Journal of Memory and Language 64 (2011) 133–152

Contents lists available at ScienceDirect

Journal of Memory and Language

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

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

ARTICLE INFO

Article history:
Received 15 November 2009
Revision received 19 October 2010
Available online 18 November 2010

Keywords:
Argumentation
Bayesian
Slippery slope
Reasoning
Categorisation

ABSTRACT

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.

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).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.10.002

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).

('causal' SSAs), some set precedents ('precedent' SSAs), while others are attributable to the vagueness of concepts and categories ('sorites' SSAs). A fourth type combines features from each of these SSAs ('full' SSAs). Other authors have opted to avoid such a detailed taxonomy, choosing instead to list 'core features' that SSAs generally seem to possess. For example, Rizzo and Whitman (2003) identified three components they claim are common to all SSAs; (1) an initial, seemingly acceptable, decision; (2) a 'danger case' that is clearly unacceptable; and (3) a process or mechanism by which the initial decision will raise the likelihood of the danger case.

We will attempt, however, to give a definition of SSAs that is useful from a psychological perspective – that is, useful specifically for a psychological analysis of SSA strength. SSAs are a particular breed of *consequentialist* argument (Hahn & Oaksford, 2007; for analysis of SSAs as consequentialist arguments see Oakley & Cocking, 2005; Walton, 1992; for recent experimental work on other forms of consequential conditional see Bonnefon & Hilton, 2004; Thompson, Evans, & Handley, 2005; and for a recent approach to 'utility conditionals' in general, Bonnefon, 2009). A dissuasive consequentialist argument (or deterrent) warns against a particular course of action on the grounds that it will lead to an undesirable outcome, or consequence. An SSA, however, posits not only a negative outcome but the idea that this outcome might in the future be *re-evaluated as positive*, if an initial proposal goes ahead.

A general consequentialist argument might oppose the legalisation of cannabis because it would lead to an increase in smoking related respiratory problems. A slippery slope argument would oppose legalisation on the grounds that attitudes towards harder drugs might become more positive in the process, and in the future a substance like cocaine might also become legal. This gives the slippery slope four distinct components:

i. An initial proposal (A).
ii. An undesirable outcome (C).
iii. The belief that allowing (A) will lead to a re-evaluation of (C) in the future.
iv. The rejection of (A) based on this belief.[2]

The alleged danger lurking on the slippery slope is the fear that a presently unacceptable proposal (C) will (by any number of psychological processes—see, e.g., Volokh, 2003) in the future be re-evaluated as acceptable. If we withhold the right to free speech from a neo-Nazi organisation, what will prevent us from censoring legitimate political dissent in the future? The proponent of this argument is inherently appealing to the malleability of public opinion to reject an otherwise appealing course of action. The uncertainty of the future is such that any reasoning about it is at best presumptive. Yet SSAs trade on the uncertainty of the future, and appear to be acceptable in a number of contexts (Lode, 1999; Volokh, 2003). The main purpose of the experiments we report here is to investigate one possible psychological mechanism that leads to the belief in (iii).

The decision-theoretic approach to SSAs

Unlike the fallacies of formal logic, informal arguments typically do not go wrong by violating logical norms (Hahn & Oaksford, 2007; Ricco, 2007; van Eemeren & Grootendorst, 2004). There are many arguments that are logically valid but that are regarded as informally unacceptable: for example, "god exists, because god exists" though valid, provides insufficient reason for *changing* one's beliefs in the existence of god. Conversely, classical logic simply has nothing to say about the majority of informal arguments, including the slippery slope. Because of this discrepancy between logical validity and informal acceptability (and a wide variety of other considerations), there has been a widespread philosophical rejection of formal logic as providing either necessary or sufficient criteria for evaluating informal argumentation (Boger, 2005; Hamblin, 1970; Heysse, 1997; Johnson, 2000). This has been buttressed by mounting psychological evidence that people do not naturally or consistently reason according to the rules of formal logic (Evans, 2002; Oaksford & Chater, 2001).

