth and final linguistic test of principled connections, which concerned formal explanations: principled connections should license them, and non-principled connections should not.

5. Experiment 4

Principled connections should license formal explanations, i.e., explanations that explain properties by referring to the kind in question (Prasada, 2017). Formal explanations can often seem shallow, uninformative, and circular. The following, for instance, presents a dialogue in which the answer is a formal explanation:

4. Q: Why does that (pointing to a tiger) have stripes?
   A: Because it's a tiger.

The response may not be a scientifically satisfying explanation – nevertheless, it's natural and felicitous because it serves to convey a link between being a *tiger* and its characteristic property of *having stripes* (Haward et al., 2018; Prasada & Dillingham, 2009). In contrast, the following response seems unnatural and infelicitous because no link exists between the kind and the property:

5. Q: Why does that (pointing to a tiger) have a tongue?
   A: Because it's a tiger.

Experiment 4 further tested prediction 1 by providing participants with dialogues that included formal explanations for teleological phenomena; participants rated how natural the dialogues were.

5.1. Method

5.1.1. Participants

52 participants (22 female, mean age = 32.0 years) completed the experiment on Amazon Mechanical Turk. All participants were native speakers and adhered to the constraints described in Section 2.1.1.

5.1.2. Materials, design, and procedure

Experiment 4 used the same nouns (e.g., *eyes*) and the same experimental and control verbs (e.g., *seeing* and *blinking*, respectively) as shown in Table 1. However, participants in Experiment 4 did not consider individual statements. Instead, they considered a dialogue between two people in which the second person provided a formal explanation, e.g.,

6. Person 1: Why is that (pointing to an eye) for seeing?
   Person 2: Because it's an eye.

They then evaluated the formal explanation by rating to what extent Person 2's answer was a natural response to Person 1's question on a scale of −3 (very unnatural) to 3 (very natural). Participants served as their own controls and rated dialogues for a total of 44 items, half of which described experimental verbs and half of which described control verbs.

5.2. Results and discussion

Fig. 2 (last column) shows the results of participants' ratings for the items in Experiment 4. Participants rated dialogues that focused on experimental verbs to be more natural than dialogues that focused on control verbs ($M_{experimental} = 1.62$ vs. $M_{control} = -0.28$, Wilcoxon test, $z = 6.19$, $p < .0001$, Cliff's $\delta = 0.73$). These results further corroborate prediction 1. Nevertheless, as the effect size shows, the difference between participants' evaluations for experimental and control formal explanations was smaller than in the previous experiments. The smaller effect may be due to the fact that participants evaluated the dialogue on how natural it was, not on whether it was acceptable or not, and participants may have hesitated to mark anything as fully natural or unnatural.

Experiments 1–4 served to assess the four linguistic tests for principled connections (Section 1.2). The experiments provided converging evidence that, across the four linguistic tests, people distinguish experimental verbs (i.e., those that yield acceptable teleological generics) from control verbs (i.e., those that yield unacceptable generalizations). But, are the four linguistic tests predictive of people's tendency to accept teleological generics? In other words, does the extent to which people accept a formal explanation predict their acceptance of a corresponding teleological generic, or can some mitigating factor better account for why participants distinguished experimental from control verbs in Experiments 1–4? To answer these questions, we conducted a hierarchical regression analysis that considered four alternative accounts of teleological generics.

6. Alternative accounts of teleological generics

Experiment 1 showed that people accept teleological generics, e.g., *cars are for driving*, for experimental verbs but not control verbs. All four experiments showed, analogously, that experimental verbs are acceptable when formulated as specified by the four linguistic tests of principled connections, and that control verbs are unacceptable when expressed in those formulations. But the experiments do not rule out alternative accounts of teleological generalization. Four such accounts are worth considering; the next section addresses each in turn.

6.1. Alternative accounts

What alternative accounts can explain why people accepted teleological generics for experimental verbs in Experiments 1–4 above? First, according to Wright's (1976) philosophical account as well as Lombrozo and Carey's (2006) psychological re-interpretation, people could consider the causal history of how a kind in question developed in order to yield the functional property. In Section 1.1.1, we noted that Wright's (1976) account stipulated that a valid function of an object is one in which the object came about in order to carry out the function, as stated in (2). Hence, people might consider *eyes are for seeing* to be an acceptable generic because *eyes* came about in order for animals to *see*, not so that animals can *blink*. As we noted above, the account may have difficulty explaining how people construe generics about artifacts – but the causal history account may nevertheless explain the acceptability of teleological generics better than the principled connection tests examined in Experiments 1–4.

Second, a recent account by Liquin and Lombrozo (2018) argues that people accept teleological explanations when the structure of a particular object coheres with its function. The authors intended their theory to account for explanations, not generalizations (see Section 1.1.2 above), but perhaps their idea can be applied to predict when people accept teleological generics. For instance, under their analysis, people might accept *knives are for cutting* because the structure of knives – their sharpened edges, the materials they're made of, their manipulability – make them well-suited for the function of cutting. The mapping between the structure of a kind and its property also suggests that the function no longer applies when the structure changes. For instance, a butter knife is not for cutting because it does not have properties that permit it to cut, e.g., sharp edges.

