Neuroscience and Biobehavioral Reviews 81 (2017) 16–28

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Review article

Following new task instructions: Evidence for a dissociation between knowing and doing

Marcel Brass[a,*], Baptist Liefooghe[b], Senne Braem[a], Jan De Houwer[b]

[a] Department of Experimental Psychology, Ghent University, Belgium
[b] Department of Experimental Clinical and Health Psychology, Ghent University, Belgium

ARTICLE INFO

Article history:
Received 3 November 2016
Received in revised form 30 January 2017
Accepted 12 February 2017

Keywords:
Instruction following
Prepared reflex
Frontoparietal network
Dissociation of knowing and doing
Goal neglect

ABSTRACT

The ability to follow new instructions is crucial for acquiring behaviors and the cultural transmission of performance-related knowledge. In this article, we discuss the observation that successful instruction following seems to require both the capacity to understand verbal information, but also the ability to transform this information into a procedural format. Here we review the behavioral and neuroimaging literature on following new instructions and discuss how it contributes to our understanding of the functional mechanisms underlying instruction following. Based on this review, we distinguish three phases of instruction following. In the instruction phase, the declarative information of the task instruction is transformed into a task model consisting of a structured representation of the relevant condition-action rules. In the implementation phase, elements of this task model are transformed into a highly accessible state guiding behavior. In the application phase, the relevant condition-action rules are applied. We discuss the boundary conditions and capacity limits of these phases, determine their neural correlates, and relate them to recent models of working memory.

Contents

* Corresponding author at: Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000, Ghent, Belgium.
E-mail address: marcel.brass@ugent.be (M. Brass).

http://dx.doi.org/10.1016/j.neubiorev.2017.02.012
0149-7634/

1. Introduction

In their seminal paper, Nakahara et al. (2002) scanned both macaque monkeys and humans while carrying out a simple version of the Wisconsin Card sorting task, a common neuropsychological test of executive control. They found that similar brain regions were active in monkey and man, leading to the conclusion that performance in this task must be based on similar neurocognitive mechanisms. However, in a commentary on this study, Roepstorff and Frith (2004) raised the question whether one can actually compare the task-relevant processes between both species: while monkeys were trained for months to carry out this task, humans learned the task within a few minutes. This commentary pointed to a fundamental difference between humans and other species: while humans can execute a given instruction almost instantaneously, often without practice, any other species needs effortful trial and error learning to learn new tasks. Admittedly, studying learning via instructions in animals remains difficult as there will always be a language barrier between humans and non-human animals. Furthermore, there is some evidence for simple forms of learning new skills via instructions in some animals (Whiten et al., 1999). However, it is unquestionable that humans have a uniquely developed ability of instruction following that allows for easy cultural transmission of rules, forms the basis for most technological developments of modern societies, and separates them from other non-human animals. This raises the question why humans can follow instructions so easily while this ability is very restricted in other animals?

For one, it seems rather straightforward that our language capacity to represent and understand abstract content in a verbal format (Deacon, 1997) is vital to instruction following. However, as we will argue below, the ability to *understand* instructions is a necessary but insufficient condition to successfully follow instructions. Following new instructions not only requires understanding these instructions but also the translation of these instructions into actual behavior. For example, it is one thing to read and understand the instruction manual of your new smartphone while it is another to actually operate it. Such dissociation between understanding instructions ('knowing') and following instructions ("doing") has been first proposed more than half a century ago by demonstrating that frontal patients sometimes fail to follow instructions even though they are perfectly able to recapitulate what they were supposed to do (Milner, 1963).