The Bayesian approach to informal argumentation and the fallacies (Hahn & Oaksford, 2007) seeks to interpret reasoning patterns as probabilistic changes in subjective degrees of belief, using probability theory as a normative framework. Using this model of argument strength an SSA should be viewed as strong (and therefore convincing) to the extent that its consequences are made *probable* by the currently disputed initial proposal. The idea that the putative outcomes of SSAs are not as probable as their proponents claim is central to its reputation as a scare-mongering, fallacious argument (Oakley & Cocking, 2005).

However, perhaps the most important aspect of SSAs is that they advocate *decisions*, and as such are not just arguments about factual claims. Philosophers interested in applied domains such as law or bio-ethics where SSAs are popular have implicitly recognised that probabilistic and utilitarian concerns are crucial determinants of consequential and slippery slope argument acceptability (e.g., Holtug, 1993; Lode, 1999), which distinguishes SSAs from most other fallacies of argumentation (Woods et al., 2004). Holtug (1993), for example, claimed that in relation to SSA strength, "the more probable the causal connection is, and the more we want to avoid (B), the stronger the argument" (Holtug, 1993, p. 404). The tools of decision theory may provide a formal framework for these intuitions.

Bayesian decision theory (Keeney & Raiffa, 1976; Savage, 1954) is a normative framework for decision making in situations where outcomes are uncertain, based on the subjective probabilities and utilities involved. According to its prescriptions, agents should seek to maximize subjective expected utility (i.e., potential gain) in their

[2] This SSA definition is not designed to correspond to any one of the types of SSA identified by Walton (1992), although it might best be thought of as embodying the 'precedent' and 'sorites' SSAs in Walton's taxonomy. Seemingly essential philosophical distinctions do not necessarily map precisely on to psychological mechanisms – and while this does not undermine the usefulness of Walton's taxonomy, it is also clear that empirical, psychological work should not be bound by it.

choices. This means that different agents can rationally choose different courses of action if their respective assessments of probabilities and utilities differ. However, there is still a normative standard in operation, in that the evaluation of decisions by a given rational agent must be derivable from more fundamental valuations—namely the probabilities and utilities they assign. Specifically, the subjective expected utility of a decision (SEU) corresponds to the probability-weighted sum of the utilities associated with a particular course of action. Eq. (1) shows the *SEU* for any number of outcomes (x_i), where "Pr()" is the subjective probability and "*U*()" is the subjective utility of each outcome:

$$\mathrm{SEU} = \sum \Pr(x_i)U(x_i) \tag{1}$$

Given then, that SSAs advocate particular actions based on their putative consequences, this framework can be brought to bear directly on the assessment of their strength (Corner & Hahn, 2009; see also Over, Manktelow, and Hadjichristidis (2004) and Perham and Oaksford (2005) for related work using SEU to predict the acceptance of deontic conditionals). The higher the probability that some feared outcome will be brought about by an initial course of action and the greater the utility of avoiding this feared outcome, the stronger the SSA will be.

For warnings, Evans et al. (2008) argued that the decision about whether to perform the action described in the antecedents of (1)–(3) is based on the prima facie utility of the action itself, *U*(A), less the expected disutility of the action to which A could lead, Pr(C|A)*U*(C), i.e.,

$$U(\mathrm{A}) - \Pr(\mathrm{C}|\mathrm{A})U(\mathrm{C}) \tag{2}$$

This formulation also applies directly to SSAs. That is, the normative or computational level theories (Marr, 1982) driving both our research on SSAs and Evans et al.'s (2008) research on other utility conditionals are the same. Experiment 1 sought to examine the relation between Evans et al.'s (2008) results and data on varying both probability and utility for SSAs.

The mechanism of the slippery slope

What distinguishes a very important class of SSAs from warnings is the implied mechanism by which A can lead to the undesirable consequent action C. For SSAs it appears that there is some *objective*, empirical basis underlying the belief, (iii) above, that people will slip down the slope. That is, while the computational level theory of SSAs and warnings is the same, the underlying cognitive algorithms by which people assess Pr(C|A) may be different. For warnings, Evans et al. (2008) consider the generic Ramsey test for conditionals (Edgington, 1995; Ramsey, 1931), i.e., add the antecedent, A, to your stock beliefs, make adjustments to accommodate this belief, read off your degree of belief in the consequent, C, and this is Pr(C|A). However, the Ramsey test is a philosophical prescription crying out for a psychological, algorithmic level explanation (Oaksford & Chater, 2007, 2010a, 2010b). Constraint satisfaction processes in neural networks provide one possible implementation (Oaksford & Chater, 2010b). We have argued that for many SSAs, Pr(C|A) is determined via category boundary re-appraisal processes which may represent a further algorithmic instantiation of the Ramsey test (Hahn & Oaksford, 2007). The importance of category boundary re-appraisal is also consistent with discussions of SSAs in the philosophical literature.