Third, certain generics are acceptable because they apply to most instances of the kind in question, i.e., they have a high probability of occurrence (Prasada et al., 2013). For instance, the generic *cars have radios* is true because most cars have radios. The fact may be an accident of history – there is nothing abnormal about a car without a radio – yet the generic remains perfectly acceptable. The results from Experiments 1–4 may have been due to people's underlying statistical and probabilistic knowledge and not because they mentally represented principled connections. For example, those who accepted *cars are for driving* as a true teleological generic may have done so because they reasoned that cars, more often than not, are used for driving (see, e.g., Tessler & Goodman, 2019, for such a prevalence-based account of generics).

Finally, as an anonymous reviewer suggested, people may have accepted generics based on the frequency of the activity described in the statement. The concern applies to some of the materials more than others. For instance, consider *knives are for cutting* vs. *knives are for sharpening.* Provided that people sharpen knives because they get dull after repeated use, cutting must occur more frequently than sharpening. For other materials, the false generalization may occur more often than the primary function, e.g., *ladders are for climbing* vs. *ladders are for leaning.* Arguably, ladders lean more frequently than they're climbed, because it's only possible to climb ladders that lean against something (rope ladders notwithstanding). In some cases, the frequency of the activity may be difficult to ascertain, e.g., *forks are for washing.* The frequency of the activity depends on whether the fork is disposable or not.

6.2. Norming data for the alternative accounts

Five different accounts may explain people's tendency to accept teleological generics: one based on causal history, one based on structure-function fit, one based on probabilistic knowledge, one based activity frequency, and one based on principled connections. The data on people's evaluations of the linguistic tests of principled connections were available from Experiments 1–4 above. To obtain data for the remaining accounts, we conducted four different norming studies that collected participants' estimates of causal history, structure-function fit, two different kinds of probabilistic knowledge, and activity frequency. Each norming study collected data from a separate, new batch of participants from Amazon Mechanical Turk. The norming studies adapted the 44 items we provided to participants in Experiments 1–4.

The first norming study ($n = 50$) collected participants' estimates of causal history by focusing on consequentiality (see (2) above). Participants were asked to assess the acceptability of statements akin to, e.g., "Driving is a consequence of cars" and responded by selecting from a 7-point Likert scale (from -3 = definitely false to $+3$ = definitely true). The second norming study ($n = 51$) collected participants' estimates of structure-function fit by asking them a question adapted from Liquin & Lombrozo (2018), e.g., "To what extent do cars have a good fit with the function of driving?" Participants rated their responses on a 7-point Likert scale (from 1 = very bad fit to 7 = very good fit). The third norming study ($n = 39$) collected two kinds of conditional probability estimates: cue validity and prevalence (see Khemlani et al., 2012, for a similar analysis on non-teleological generics). Cue validity refers to the probability that an instance belongs to a kind given that it has a particular property (Murphy, 1982; Rosch, 1978); it can be construed as P(*car* | *for-driving*). For example, on one problem, participants evaluated cue validity by answering the following question: "Suppose a particular thing is for driving. What is the probability that that thing is a car?" Prevalence estimates refer to the opposite conditional probability, P(*for-driving* | *car*), i.e., the probability that an object has a particular property given that it belongs to a particular kind. Participants generated prevalence estimates by answering the following question: "What percentage of cars are for driving?" For both sorts of question, participants provided estimates on a movable slider ranging from 0% to 100%. The fourth and final norming study (n = 50) asked participants to assess how frequently a particular activity took place, e.g., "How often do cars get driven?" They rated their responses on a 7-point Likert scale (from −3 = very infrequently to +3 = very frequently). We omit additional details of the norming studies for brevity, but their corresponding data, materials, and code are available online (https://osf.io/92qp3/).

6.3. Hierarchical analysis

A regression model that predicts the acceptance of teleological generics can be constructed by aggregating only the data from people's evaluations of alternative accounts (i.e., from the norming studies): such a "baseline" model can serve as the strongest possible consideration of the alternatives. Hence, if the baseline model is outperformed by the addition of a predictor, then that predictor must explain people's acceptance of teleological generics beyond what can be achieved by all the competing alternative accounts.

