# The psychological mechanism of the slippery slope argument

Adam Corner [a,*], Ulrike Hahn [a], Mike Oaksford [b]

[a] School of Psychology, Cardiff University, United Kingdom
[b] School of Psychology, Birkbeck College, University of London, United Kingdom

A B S T R A C T

Slippery slope arguments (SSAs) have a bad philosophical reputation. They seem, however, to be widely used and frequently accepted in many legal, political, and ethical contexts. Hahn and Oaksford (2007) argued that distinguishing strong and weak SSAs may have a rational basis in Bayesian decision theory. In this paper three experiments investigated the mechanism of the slippery slope showing that they may have an objective basis in category boundary re-appraisal. When the beginning and the end of a slippery slope are more similar, the probability that they are perceived to belong in the same category is higher and the SSA is stronger. Experiment 1 established a robust effect of probability on SSA evaluation. Experiments 2 and 2A showed that when similar items are classified in the same category this leads to stronger SSAs. In Experiment 3, in a correlational analysis, it was shown that participants' confidence in their categorisation judgements predicted the perceived strength of an SSA and that this relationship was moderated by similarity between the ends of the slippery slope. We conclude that an important aspect of many SSAs may have an objective basis in well-established and rational cognitive theories.

© 2010 Elsevier Inc. All rights reserved.

## Introduction

The 'slippery slope' is an intuitive metaphor that is used to refer to a class of arguments with a distinctive form, but varied content. Classified as a fallacy of reasoning in most critical thinking textbooks (Woods, Irvine, & Walton, 2004) and by many philosophers (e.g., Enoch, 2001), yet frequently used and widely accepted in applied domains such as politics (van Der Burg, 1991), law (Lode, 1999) and bio-ethics (Lamb, 1988; Launis, 2002), the slippery slope argument (SSA) is a controversial topic in the field of argumentation. For most, the argument possesses the somewhat undignified status of "wrong but persuasive", and therefore fits neatly into the category of arguments that argumentation theorists call fallacies (although see Corner & Hahn, 2007; Hahn & Oaksford, 2007).

Examples of SSAs show that they can vary greatly in strength:

1. If we allow gay marriage, then in the future people will want to marry their pets.
2. If voluntary euthanasia is legalised, then in the future there will be more cases of 'medical murder'.
3. If we accept voluntary ID cards in the UK, we will end up with compulsory ID cards in the future.

Few would agree that homosexual marriages are the beginning of a slippery slope to inter-species marriages, although this argument was put forward by a group called the American Family Research Council (2004). Example (2) seems more plausible, although not sufficiently plausible to prevent the Dutch Government from legalising certain forms of voluntary euthanasia. In (3), it seems extremely likely that ID cards in the UK would become compulsory if they were introduced—in fact, if they were to function as an effective security measure, this would be a necessity. From the dubious logic of (1), through the calculated

 * Corresponding author. Address: School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom.
   E-mail address: corneraj@cardiff.ac.uk (A. Corner).

risk-taking of (2), to the almost inevitable consequence of (3), SSAs display an impressive variation in their persuasiveness.

Recently, Hahn and Oaksford (2007) provided a Bayesian analysis of the fallacies, including SSAs, arguing that many could be viewed as inductive arguments of varying strength, depending on a variety of Bayesian factors. The slippery slope was regarded as a form of consequentialist argument whose conclusion was the denial of the antecedent of the conditional sentences (i.e., *if…then* sentences) used to frame the above examples. Each of (1)–(3) is expressed using a conditional, *if antecedent* (A), *then consequent* (C), and the conclusion people are invited to draw is not-A, e.g., one should not allow gay marriage. The strength of the argument depends on the probability that the antecedent (A) would lead to the consequent (C), i.e., inter-species marriage, (Pr(C|A)), and the utility of the consequent, *U*(C). The higher Pr(C|A) and the more negative *U*(C), the more persuasive a slippery slope argument should be that the action in A should not be taken. Example (1) is weak because of the very low value of Pr(C|A) whatever we may think of the merits of inter-species marriage. Example (2) is stronger because this probability is higher but also because 'medical murder' is clearly so undesirable. Example (3) is even stronger because Pr(C|A) seems very close to 1 and the consequent is highly undesirable (for some).

