THE “SPATIAL TURING TEST”

Stephan Winter and Yunhui Wu
Department of Geomatics, The University of Melbourne, Australia
winter@unimelb.edu.au, y.wu21@pgrad.unimelb.edu.au

Abstract

How far are current route planners from imitating humans in giving route directions? And what can we learn, in terms of gaps of knowledge and need for further research, from this question?

1. INTRODUCTION

Turing starts his landmark paper Computing Machinery and Intelligence with the words: “I propose to consider the question, ‘Can machines think?’” (1950, p. 1). He approaches the question by an anthropomorphic imitation game: a game in which a person is supposed to find out whether it is communicating with a machine or another person.

In this paper we will borrow from this idea, and investigate the question ‘Can machines think spatially’ to an extend that a person, in search for some spatial information, cannot find out from the response whether they are communicating with a machine or another person.

The investigation will be along a case study: the interactions of persons with a web-based route planner. The underlying question (“Can you tell me the way to …?”) leads to typical everyday communication, either with other persons or with a computing machine in form of dedicated route planners. Still these questions form quite a challenge for computing machines, despite their apparently simple interface (Figure 1) and their across platforms similar response styles.

Figure 1: User interaction with Melbourne’s public transport planner.
The paper provides a review of recent work, a critical investigation of gaps, and it develops a research program for the future.

2. BACKGROUND

The Turing test of whether machines can think is an imitation game: a computer has to convince players that they are communicating with a person. By this way Turing avoids to define thinking or intelligence. Instead he suggests a test of anthropomorphic qualities in the response of computers. In fact, this approach includes imitating some behaviors that people do but that are not particularly intelligent. It also does not allow for superhuman qualities: a computer that can solve problems people cannot solve would fail Turing’s test.

A spatial Turing test is only sensible if we accept a priori that human communication about space is intelligent. Human communication has qualities challenging computers, such as adapting to the communication situation, or talking about places. Research demonstrates that cognitively ergonomic route directions (as for example given by persons) have some redundancy (Richter and Klippel 2005), have varying, if not asymmetric granularity (Klippel et al. 2003; Tenbrink and Winter submitted; Tomko and Winter submitted), relate to the perceptual experience of the environment (Denis 1997), and refer to a selected range of features depending on the traveling mode (Maaß 1996; Burnett 2000). Current route planners do not yet imitate persons in these ways. They also solve the problem of finding a route based on spatial conceptualizations that are different from people’s conceptualizations. Accordingly, in the communication of these routes ontological mismatches will be expected.

However, Turing’s interest was a philosophical one, on the nature of artificial intelligence. His test does not facilitate a real measure of intelligence of programs, nor does it provide an objective way of testing. In this sense, we will be able to identify gaps leading to failure of the Turing test, but we will not be able to measure the quality of a route planner.

3. COMMUNICATING WITH A TRANSPORT PLANNER

We have chosen Melbourne’s public transport planner for the case study. This state-of-the-art planner is based on the EFA software by Mentz DV (Einert 2006), which is also broadly used by German and Austrian transportation networks. We call it planner for short in the remainder of
In this paper, the planner is able to make travel plans across various transportation modes, such as trains, trams and buses, and allows either walking or cycling to cover the transfers.

In this case study, a user plans a typical trip, here from home to work, using the planner. After being provided with a travel plan, the user will follow the plan and examine how well the plan helps her to complete the trip.

The communication between the user and the planner starts on the planner’s homepage (Figure 1). The user has to choose between Station/Stop, Address and Landmark to locate from and to locations, as well as either the departure or the arrival time. In general, the distinction of location types is not one that people make, and they even have difficulties to say whether a reference such as university is a stop name or a landmark (in fact, it can be both, and it can be ambiguous as well), or what the address of their particular location on the campus of the university is. Here, the spatial Turing test fails already.