We accordingly subjected the data from the norming studies as well as the data from Experiments 1–4 to a hierarchical regression analysis. Participants in Experiment 1 had rated teleological generics on their acceptability, and so their data were collected and aggregated across the separate items to yield 44 different data points to serve as the outcome variable (see column 6 in Appendix A). The data from Experiments 1–4 (columns 7–12 in Appendix A) and each of the norming studies (columns 1–5 in Appendix A) were aggregated in an analogous way. The regression equation for the baseline model was as follows:

$$Y_{TG} = \alpha + \beta_{CH}X_{CH} + \beta_{SFF}X_{SFF} + \beta_{CV}X_{CV} + \beta_P X_P + \beta_{AF}X_{AF} \qquad \textit{Model}_{\text{Baseline}}$$

where

Y_{TG} = the outcome variable, i.e., the tendency to accept teleological generics
X_{CH} = normed estimates of *causal history* (alternative #1)
X_{SFF} = normed estimates of *structure-function fit* (alternative #2)
X_{CV} = normed estimates of *cue validity* (alternative #3)
X_P = normed estimates of *prevalence* (alternative #3)
X_{AF} = normed estimates of *activity frequency* (alternative #4)

The regression was conducted using the *lme4* package in R (Bates, Maechler, Bolker, & Walker, 2015), and the code and results of the analysis are provided online (https://osf.io/92qp3/). The resulting baseline regression predicted the data well ($R^2 = 0.92$), and the following predictors significantly accounted for participants' acceptance of teleological generics: causal history ($\beta_{CH} = -0.44$, $p = .04$), structure-function fit ($\beta_{SFF} = 0.59$, $p = .007$), prevalence ($\beta_P = 0.06$, $p < .0001$). The other two predictors were unreliable, which suggests that neither cue validity nor activity frequency explain why people agree with some teleological generics and reject others.

The baseline regression model can be augmented by adding predictors derived from participants' evaluation of linguistic tests of principled connections. For instance, the baseline model can be augmented with their ratings of the acceptability of the 44 self-referential generalizations in Experiment 1. After aggregating those data across the different materials (see column 7 in the Appendix A), the baseline regression model was updated as follows:

$$Y_{TG} = \alpha + \beta_{CH}X_{CH} + \beta_{SFF}X_{SFF} + \beta_{CV}X_{CV} + \beta_P X_P + \beta_{AF}X_{AF} + \boldsymbol{\beta_{SRG}X_{SRG}} \qquad \textit{Model}_{\textit{Self-referential}}$$

where

X_{SRG} = agreement ratings of self-referential generalizations (Experiment 1)

The resulting regression model provided a better fit to the data than the baseline model ($R^2 = 0.98$). The baseline model, *Model*$_{\text{Baseline}}$, and the augmented model, *Model*$_{\text{Self-referential}}$, were subjected to an analysis of deviance, which is used to compare models that have nested structures to ascertain whether one model provides a significantly better fit to the outcome variable than another. The results of an analysis of deviance revealed that *Model*$_{\text{Self-referential}}$ provided a closer fit than *Model*$_{\text{Baseline}}$ ($\chi^2 = 62.23$, $p < .0001$). The predictors in the model that reliably accounted for generic acceptance were participants' evaluations of self-referential generalizations ($\beta_{SRG} = 0.75$, $p < .0001$) and their evaluations of prevalence ($\beta_P = 0.01$, $p = .05$); other predictors that had been highly reliable in *Model*$_{\text{Baseline}}$ were no longer significant in *Model*$_{\text{Self-referential}}$ (i.e., $\beta_{CH} = -0.09$, $p = .40$ and $\beta_{SFF} = 0.06$, $p = .59$), which suggests that the acceptability of self-referential generalizations is a better predictor of participants' acceptance of teleological generics than normed ratings of causal history or structure-function fit. The result corroborates prediction 2.

Five analogous regression models were constructed as follows:

$$Y_{TG} = \alpha + \beta_{CH}X_{CH} + \beta_{SFF}X_{SFF} + \beta_{CV}X_{CV} + \beta_P X_P + \beta_{AF}X_{AF} + \boldsymbol{\beta_{Nm}X_{Nm}} \qquad \textit{Model}_{\text{Normality}}$$

$$Y_{TG} = \alpha + \beta_{CH}X_{CH} + \beta_{SFF}X_{SFF} + \beta_{CV}X_{CV} + \beta_P X_P + \beta_{AF}X_{AF} + \boldsymbol{\beta_{Nta}X_{Nta}} \qquad \textit{Model}_{\text{Normativity}} \text{ (Exp. 2a)}$$

$$Y_{TG} = \alpha + \beta_{CH}X_{CH} + \beta_{SFF}X_{SFF} + \beta_{CV}X_{CV} + \beta_P X_P + \beta_{AF}X_{AF} + \boldsymbol{\beta_{Ntb}X_{Ntb}} \qquad \textit{Model}_{\text{Normativity}} \text{ (Exp. 2b)}$$

$$Y_{TG} = \alpha + \beta_{CH}X_{CH} + \beta_{SFF}X_{SFF} + \beta_{CV}X_{CV} + \beta_P X_P + \beta_{AF}X_{AF} + \boldsymbol{\beta_A X_A} \qquad \textit{Model}_{\text{Aspect}}$$

$$Y_{TG} = \alpha + \beta_{CH}X_{CH} + \beta_{SFF}X_{SFF} + \beta_{CV}X_{CV} + \beta_P X_P + \beta_{AF}X_{AF} + \boldsymbol{\beta_{FE}X_{FE}} \qquad \textit{Model}_{\text{FormalExplanation}}$$

where

X_{Nm} = agreement ratings to statements about normality (Experiment 2a)
X_{Nta} = agreement ratings of statements about normativity (Experiment 2a)
X_{Ntb} = agreement ratings of statements about normativity (Experiment 2b)
X_A = agreement ratings of statements about aspect (Experiment 3)
X_{FE} = naturalness ratings of formal explanations (Experiment 4)

For brevity, we focus only on the results of five separate analyses of deviance, which compared the four new models against the baseline model (see Table 2), and we report the details of the individual models in an online supplement (https://osf.io/5hjau/).