It is often claimed by those authors who have been positive about SSAs that conceptual vagueness (e.g., the difficulty of providing a precise definition of "terminally ill") and a fear of constructing arbitrary distinctions (e.g. deciding which terminally ill patients' lives are "worthwhile") provides the rationale for many SSAs (e.g., Lode, 1999, p. 1499). For example, Govier (1982) suggests that the process of psychological assimilation acts as a catalyst for slippery slope arguments and that the *Sorites* paradox provides an example:

> It is morally wrong to kill a sentient being, and a foetus at the time of birth (T) is a sentient being. A foetus at one second ($T-1$) before the time of birth is also a sentient being, as the addition or subtraction of one second cannot affect a being's sentience. Therefore, a foetus at ($T-2$) is also a sentient being. Therefore, a foetus at ($T-n$) is also a sentient being; a foetus at the moment of conception is a sentient being.

The *Sorites* argument plays on the vagueness of the predicate "sentient" and the inevitability of the logical inference of *modus ponens* to achieve paradox. The idea that certain SSAs might be based on some kind of category boundary re-appraisal mechanism has been articulated implicitly by many authors (Holtug, 1993; Lode, 1999; Rizzo & Whitman, 2003; see also Walton (1992), who distinguishes *sorites* as a unique type of SSA). Indeed, the very notion that a slippery slope might exist between an ostensibly positive proposal and a negative outcome directly implies an extension process of some kind. When advances in gene therapy are discussed, the spectre of Nazi eugenics is raised precisely because the concept of pro-social genetic engineering is vague (Holtug, 1993), and membership of the category "acceptable practice" is a dynamic and fluctuating process.

On the basis of decades of empirical investigation of our everyday concepts it is clear that the majority of the concepts that pervade our everyday argumentation are indeterminate (e.g., Labov, 1973; McCloskey & Glucksberg, 1978; Rosch, 1973; Rosch & Mervis, 1975; see also Estes, 2003; Kalish, 1995, 2002). Our everyday concepts lack necessary and sufficient features and do not, as a consequence, have clear-cut boundaries. As a result classification is heavily dependent on the set of instances to which the category label has been applied.

As we have already observed, it is fundamental to a wide range of current theories of conceptual structure that encountering instances of the category at the category boundary will extend that boundary for subsequent classifications. There is, then, a feedback loop inherent in the classification of new data into an existing category, whereby that classification also affects and alters the category itself (see, e.g., Lakoff, 1987). In appropriate circumstances this extends the category boundary in a way that could naturally give rise to slippery slope arguments (Hahn & Oaksford, 2007). This suggests that SSAs will have an empirical basis

in many cases; extending the cases that fall under a conceptually vague term will genuinely facilitate future extensions. In other words, some slopes really are slippery. However, this leaves open the question of whether or not people are also naturally aware of this in argument evaluation. Experiments 2–4 set out to test directly whether parallel effects could be obtained in a categorisation and an argument evaluation task that manipulated exemplar similarity. However, first we seek to establish the decision-theoretic approach to SSAs by showing effects of probability and utility on judgements of their strength.

Experiment 1

As a preliminary to examining the mechanism of the slippery slope, we first establish the relevance of the decision theoretic factors to SSA evaluation and compare our results to those obtained by Evans et al. (2008) for conditional warnings. This experiment replicated Corner et al.'s (2006) Experiment 1 with a larger sample size. Participants were required to read several short scenarios containing slippery slope arguments, and provide a rating of argument strength. These slippery slope arguments were designed to vary in persuasiveness by manipulating (i) the probability, and (ii) the utility of the predicted future outcome.