In the present case, the user knows the addresses of home and work, and requests a trip from 9 Beaumonde Street, Coburg to 1 Southbank Boulevard, Southbank arriving at 9:00am on 19 March, 2008. The planner does not recognize the second address, but accepts 1 Southbank Bvd, Southbank. All the guesswork was on the user’s side. While abbreviations of common street types are managed properly by the planner, the rarely used type Boulevard and its abbreviation seem to be missing in the table of synonyms.

The planner then provides five travel options, and highlights the one closest to the specified arrival time. Brief information on each option, such as trip duration and numbers of transfers, is also available to assist decisions. Once the user picks the highlighted one, the planner provides more details of the selected travel option (Figure 2).

The trip plan instructs the user to walk about 570 meters from the start to (tram) Stop 129, then take Route 8 towards Toorak at 8:03am, get off at Stop 13 and walk about 360 meters to the destination. Turn-by-turn route descriptions (TBT) and leg maps (LM) are offered for more detail on the walking parts. Stop maps (SM) are offered for more detail at transfer points, but are almost the same as LM.

However, TBT (expanded in the trip plan of Figure 2) show leg lengths that are inconsistent with the detailed trip plan due to rounding errors (570m vs. 560m, and 360m vs. 350m), which raises confusion on the user’s side. The segment on Flinders Walk is 30m in TBT while 20m in LM (in reality it is 220m—compare this segment with the segment on Southbank Prom on the LM—, which is a different issue). Obviously the
planner uses different algorithms in different modules to compute leg lengths. Furthermore, TBT lack an orientation at the start: *Continue along* is ambiguous. For example, when arriving at Stop 13, which is on *Swanston St*, it is left to the user to find *Flinders Walk* to realize the first instruction *Continue along Flinders Walk*. This example is particularly difficult because *Flinders Walk* can be reached only via stairs from Stop 13. The stairs have to be in the database as a link, although they are not part of *Flinders Walk*, and they are not mentioned.

![Figure 2: The computed trip plan for the recommended travel option.](image)

Also, TBT and LM use a different language (compare with Figure 3). For example, TBT read *Continue along* and *Turn ... at*, while LM read *Straight Up* and *Turn ... Into*. Similarly, the TBT is precise in its reference to a tram stop by *Stop 13-Flinders Street Railway Station / Swanston St*, while the LM is titled *from Flinders Street Railway Station (Melbourne City)*, which verbally refers to the train station, although the shown leg starts correctly at the tram stop. Furthermore, the LM maps the leg by a polyline, but fails to label or characterize the named segments of the TBT (e.g., *Flinders Walk, Southbank Prom*), which
produces lack of correspondence. Additionally, TBT refers to an *Unnamed Road*, which after visual inspection of the LM turns out to be a bridge. Ontological mismatches of this sort in the computer’s database were discussed already by Raubal and Kuhn (2004); they do not only confuse a user but also neglect to take advantage of the salience of, in Lynch’s terminology (1960), a path—the bridge—crossing an edge—the river. Another language conflict concerns the characterization of the first segment of the leg in the LM by a dotted line, which is a symbolization that does not exist in the legend.

Realizing the given plan, the user will experience mismatches between the provided information, the information in the environment, the perceived affordance of the environment, and the layout of the environment. Among these are the already mentioned mismatches between *Flinders Walk* and stairs, and between *Unnamed Road* and a bridge. Taking the stairs is also prohibited by a negative affordance: the
stairs can be reached from Stop 13 only through a part of the train station building.

4. **Discussion**

In summary, the experiment demonstrates fundamental mismatches between provider ontologies and user ontologies, as well as between provider language and user language. Inconsistent computing results, inconsistent language and lack of correspondences between verbal and graphical communication means, and also reliance on metric measures unmask the computing machine as communication partner. Metric measures almost never occur in human route directions (Daniel and Denis 1998). Where leg maps show a lack of detail for relevant information (stairs, segment names), they show irrelevant details. Human-generated sketch maps look different (Blades 1990).

A full paper will elaborate the research questions arising from these and more observations in the case study.

**References**

Tenbrink, T.; Winter, S., submitted: Granularity in Route Directions.