Table 2 shows that every model that included a predictor corresponding to a principled connection test outperformed the baseline model; supplementary materials likewise show that the principled connection tests were the strongest predictors in each of the five models

Table 2
Analysis of deviance table for the five models that include predictors for statements diagnostic of principled connections. Each row reports the results of an analysis of deviance against the baseline model, *Model*$_{\text{Baseline}}$. A significant *p*-value in the rightmost column serves as a corroboration of prediction 2.

	AIC	Deviance	R^2	χ^2	Significance
Model$_{\text{Self-referential}}$	16.75	−1.255	0.98	62.23	$p < .0001$
Model$_{\text{Normality}}$	52.63	34.63	0.96	26.35	$p < .0001$
Model$_{\text{Normativity (Exp 2a)}}$	38.97	20.97	0.97	40.01	$p < .0001$
Model$_{\text{Normativity (Exp 2b)}}$	43.67	25.67	0.96	35.30	$p < .0001$
Model$_{\text{Aspect}}$	35.24	17.24	0.97	43.74	$p < .0001$
Model$_{\text{FormalExplanation}}$	74.80	56.80	0.93	4.18	$p = .04$

shown in Table 2. With one exception, each model yielded a value of R^2 close to ceiling. The exception was the model that incorporated participants' evaluations of formal explanations, $Model_{FormalExplanation}$. The weaker results for formal explanations may be due to how different Experiment 4 was from the previous three experiments: it presented a dialogue, not a single sentence; it asked for naturalness ratings, not agreement ratings; and it made use of deictic parentheticals, e.g., "(pointing to a car)". All of these superficial differences may have perturbed participants' interpretations. In addition, the formal explanations presented in our studies may have struck participants as awkward because they were redundant or uninformative: participants may have judged that the explanation seeker already knew the item's label. Nevertheless, $Model_{FormalExplanation}$ outperformed $Model_{Baseline}$ ($\chi^2 = 4.18$, $p = .04$), which shows that even the weakest model incorporating a principled connection test outperformed a model that integrates all the alternative accounts. This result and the results of the other models in Table 2 corroborate prediction 2.

The results of the analyses of deviance provided strong evidence against alternative accounts of the acceptability of teleological generics. The baseline model included predictors that correspond to normed estimates of causal history (Wright, 1976), structure-function fit (Liquin & Lombrozo, 2018), probabilistic knowledge (Tessler & Goodman, 2019), and activity frequency. The analyses suggest that such alternatives do not adequately explain participants' tendencies to accept or reject teleological generics. Instead, the regressions revealed that the linguistic tests of principled connections best predict people's tendency to accept teleological generics.

7. General discussion

The statement *cars are for driving* seems to be a reasonable generalization. It concerns the "teleology" of cars, i.e., their purpose or function. As a generalization, it permits exceptions – concept cars and malfunctioning cars aren't used for driving – but the existence of abnormal instances does not make the generalization any less acceptable. Cars can serve multiple purposes: they're used to go to work, they're used as homes, they're used as shields in warzones. But most of those purposes don't yield compelling generalizations, e.g., it's false to say *cars are for going to work*, even if they're frequently used as such. It's likewise false to say *cars are for parking*, even though they're necessarily parked each time they're used. Only some teleological generalizations are acceptable, e.g., *cars are for driving,* but no existing theory explains why. To address the deficit, we developed a theory based on the idea that generic language such as *cars have four wheels* or *cars are for driving* is a byproduct of the way people mentally represent kinds and their properties (Prasada, 2017; Prasada et al., 2013; Prasada & Dillingham, 2006, 2009). The theory posits that people mentally represent "principled" connections between certain kinds (*cars*) and certain privileged properties (e.g., *having four wheels*). Principled connections can be self-referential, they help establish norms, they license aspectual construals – e.g., the expectation that having four wheels is an aspect of being a car – and they license formal explanations. We adapted the analysis to cover teleological generics (e.g., *cars are for driving*). Under our account, people accept *cars are for driving* because they represent a principled connection between *cars* and *being for driving*; they don't represent any such connection between *cars* and incidental properties (e.g., *being for going to work* or *being for parking*).

