

Language, Duration Estimation, and Causation

DANIEL B. WRIGHT^{1*} and ALICE C. I. PEDERSEN²

¹*University of Sussex, UK*

²*Institute of Education, UK*

SUMMARY

Burt (2002) raises several interesting points in his commentary on our paper. However, the most important difference between his and our studies, the design, was not addressed. We argue that our experimental approach makes more realistic assumptions for causal interpretation. The other points that he raises—the relationship between practical and statistical significance, the verbs used by our participants, and definitions of memory—are all discussed. Copyright © 2002 John Wiley & Sons, Ltd.

We agree with Burt's (2002) final sentence: 'the jury must remain "out" on the impact that language use has on duration estimation', but arrived at the conclusion differently. While this appears a relatively simple hypothesis, it is in fact a very difficult hypothesis to evaluate with confidence. In this reply we describe why our experimental approach is a better method to assess whether there is a causal relationship between language use and duration estimation than Burt's (1999, Experiment 2) non-experimental approach. Burt's approach is better for evaluating other types of hypotheses, and we make clear this importance. The idea that language use affects duration estimation is appealing on theoretical grounds and is of great practical importance, but is one not demonstrated by Burt's study. Our experiments were designed to test whether language use did affect duration estimation. However, the results showed that the jury is still out on whether language use affects duration estimation. Before discussing the difficulties in attributing causation we respond to three other points raised by Burt: the relationship between practical and statistical significance, the impact level of the verbs used by our participants, and our views about reconstructive memory.

POWER: PRACTICAL AND STATISTICAL SIGNIFICANCE

Burt (1999, Experiment 2) showed participants a bank robbery and then had them describe the event. He found a statistically significant association (r^2 of 17%, 95% CI = 0%–48%) between the number of action words people used and their duration estimates (in one of two conditions). He argues that our findings, while not statistically significant, could still be of practical significance. However, by considering the 95% confidence interval rather than simply the dichotomous accept/reject significance testing approach, as the APA task force recommends (Wilkinson *et al.*, 1999), it is clear that Burt's study did not show that

*Correspondence to: Daniel B. Wright, Psychology Group, Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton BN1 9QH, UK. E-mail: DanW@cogs.susx.ac.uk

the effect was of practical significance. It showed that the effect *might* be of practical significance and that it *might* be of no practical significance. Of course, a confidence interval on its own should not be over-interpreted. Replication using different scenarios provides further important evidence on the size and robustness of any effect (Wright, 2002). This was made clear in Pedersen and Wright (2002) by looking for systematic relationships across several effects in order to reach any conclusions.

Some of the arguments regarding the relationship between effect size and practical significance in Burt (2002) could be misinterpreted through his example. First, Burt notes that a correlation of -0.41 was found (in one condition), which equates with 17% of shared variance. Next, he describes that a 10% reduction in duration is a 'conservative' estimate. It is unclear where this estimate comes from, but because it is mentioned right after reporting the correlation readers may think that one is derived from the other. The 17% refers to a percentage of shared variance, not a percentage reduction. There could be a 1% reduction that accounts for 50% of the variance, and a 50% reduction that accounts for 1% of the variance. It depends how spread out the data are around the regression line. The 6-second reduction per minute refers to a slope of the regression line, and cannot be derived from the correlation alone. There is another problem with this example for Burt's hypothesis. Burt does not mention the person using any language. The hypothesis under consideration is not that people who experience events that include many activities experience time differently from those who experience few activities. Ornstein (1969) and others have conclusively shown this.

When there is concern about practical and statistical significance, power analyses are usually conducted to help determine the sample size. Pedersen and Wright (2002) discuss how concern for power helped determined the sample sizes for their studies. We do note that there was a slight tendency for effects to be in the direction of Burt's hypothesis, but this level of evidence is not enough to conclude that the effect is either present or of practical significance.

ESTIMATING IMPLIED ACTION

Burt (2002) was critical that Pedersen and Wright's (2002) participants were 'using verbs which implied very little in the way of behavioral velocity'. We found this comment curious for several reasons. First, while Burt (1999) mentioned that the verbs used by his participants varied in their implied action, he failed to differentiate verbs on their impact. We choose to differentiate and think that this is vital for studies of this sort. Second, to say that the scores are low it is necessary to say with what they are low in comparison. If they turned out to be much lower or higher than the norm we would want to take care generalizing to other event descriptions, but without comparison values we cannot say whether they are much lower or higher. Finally, it is not the overall level that is of interest but the differences among the conditions. The technique we used to create differences among the conditions did work. We would have preferred larger differences, and encourage others to develop methods which create larger differences in language usage.

THEORIES OF MEMORY

There are two points Burt raises about memory which we address. The first is: 'the words used to describe an event *are* the reconstructed elements of the event memory' (*italics* in

the original). This is a narrow definition of memory, which excludes, for example, human infants and non-speaking animals from having memories. In our Figure 1 (Pedersen and Wright, 2002) we very purposefully separated memory and language. Language is a method that can be used to describe a memory, but most memories are not described. This does not mean that language cannot affect with memory, as it does, for example, with verbal overshadowing (Schooler and Engstler-Schooler, 1990). The second issue is that a memory exists only at the time of reconstruction. Work on the postevent information effect (e.g. Wright *et al.*, 2001) shows that memory can be altered prior to being recalled. It is a matter of definition if one wishes to restrict the phrase 'memory' to apply only to conscious recollections.