Regarding (i), an argument where the probability of the outcome (B) following the initial action is high should be more convincing than an argument where this probability is low. In the present experiment, the probabilities presented to participants were varied by describing either a probable or an improbable mechanism by which the proposed outcome of the argument could occur.[3] Regarding (ii), a predicted outcome is a necessary component of slippery slope argumentation, but predicted outcomes that are perceived to have only a moderately negative expected utility will not be "feared" or avoided as much as outcomes with very negative expected utility. Predicted outcomes with very negative utilities will provide a stronger argument against the proposed course of action. In the present experiment, the outcome utilities of the arguments presented to participants were either moderately negative or very negative (reliable differences in outcome utility were established during pilot testing of the materials).

Method

Participants. 97 sixth form students from four colleges across South East Wales took part in the experiment, as part of a National Science and Engineering Week project.

Design, materials and procedure. The experiment was a 2 (probable/improbable mechanism) × 2 (moderately/very negative outcome utility) Latin Square Confounded design (Kirk, 1995). Both variables were manipulated across four different argument topics, creating a total of sixteen distinct arguments. Each participant received an experimental booklet containing four slippery slope arguments on different topics. The topics were (i) the legalisation of voluntary euthanasia, (ii) developments in cloning technology, (iii) the introduction of ID cards, and (iv) the cessation of postal deliveries to houses inhabited by vicious dogs (see Appendix A for all the materials used in Experiment 1). Topics (i), (iii) and (iv) were selected from previous experiments (see, e.g., Corner et al., 2006).

Related social psychological work on persuasion examining the effects of probability and utility manipulations in persuasive messages has repeatedly found effects of outcome utility only (e.g., Areni & Lutz, 1988; Johnson, Smith-McLallen, Killeya & Levin, 2004; but see also Albarracin & Wyer, 2001). This has led some authors to doubt whether probabilistic information is detectable by ordinary participants (van Enschot-Van Dijk, Hustinx, & Hoeken, 2003). It was important for this experiment therefore that any effects of the probability manipulation could not be attributed to unintended changes in outcome utilities. Topics (i), (iii) and (iv) had been pre-tested for utilities to ensure that this was not the case. Comparing utility ratings between the high and low probability versions of these topics revealed no significant effects such that the high probability version was also rated as having a higher utility: for euthanasia, $t(57) = 0.19$, $p = .85$ (high probability: mean utility = 8.14, SE = .60; low probability: mean utility = 8.00, SE = .44), for ID cards, $t(58) = 1.42$, $p = .16$ (high probability: mean utility = 4.34, SE = .88; low probability: mean utility = 2.36, SE = 1.11), for vicious dogs, $t(59) = 0.32$. $p = .75$ (high probability: mean utility = 6.19, SE = .54; low probability: mean utility = 6.41, SE = .45). These analyses confirm that for these materials probability is not confounded with utility.

Because of the Latin Square Confounded design (where each cell of the design contains different materials), we also added a new set on cloning technology. This topic (ii) was included as a 'scientific' topic because the study formed part of a National Science and Engineering Week project. The topics of the arguments and the order they were presented in were randomised for each participant using a Latin Square Confounded design and participants were required to provide a rating of argument strength for each argument on a scale from 0 (unconvincing) to 10 (very convincing).

The materials for all four topics are presented in Appendix A. Fig. 1 shows the four conditions of the experiment for topic (i) to illustrate the general design. In the first version of the scenario the probability of the outcome (B) given the initial proposal (A) is high (because of the alleged difficulty of formulating clear medical guidelines), whilst the utility of the predicted outcome is very negative (in the form of an increase in involuntary euthanasia). In the second version the probability of this negative outcome occurring is designed to be lower. In the third version the predicted outcome is less negative (other patients on the ward feeling less comfortable knowing that euthanasia is taking place), but probable. Finally, in the fourth version the outcome is both less negative and less probable.

[3] An alternative method of manipulating the probability of the outcome would have been to provide participants with a numerical estimate of the conditional probability of the outcome given the initial proposal, $P(B|A)$. However, people do not typically receive arguments that contain explicit probabilistic information in their daily lives, and we wanted to ensure that the materials were as naturalistic as possible. Thus, we opted to design arguments where the predicted outcome seemed either more or less likely to occur as a consequence of the initial action.