Prasada et al. (2013) introduced several linguistic tests to diagnose principled from non-principled connections (see Fig. 1 above). We accordingly adapted the tests into a series of four experiments in which participants evaluated self-referential generalizations, e.g., *cars, by virtue of being cars, are for driving* (Experiment 1); statements about normality, e.g., *all normal cars are for driving* and normativity, e.g., *cars are supposed to be for driving* (Experiment 2a) and *a car should be for driving* (Experiment 2b); statements about aspect, e.g., *one aspect of cars is that they're for driving*; (Experiment 3); and formal explanations, e.g., [that particular car is for driving] *because it's a car* (Experiment 4). Participants rated test statements higher when the corresponding teleological generic was acceptable (e.g., *cars are for driving*) than when it was unacceptable (e.g., *cars are for parking*), and so the experiments provide provisional evidence that principled connections explain what makes a teleological generic acceptable or unacceptable.

Can other accounts explain people's tendency to accept some teleological generics and reject others? No theory has ever addressed the matter, because existing theories of teleological thinking tend to focus on what makes teleological explanations good or bad. We re-interpreted four ideas in philosophy and psychology to yield deflationary alternative accounts of teleological generics. One account states that people should consider the causal history of teleological generics, e.g., people should accept *cars are for driving* because cars were developed because they result in driving (adapted from Wright, 1976). A second account states that people should accept the generic because the physical structures of cars – the fact that they have seats and wheels – make them conducive to driving (adapted from Liquin & Lombrozo, 2018). A third account posits that people's background knowledge about probabilities, e.g., the probability that something is a car given that it's for driving, can explain generalizations (adapted from Tessler & Goodman, 2019). And a fourth account, due to an anonymous reviewer, proposes that the frequency of an activity, e.g., the fact that cars are often used for driving, can explain why people tend to agree with *cars are for driving.* We ran a series of norming studies to reflect people's intuitions about these four deflationary accounts, and we entered the data from those studies into hierarchical regression analyses. The analyses revealed that while some of these alternative accounts, and probabilistic knowledge in particular, were meaningfully predictive of generic endorsement, assertions diagnostic of principled connections better predicted people's tendency to endorse teleological generics than any of the four separate deflationary accounts.

Psychological accounts of teleology agree that both artifacts and parts of biological natural kinds, e.g., body parts, can serve functions (e.g., Keil, 1994; Kelemen, 1999). Accordingly, in all of the experiments we report, we provided participants with an equal number of artifact kinds and body parts (see Table 1). In retrospect, it may not be surprising that deflationary accounts did not predict people's agreement to teleological generics as well as the principled connection tests, because at least two of the four alternative accounts have difficulty explaining teleological generics about artifact kinds (Wright, 1976; Liquin & Lombrozo, 2018). For instance, Wright's causal history account states that people should construe an object's function as being acceptable if that function is a consequence of the object's existence. But driving is not a consequence of the existence of cars: people drove carriages and bicycles before they drove cars. Likewise, Liquin and Lombrozo's structure-function fit account describes a psychological heuristic that assesses the compatibility between an object's structure and its function. But the account would permit bizarre and false generalizations, such as *noses are for holding up glasses*, since noses have a structure conducive to this function. Whatever success the structure-function fit account might have in accounting for people's teleological explanations (Liquin & Lombrozo, 2018), it cannot be extended to account for the acceptability of teleological generics. In contrast, principled connections explain the acceptability of generics that describe the functional properties of both biological parts and artifacts. Though differences may exist between artifacts and biological parts, our theory posits that people represent them both in a uniform way such that principled connections link representations of kinds to their functional properties.

Experiments 1–4 show that when a principled connection exists between a kind and a function, the corresponding teleological generic should be acceptable. Does the result suggest that *only* principled connections license teleological generics? As Prasada et al. (2013) argue, other types of connections, such as statistical connections, can also license generics (at least in the case of non-teleological generics, e.g., *cars have radios*; see Fig. 1 above). The present studies do not rule out the possibility that some acceptable statistical teleological generalizations may exist, and that they may be licensed by statistical, rather than principled, connections – such a result would be in line with Prasada et al.'s claim. But the generics in Experiment 1 all referred to a single, primary function (e.g., *knives are for cutting*), and for many kinds, such as *knives* and *eyes*, it is difficult to think of acceptable generalizations that refer to alternative functions. It may be that the vast majority of generalizations about functions that people consider in daily life refer to one primary function (as anticipated by Millikan, 1989).

However, one class of generalizations might challenge all existing accounts of generics: artifacts such as *plastic* are used for multiple purposes (e.g., *constructing packaging, storing food, making clothing*), none of which constitute a primary function. As the present theory suggests, no principled connection should exist between *plastic* and *storing food*, because *storing food* is not part of what it means to be plastic. But future studies should investigate whether or not people accept generics for non-primary functions, e.g., *plastic is for storing food.* If they reject such statements, then perhaps only principled connections yield acceptable teleological generics. If they accept them, then perhaps such generics operate akin to the statistical connections proposed by Prasada and colleagues (see Section 1.2).

We conclude by considering what is at stake in advancing a theory of teleological generics. Far from being a subsidiary phenomenon of teleological thinking, we argue that teleological generics represent its core. The way in which people interpret and produce generics sheds light on how they mentally represent functions and purposes, because generics express commonsense conceptions of objects and entities in the world. To our knowledge, the only theory of how the mind represents those conceptions for functional attributes is the one presented above.