ATTRIBUTING CAUSATION

Our central disagreement with Burt (2002) is the type of inference that he is claiming from his 1999 paper: 'the verbs an individual uses to describe an experience can influence their estimate of the experience's duration'. This is a causal hypothesis because of the word 'influence'. Usually this type of hypothesis is framed as a *ceteris paribus*, or all other things being equal, hypothesis (Fodor, 1991). The first question is whether it makes sense to say that language use affects duration estimation. Some law-like phrases do not. Using Rubin's model of causation: 'causes are only those things that could, in principle, be treatments in experiments' (Holland, 1986, p. 954). Language use, in principle, could be experimentally manipulated, thus the hypothesis seems valid to consider.

Neither Burt (1999) nor Pedersen and Wright (2002) can directly evaluate this hypothesis without making some assumptions. Here we argue that the assumptions needed for Pedersen and Wright's experiments are more tenable. Next we describe the hypotheses that Burt's and Pedersen and Wright's designs do directly evaluate. Burt's (1999) design was non-experimental. We do not take a hard-line view that experiments are necessary to make causal statements, otherwise many of the causal statements made in astronomy, anthropology, psychiatry, etc. would be inappropriate. However, experiments make it easier to make causal statements: 'random assignment is the great *ceteris paribus*—that is, other things being equal—of causal inference' (*italics*, in the original, Cook and Campbell, 1979, p. 5). In order for Burt to reach his causal conclusion he must assume that at each level of action word use, the people are otherwise the same. Clearly some trivial differences which are highly unlikely to affect duration estimation should be admissible, but if there are variables that are likely to influence both language use and duration estimation then this can produce a spurious correlation (see Figure 1 in Pedersen and Wright, 2002). People will have different memories for the same event and one difference will be the remembered speed of the event. Burt is implicitly arguing that differences along this dimension would not be associated with differences in action word use and duration estimation. We do not feel that this assumption is tenable, and is why we ran our experiments.

In order for Pedersen and Wright's (2002) experiment to be able to assess the hypothesis, language use influences duration estimation, we have to assume that our manipulation affected language use but nothing else. This is the *ceteris paribus* part of the hypothesis. Again, trivial differences should be acceptable to most. The important thing is that none of the differences, other than language use, affect duration estimation. We are more willing to accept this assumption.

Both Burt (1999) and Pedersen and Wright (2002) are testing important hypotheses that do not rely on these assumptions. Burt (2002) notes that: 'it seems likely that individuals

who recall (reconstruct) a description of an event that varies in the action verbs used might generate rather different estimates of the event's duration'. Finding a correlation is testing this hypothesis. This is an important hypothesis as courts and police investigators may wish to assess the accuracy of a duration estimate by content analysing the event description. Further research would be necessary on this, but Burt (1999) is a good start. Pedersen and Wright's design directly assesses whether a manipulation, like how the person is told to describe the event, can affect duration estimation. This also is important as often eyewitnesses are asked to describe an event in different ways, for example the cognitive interview, to the police officer at the scene, etc. If future research confirms that there is little or no effect of this on duration estimation, then this is also of applied importance for scientific expert testimony.

SUMMARY

Burt's (2002) commentary on Pedersen and Wright (2002) raised several interesting points, but failed to discuss the main difference between the conclusions that could be reached from Burt's (1999) correlational study and our experiments. The assumptions necessary to reach causal conclusions from Burt's study are untenable. To assume that some form of temporal memory does not affect the language used to describe an event and does not affect duration estimation seems to go against much of Burt's research over the past decade. The assumption necessary to reach causal conclusions about the relationship between language use and duration estimation for our experiments do seem acceptable. Of course, these are assumptions so others may feel differently and we welcome researchers to test these assumptions. The take-home message of Pedersen and Wright (2002) was that Burt's (1999) study did not show whether there was a causal link between language use and duration estimation. After running three studies, we still feel this link has not been established, and thus agree that the jury is still out.

REFERENCES

- Burt CDB. 1999. Categorisation of action speed and estimated event duration. *Memory* **7**: 345–355.
- Burt CDB. 2002. A commentary on Pedersen and Wright (2002). *Applied Cognitive Psychology* **16**: 785–787.
- Cook TD, Campbell DT. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin Company: Boston, MA.
- Fodor JA. 1991. You can fool some of the people all of the time, everything else being equal, hedged laws and psychological explanations. *Mind* **100**: 19–34.
- Holland PW. 1986. Statistics and causal inference. *Journal of American Statistical Association* **81**: 945–960.
- Ornstein RE. 1969. *On the Experience of Time*. Penguin Books: Harmondsworth.
- Pedersen ACI, Wright DB. 2002. Do differences in event descriptions cause differences in duration estimates? *Applied Cognitive Psychology* **16**: 769–783.
- Schooler JW, Engstler-Schooler TY. 1990. Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology* **22**: 36–71.
- Wilkinson L, the Task Force on Statistical Inference, APA Board of Scientific Affairs. 1999. Statistical methods in psychology journals: guidelines and explanations. *American Psychologist* **54**: 594–604.
- Wright DB. 2002. *First Steps in Statistics*. Sage Publications: London.
- Wright DB, Loftus EF, Hall M. 2001. Now you see it; now you don't: inhibiting recall and recognition of scenes. *Applied Cognitive Psychology* **15**: 471–482.