Whereas the dissociation between knowing and doing seems straightforward at first, understanding the neurocognitive dynamics at the origin of this dissociation has become a major challenge in recent years (Demanet et al., 2016; Duncan et al., 1996; Duncan et al., 2008; Liefooghe et al., 2012; Muhle-Karbe et al., 2016). Accordingly, the aim of the current review is to provide an overview of the current state of knowledge on the dissociation between knowing and doing. To this end, three research domains are considered. First, research on 'goal neglect', which argues that participants sometimes fail to implement specific instructions even though they are perfectly able to remember them (Bhandari and Duncan, 2014; Duncan et al., 1996; Duncan et al., 2008). Second, behavioral research on the 'prepared reflex' (Hommel, 2000) or 'intention-based reflexivity' (Meiran et al., 2012) which examines the automatic effect of instructions to respond to stimuli. One important question within this line of research is whether instruction-based automatic response activation depends on the intention to implement a specific instruction or whether it is enough to simply remember the instruction (Liefooghe et al., 2012). Finally, we review brain imaging research, which tried to reveal the functional neuroanatomy of instruction following (Brass et al., 2009; Hartstra et al., 2011; Ruge and Wolfensteller, 2010) and attempted to dissociate between the implementation and memorization of instructions (Demanet et al., 2016; Muhle-Karbe et al., 2016).

Based on this literature review we will argue that instruction following can be decomposed in three different phases: the instruction phase, the implementation phase and the application phase. The instruction phase refers to the translation of the instruction into a task model. Research on goal neglect and neuroimaging research on complex rule following has extensively investigated this phase (Cole et al., 2010; Duncan et al., 2008). The implementation phase refers to active maintenance of specific aspects of the task model that need to be implemented. The literature on instruction-based congruency and some imaging studies have investigated this phase (Liefooghe et al., 2012; Muhle-Karbe et al., 2016). Finally, the application phase refers to the execution of the instruction. While the application phase is not at the core of the current review, we discuss some interesting findings that are relevant for our broader understanding of instruction following (Bhandari and Duncan, 2014; Ruge and Wolfensteller, 2010).

2. The study of goal neglect

First evidence for the idea that instruction following goes beyond instruction understanding stems from neuropsychological research in prefrontal patients and refers to the dissociation of 'knowing and doing' (Luria, 1980; Milner, 1963). Milner (1963) reported that her frontal leucotomy patients accompany their incorrect actions with correct verbal comments. Teuber (1964) referred to this as the 'curious dissociation of knowing and doing' (page, 333). According to Luria (1980), this dissociation between knowing and doing is neither caused by a lack of instruction understanding nor by motor deficits. While these findings have been discussed in the literature for decades, little systematic research was conducted to further understand the neurocognitive mechanisms that underlie this dissociation.

2.1. Goal neglect in the cognitive literature

In order to fill this empirical gap, Duncan and colleagues (Duncan et al., 1995; Duncan et al., 1996) introduced the concept of *goal neglect* which tried to capture the dissociation between knowing and doing on an experimental level. Goal neglect is defined by three properties (Bhandari and Duncan, 2014): (a) it reflects a gross failure to follow task rules; (b) performance is limited by the complexity of task instructions rather than by the complexity of task execution; and (c) performance is not explained by a failure of explicit rule recall.

Duncan et al. (1996) were the first to investigate goal neglect by using a letter-monitoring task (Fig. 1). In this task, a pair of letters or a pair of numbers is presented in each trial. One character is presented on the left side of the screen, the other on the right side of the screen. At the onset of the task, participants are cued which screen side is relevant and they have to read out loud the letters that are presented on that side. Digits on the same side and letters on the other side have to be ignored (Fig. 1). After a few trials, participants receive a symbol that either indicates that they have to switch to the other side or continue the task on the same side. Duncan et al. (1996) observed that some participants did not switch to the other side when they were required to do so, even though they were able to repeat the instructions verbally at the end of the task. Duncan et al. (1996) interpreted this failure to follow the instruction as 'goal neglect' and related it to general intelligence (g) and frontal brain damage. In order to investigate whether goal neglect also depended on task difficulty, a secondary task was introduced during the letter-monitoring task. During the stream of character pairs a dot could briefly flash either above or below the pairs and