Results and discussion

Every participant saw four arguments—one from each experimental condition, and one of each topic according to the Latin Square Confounded experimental design described above. To statistically analyse data from Latin Square Confounded designs, participant effects within the ratings are factored out and the analyses are conducted on the residuals (Kirk, 1995).[4] The results for the Overall data and for each Topic separately are shown in Fig. 2. An ANOVA was conducted with Outcome Probability and Outcome Utility as the independent variables, and residual ratings of argument strength as the dependent variable. There was a main effect of both independent variables. Arguments with likely outcomes were rated as stronger ($M = 6.06$; $SD = 2.05$) than arguments with unlikely outcomes ($M = 4.97$; $SD = 2.1$), $F(1, 384) = 35.99$, $MSe = 3.01$, $p < .001$, $\eta^2 = .09$. Very negative outcomes produced higher ratings of argument strength ($M = 5.86$; $SD = 2.1$) than moderately negative outcomes ($M = 5.17$; $SD = 2.2$), $F(1, 384) = 15.54$, $MSe = 3.01$, $p < .001$, $\eta^2 = .04$. The interaction between probability and utility was also significant, $F(1, 384) = 4.85$, $MSe = 3.01$, $p < .025$, $\eta^2 = .01$. The strongest arguments were probable and had very negative outcomes, while the weakest arguments were less probable and had moderately negative outcomes.

A skeptic might maintain, however, that perhaps participants were not in fact estimating the probability of the outcome as defined by the arguments at all, but rather, were estimating the probability of the outcome independently of the argument in which it was embedded. Someone might, for example, consider it highly likely that compulsory ID cards will be introduced, regardless of whether the House of Lords support it or otherwise (perhaps there is strong support among the electorate, and the issue is to be put to a referendum). Pairwise comparisons revealed, however, a strong and consistent effect of the probability manipulation, under conditions of both very negative and moderately negative utility.

When the outcome was very negative, probable arguments were rated as significantly more convincing ($M = 6.2$, $SD = 2.05$) than less probable arguments ($M = 5.52$, $SD = 2.03$), $t(191) = 2.30$, $p < .05$. Probable arguments were also significantly more convincing ($M = 5.92$, $SD = 2.06$) than less probable arguments ($M = 4.41$, $SD = 2.14$) when outcome utility was moderately negative, $t(193) = 4.99$, $p < .001$. This suggests that participants were indeed estimating the probability of the outcomes according to the actual arguments in which they were embedded—as the only difference between the two arguments was whether the initial proposal made this outcome probable or not. For all materials there was a significant main effect of probability in the predicted direction (Dogs: $F(1, 93) = 5.72$, $MSe = 5.05$, $p < .01$, $\eta^2 = .09$; Euthanasia: $F(1, 93) = 4.78$, $MSe = 3.79$, $p < .025$, $\eta^2 = .05$; Cloning: $F(1, 93) = 9.41$, $MSe = 4.12$, $p < .005$, $\eta^2 = .09$; ID: $F(1, 93) = 6.45$, $MSe = 3.72$, $p < .01$, $\eta^2 = .07$). It is the mechanism underlying this robust effect of probability that we investigate in our subsequent experiments.

At a minimum, Experiment 1 would seem to provide a concrete rebuttal of the position that using or taking heed of SSAs is *never* a good idea. Such a sceptical position has been articulated recently by Enoch (2001) who claimed that even though it is possible to construct SSAs that are strong and compelling, people are inherently poor at abiding by the distinction between good and bad SSAs. Thus there is a 'meta' slippery slope between 'good' SSAs and 'bad' SSAs, and the use of 'good' SSAs will trigger a process that will ultimately lead to the spread of 'bad' SSAs. According to Enoch (2001), this means that even good SSAs are ultimately bad. Experiments 1 suggests, however, that people have little difficulty in consistently distinguishing between strong and weak SSAs.