Framing (1)–(3) in terms of conditionals relates SSAs directly to the consequential conditional reasoning investigated by Bonnefon and Hilton (2004) and to utility conditionals more generally (Bonnefon, 2009; Evans, Neilens, Handley, & Over, 2008). Evans et al. (2008) investigated a variety of conditionals expressing conditional tips, warnings, threats, and promises. The most relevant to SSAs are warnings, e.g., "If you go camping this weekend then it will rain," which clearly depend on all the same factors. So, the higher Pr(C|A) and the more negative *U*(C), i.e., rain, the more persuasive a conditional threat is to the conclusion that action A should not be taken, i.e., you should not go camping. What differs between SSAs and warnings is that whereas in the latter the probability is assessed just by reference to prior world knowledge, in the former there seems to be an implied mechanism that leads to the consequent action from the antecedent action. This mechanism suggests that an act of categorising an item *a* (gay couples) under a category *F* (can marry), i.e., *Fa*, will lead to other items *b* (inter-species "couples") also falling under the same category, *Fb*. Hahn and Oaksford (2007) proposed that such a "category boundary re-appraisal" mechanism may explain why people find slippery slope arguments so compelling.

It is fundamental to a wide range of current theories of conceptual structure that encountering instances of a category at the category boundary will extend that boundary for subsequent classifications. Furthermore there is a wealth of empirical evidence consistent with these assumptions. In particular there are numerous experimental demonstrations of so-called exemplar effects, that is, effects of exposure to particular instances and their consequences for subsequent classification behaviour (e.g., Lamberts, 1995; Nosofsky, 1986, 1988a, 1988b). For example, observing that a dog that weighs 10 kg is considered underweight invites the conclusion that a dog that weighs 10.5 kg is also underweight. With only the information that a 5 kg dog is underweight, and a 15 kg dog is overweight, however, one might not be so compelled to draw this conclusion.[1] This is because of the similarity between 10 kg and 10.5 kg and the comparative dissimilarity with either 5 kg or 15 kg. Similarly, one may argue that (1) is a poor argument and so Pr(C|A) is low because of the dissimilarity between same sex human relations and inter-species relations and hence it is clear that there is no likelihood of slippage of the category "can marry" from one case to the other. In sum, there are a range of factors identified in the psychology of inductive reasoning that lead to a greater probability of placing an exemplar in the same category as a previously observed exemplar.

The purpose of the experiments we report in this paper was to directly test the predictions of this putative mechanism of the slippery slope by investigating the relationship between category boundary judgements and slippery slope arguments. We first introduce SSAs in more depth before turning to the justification for Hahn and Oaksford's (2007) decision-theoretic approach. We then discuss the motivation for a category boundary re-appraisal mechanism for the slippery slope. Hahn and Oaksford (2007) discussed the only experiment (from Corner, Hahn, & Oaksford, 2006) demonstrating the effects of utility and probability on slippery slope arguments. However, that experiment was under analyzed being only a brief conference proceedings report. Prior to investigating the underlying mechanism of the SSA, it is important to establish that they are indeed influenced by probabilities and utilities and how these effects relate to Evans et al.'s (2008) results on warnings. Three further experiments directly investigated the link between category boundary re-appraisal and slippery slope arguments using a uni-dimensional, quantitative category and numerically defined exemplars both in a between (Experiment 2) and a within subjects design (Experiment 2A) and using qualitative categories (Experiment 3). If SSAs have an objective basis in category boundary re-appraisal, there should be agreement between the perception of an SSA's strength and corresponding categorisation decisions, given identical data to evaluate.

## What are SSAs?

Whilst it is simple to produce an intuitive characterisation of SSAs, they have resisted attempts to provide a comprehensive definition. As Rizzo and Whitman (2003, p. 544) put it, "there is no paradigm case of *the* slippery slope argument." Authors have typically opted either to differentiate multiple independent forms of SSA (e.g., Walton, 1992) or to treat only a very select group of arguments as genuine examples of SSAs (e.g., Govier, 1982). Walton (1992), for example, distinguishes four types of SSA, suggesting that some SSAs involve causal mechanisms

---

[1] These examples of similarity-based categorisation might be viewed as a specific instance of the more general process of similarity-based induction – see, e.g., Rips (1989), Sloman (1993), or Osherson, Smith, Wilkie, Lòpez, and Shafir (1990). This process in turn might be given a Bayesian interpretation (Heit, 1998; Kemp & Tenenbaum, 2009).