The theory can be used to elucidate a wide variety of phenomena in teleological thinking. For instance, children seem to "promiscuously" endorse unacceptable generalizations, such as *mountains are for climbing* and *lions are for being in zoos* (Kelemen, 1999). Children's exposure to teleological generic language may play a key role in shaping mental representations of functional properties and the kinds of things to which they apply. Understanding those representations may help explain how adults learn to reject inaccurate generalizations. Likewise, researchers have proposed heuristics, such as structure-function fit (Liquin & Lombrozo, 2018), to explain the rapid assessment of teleological explanations. Even in adults, such rapid assessments can lead to endorsements of otherwise unwarranted explanations (Kelemen et al., 2013). If our theory of teleological generics is true, and if teleological explanations are composed from generics, then those heuristics must process the three mental representations – representations of kinds, properties, and connections – that we outline above. In sum, to understand teleological generics is to understand teleological thinking in general.

Author contribution statement

Joanna Korman: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing – Original draft preparation; Writing – Review & Editing.

Sangeet Khemlani: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing – Original draft preparation, Writing – Review & Editing, Supervision, Resources.

Acknowledgements

This work was supported by an NRC Research Associateship Award to JK and funding from the Office of Naval Research to SK. We thank Jon Baron, Krista Casler, Corey Cusimano, Noah Goodman, Geoff Goodwin, Hillary Harner, Tony Harrison, Laura Hiatt, Zach Horne, Deb Kelemen, Laura Kelly, Greg Murphy, Janani Prabhakar, M.H. Tessler, Greg Trafton, and three anonymous reviews for their insights and comments on earlier versions of the paper. We also thank Kalyan Gupta, Danielle Paterno, Kevin Zish, and the Knexus Research Corporation for their help in data collection, and the audience at the Annual Meeting of the Cognitive Science Society in 2018, where a preliminary version of this research was presented.

Declaration of competing interest

None.

Appendix A. The aggregated data from four norming studies as well as Experiments 1–4 reported above that were aggregated across the 44 different materials in each study and subjected to hierarchical regression analyses

Material	Norming studies					Outcome variable	Linguistic tests of principled connections					
	Study 1	Study 2	Study 3	Study 3	Study 4	Exp. 1	Exp. 1	Exp. 2a	Exp. 2a	Exp. 2b	Exp. 3	Exp. 4
	Causal history	Structure-function fit	Cue validity	Prevalence	Activity frequency	Teleological generics	Self-referential generalizations	Normality	Normativity (“supposed to”)	Normativity (“should”)	Aspect	Formal explanation
bags-storing	−0.02	5.60	50.67	57.05	0.52	1.68	1.65	0.51	0.79	1.70	2.16	0.92
bags-carrying	0.42	6.15	57.72	74.90	1.16	1.92	2.35	1.62	2.03	1.92	2.41	1.54
books-reading	1.70	6.23	67.03	93.03	1.54	2.80	2.69	2.44	2.44	2.66	2.76	1.79
books-packing	−1.46	3.07	17.51	15.44	−0.58	−2.20	−2.12	−1.79	−2.00	−1.86	−1.33	−1.60
brains-thinking	2.36	6.50	91.08	90.26	2.64	2.68	2.61	2.36	2.33	2.62	2.71	1.92
brains-sleeping	1.08	4.72	29.69	45.21	0.14	−1.72	−2.15	−1.21	−1.28	−1.92	−0.82	−1.25
chairs-sitting	1.36	6.29	66.74	88.15	2.10	2.80	2.69	2.26	2.72	2.54	2.82	1.83
chairs-dusting	−1.58	2.66	18.03	23.95	−1.34	−2.20	−2.35	−2.03	−2.18	−2.14	−1.55	−1.96
cars-driving	1.70	6.31	68.13	91.62	2.36	2.60	2.65	2.46	2.54	2.58	2.67	1.77
cars-painting	−1.58	2.23	10.67	26.77	−1.50	−2.16	−2.38	−1.72	−2.23	−2.34	−1.16	−2.27
cups-drinking	0.28	5.84	59.31	84.38	1.86	2.72	2.34	2.26	2.36	2.30	2.65	1.83
cups-stacking	−0.88	4.56	28.77	34.62	0.08	−1.48	−1.65	−1.44	−1.33	−1.10	−0.41	−0.75
ears-hearing	2.28	6.62	82.95	92.28	2.72	2.84	2.73	2.72	2.56	2.60	2.73	1.92
ears-plugging	−0.88	3.47	26.69	36.56	−1.08	−1.40	−2.04	−1.67	−1.85	−1.88	−0.94	−1.27
eyes-seeing	2.24	6.45	81.54	90.38	2.68	2.88	2.80	2.69	2.59	2.68	2.80	1.92
eyes-blinking	2.18	5.76	69.44	72.28	2.68	0.32	0.08	0.46	0.28	0.44	1.02	0.94
forks-eating	−0.28	6.07	40.56	84.00	1.76	2.64	2.42	2.00	2.00	2.16	2.37	1.90
forks-washing	−1.24	3.09	17.59	47.18	1.74	−1.80	−2.35	−1.03	−1.38	−1.60	−0.53	−1.75
hands-grasping	1.34	6.17	59.62	75.87	1.40	1.60	1.84	1.82	1.87	1.64	2.35	1.08
hands-clapping	1.24	6.01	71.85	62.08	−0.18	0.80	−0.04	−0.21	−0.23	0.64	1.51	0.81
knives-cutting	1.78	6.43	68.59	88.05	1.22	2.76	2.65	1.95	2.28	2.66	2.84	1.77
knives-sharpening	0.32	5.27	50.85	59.00	−0.92	−0.84	−0.19	−0.15	−0.51	−0.34	0.35	0.29
ladders-climbing	0.64	6.11	59.41	86.18	0.52	2.60	2.62	2.36	2.44	2.52	2.55	1.87
ladders-leaning	−1.12	4.25	28.26	31.46	0.14	−1.68	−1.96	−1.41	−1.13	−0.88	−0.57	−1.04
muscles-moving	1.32	5.94	41.28	74.08	2.44	1.84	1.92	1.95	1.79	1.74	1.82	0.56
muscles-flexing	1.26	5.64	65.15	58.97	1.40	0.00	−0.19	0.62	0.18	0.94	1.37	1.17
noses-smelling	2.24	6.54	86.05	91.87	2.68	2.52	2.38	2.38	2.26	2.46	2.61	1.73
noses-sniffing	1.94	6.00	87.31	83.05	1.64	1.60	1.65	1.59	1.49	1.52	2.24	1.19
pencils-writing	1.14	6.45	57.95	87.33	1.50	2.80	2.69	2.36	2.72	2.46	2.76	2.04
pencils-carrying	−1.98	2.62	18.46	26.13	0.10	−2.00	−2.12	−1.26	−1.49	−1.60	−0.98	−1.65
pianos-playing	0.36	5.94	30.77	87.82	0.14	2.36	2.73	2.51	2.64	2.48	2.53	1.35
pianos-tuning	0.40	5.00	38.56	50.85	−1.12	−0.20	−0.65	−0.26	−0.49	−0.56	0.63	−0.38
shoes-wearing	−0.06	5.62	33.54	90.21	2.04	2.36	2.26	2.72	2.59	2.22	2.47	1.17
shoes-polishing	−0.22	3.94	31.56	37.46	−1.18	−1.44	−1.23	−1.26	−1.62	−0.24	−0.24	−0.48
stomachs-digesting	2.20	6.43	84.82	91.28	2.56	2.56	2.46	2.49	2.51	2.56	2.76	1.52
stomachs-grumbling	0.60	4.78	42.28	47.46	0.20	−1.24	−1.53	−1.21	−1.38	−0.88	−0.41	−0.17
teeth-chewing	1.78	6.50	83.62	87.33	1.90	2.72	2.65	2.26	2.36	2.18	2.65	1.67
teeth-brushing	0.28	4.90	46.23	65.79	1.86	−0.80	−1.04	0.21	0.03	0.16	0.59	0.29
throats-swallowing	1.74	6.33	84.31	88.85	2.20	2.04	2.26	2.10	2.03	2.10	2.53	1.40
throats-coughing	0.88	5.13	69.90	59.05	−0.32	−0.36	−0.96	−0.18	−0.51	−0.14	0.43	0.15
tongues-tasting	2.00	6.35	79.95	85.03	2.18	2.44	2.27	2.23	2.36	2.38	2.63	1.62
tongues-licking	1.96	6.15	79.90	81.08	0.66	1.48	0.65	0.56	0.77	1.48	2.20	1.54
wings-flying	1.84	6.33	67.62	83.10	1.88	2.60	2.46	1.90	2.05	2.38	2.73	1.54
wings-flapping	1.26	5.80	54.97	75.41	1.72	0.84	0.85	0.92	1.08	1.10	1.59	1.17

Appendix B. Supplementary materials

Supplementary materials, including raw data, analysis scripts, code for the experiments, and materials, are available at https://osf.io/92qp3/.

References

Allen, C., Bekoff, A. C., & Lauder, G. (1998). *Nature's purposes: Analyses of function and design in biology.* Cambridge: MIT Press.

Atran, S. (1995). Causal constraints on categories. In D. Sperber, D. Premack, & A. J. Premack (Eds.). *Causal cognition: A multi-disciplinary debate*. Oxford, UK: Clarendon Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software, 67*(1), 1–48.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carlson, G. (1977). *Reference to kinds in English.* Amherst: University of Massachusetts (Ph.D. dissertation).

Casler, K., Terziyan, T., & Greene, K. (2009). Toddlers view artifact function normatively. *Cognitive Development, 24*(3), 240–247.

Champely, S. (2018). pwr: Basic functions for power analysis. *R package version, 1*, 2. https://CRAN.R-project.org/package=pwr.

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science, 34*, 1452–1482.

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition, 113*, 14–25.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science, 1*, 255–259.

DiYanni, C., & Kelemen, D. (2005). Time to get a new mountain? The role of function in children's conceptions of natural kinds. *Cognition, 97*, 327–335.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* New York: Oxford University Press.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8*, 404–409.

Gelman, S. A. (2010). Generics as a window onto young children's concepts. *Kinds, things, and stuff: Mass terms and generics* (pp. 100–120). .

Gelman, S. A., & Bloom, P. (2007). Developmental changes in the understanding of generics. *Cognition, 105*.

Gelman, S. A., & Brandone, A. C. (2010). Fast-mapping placeholders: Using words to talk about kinds. *Language Learning and Development, 6*(3), 223–240.

Hampton, J. (2009). Stability in concepts and evaluating the truth of generic statements. In F. J. Pelletier (Vol. Ed.), *Kinds, things, and stuff: Mass terms and generics. New directions in cognitive science. Vol. 12*. New York: Oxford University Press.

Haward, P., Wagner, L., Carey, S., & Prasada, S. (2018). The development of principled connections and kind representations. *Cognition, 176*, 255–268.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135–175.

Hussak, L. J., & Cimpian, A. (2018). Memory accessibility shapes explanation: Testing key claims of the inherence heuristic account. *Memory & Cognition, 46*(1), 68–88.

Keil, F. C. (1992). The origins of an autonomous biology. In M. R. Gunnar, & M. Maratsos (Vol. Eds.), *Modularity and constraints in language and cognition. Vol. 25*. Hillsdale, NJ: Erlbaum.

Kelemen, D. (1999). The scope of teleological thinking in preschool children. *Cognition, 70*, 241–272.

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *JEP: General, 142*, 1074–1083.

Khemlani, S., Leslie, S. J., & Glucksberg, S. (2012). Inferences about members of kinds: The generics hypothesis. *Journal of Language and Cognitive Processes, 27*, 887–900.

Keil, F. C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. A. Hirschfeld, & S. A. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). New York: Cambridge University Press.

Lederer, C. (1939). His Girl Friday. Retrieved from http://nldslab.soe.ucsc.edu/charactercreator/film_corpus/film_2012xxxx/imsdb.com/His-Girl-Friday.html.

Leslie, S.-J. (2007). Generics and the structure of the mind. *Philosophical Perspectives, 21*, 375–403.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical Review, 117*(1), 1–47.

Leslie, S.-J. (2012). Generics articulate default generalizations. *Recherches Linguistiques de Vincennes, 41*, 25–44.

Leslie, S.-J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology, 64*, 186–214.

Leslie, S.-J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? The generic overgeneralization effect. *Journal of Memory and Language, 65*, 15–31.

Liquin, E. G., & Lombrozo, T. (2018). Structure-function fit underlies the evaluation of teleological explanations. *Cognitive Psychology, 107*, 22–43.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167–204.

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science, 56*, 288–302.

Murphy, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin, 91*(1), 174.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation.* New York: Harcourt-Brace & World Inc.

Neander, K. (1991). The teleological notion of 'function' *Australasian Journal of Philosophy, 69*, 454–468.

Pelletier, F. J. (2009). Kinds, things, and stuff: Mass terms and generics. *New Directions in Cognitive Science* (Vol. 12). New York: Oxford University Press.

Perlman, M. D. (2004). The modern philosophical resurrection of teleology. *The Monist, 87*, 3–51.

Prasada (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences, 4*, 66–72.

Prasada, S. (1999). Names of things and stuff: An Aristotelian perspective. In R. Jackendoff, P. Bloom, & K. Wynn (Eds.). *Language, logic, and concepts: Essays in memory of John Macnamara*. Cambridge, MA: MIT Press.

Prasada, S. (2010). Conceptual representations and some forms of genericity. In F. J. Pelletier (Vol. Ed.), *Kinds, things, and stuff: Mass terms and generics. New directions in cognitive science. Vol. 12*. New York: Oxford University Press.

Prasada, S. (2012). Mechanisms for characterizing kinds and classes. *Recherches Linguistiques de Vincennes, 41*, 45–62.

Prasada, S. (2017). The scope of formal explanation. *Psychonomic Bulletin and Review, 24*, 1478–1487.

Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*, 73–112.

Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science, 33*, 401–448.

Prasada, S., Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition, 126*, 405–422.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.). *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Sober, E. (1984). *The nature of selection: Evolutionary theory in philosophical focus.* Chicago: University of Chicago Press.

Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review, 126*, 395–436.

Wright, L. (1976). *Teleological explanations.* Berkeley, CA: University of California Press.

Wright, L. (2013). Revisiting teleological explanations: Reflections three decades on. In P. Huneman (Ed.). *Functions: Selection and mechanisms* (pp. 363). Springer.