Despite participants' ability to use probabilities and utilities to make this distinction, do they do so in a normatively appropriate way, i.e., do their decisions respect the dictates of Bayesian decision theory? Eqs. (1) and (2), show that probabilities and utilities combine multiplicatively in *SEU* theory. Multiplicative effects are often regarded as present when a significant interaction is observed (but see, Landsheer, van den Wittenboer, & Maassen, 2006). So the observed interaction seems to indicate that participants are combining utilities and probabilities as normative *SEU* theory predicts. However, as can be seen from Fig. 3, the ID materials were anomalous. While the effect of high probability is as predicted the effect of utility appears inverted. If the ID materials are removed, and the residuals re-computed the interaction is no longer significant, $F(1, 287) < 1$. However, although a simple additive effect of both fixed factors can be rejected if an interaction is observed, a multiplicative effect cannot be rejected if no interaction is observed (Landsheer et al., 2006). In part, this is a sensitivity issue: the sample size needed to detect interactions is seven to nine times higher than for main effects (Wahlsten, 1991).

Our results were consistent with Evans et al. (2008), who also found main effects of the same factors for utility conditionals but did not observe any consistent interaction effects, although many were significant (but not in by-items analyses). They put these interactions down to "spurious materials effects" (Evans et al., 2008, p. 108). As a precaution, they also generated a set of ordinal predictions for the cell means. We took a similar precaution. Assuming a multiplicative or an additive effect of the levels of two fixed factors, X and Y, such that $X_1 < X_2$ and $Y_1 < Y_2$, leads to an ordering over the products $a = X_1Y_1$, $b = X_1Y_2$, $c = X_2Y_1$, $d = X_2Y_2$ (or the sums) such that: $a < b$, $a < c$, $a < d$, $b < d$, and $c < d$. It is not a complete order because the relationship between b and c is not determined. So there are five ordinal predictions for each topic. In Experiment 1, 18 out of 20 of these cases were as predicted

[4] Computing residual values is necessary because although participants provide data in every condition of the experiment, the combination of topic and experimental condition differs between participants. Computing a residual transformation permits standard, between-subjects analyses to be conducted. Though this changes the absolute numerical values, it typically leaves the overall shape of the data unaltered. In all the data reported here, analyses of variance on raw and residual values produced the same statistical effects.

(Likely/Very Negative Outcome)
Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to an increase in cases of involuntary euthanasia – or "medical murder". The British Medical Association has warned that once voluntary euthanasia is permitted in some cases, it will be difficult to formulate clear guidelines about when doctors can euthanize patients. The article ends with the view of the author about the future of euthanasia legislation;

"We should oppose the legalisation of euthanasia in the UK, as it will lead to an increase in the number of instances of 'medical murder'".

(Unlikely/Very Negative Outcome)
Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to an increase in cases of involuntary euthanasia – or "medical murder". The British Medical Association has indicated, however, that there will be extremely clear and strict guidelines about if and when doctors may euthanize patients, and those who break them will be removed from the medical register. The article ends with the view of the author about the future of euthanasia legislation;

"We should oppose the legalisation of euthanasia in the UK, as it will lead to an increase in the number of instances of 'medical murder'".

(Likely/Less Negative Outcome)
Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to other hospital patients feeling that their lives are not as worthwhile. The British Medical Association has warned that once voluntary euthanasia is permitted, terminally ill patients may start to view their lives as of less worth than healthy individuals. The article ends with the view of the author about the future of euthanasia legislation;

"We should oppose the legalisation of euthanasia in the UK, as it will lead to other terminally ill patients feeling psychologically damaged by the process"

(Unlikely/Less Negative Outcome)
Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to other hospital patients feeling that their lives are not as worthwhile. The British Medical Association has indicated, however, that most hospital patients are unconcerned by the thought of voluntary euthanasia in hospitals. The article ends with the view of the author about the future of euthanasia legislation;

"We should oppose the legalisation of euthanasia in the UK, as it will lead to other terminally ill patients feeling psychologically damaged by the process"

Fig. 1. Example materials from Experiment 1 (Euthanasia).

($p < .001$, binomial test).[5] This is consistent with the qualitative, or ordinal assumptions of a number of approaches to informal argumentation which are all inspired by a decision-theoretic approach (see e.g., Bonnefon, 2009; Evans et al., 2008).

While we need not go beyond this conclusion to motivate our further investigation of the mechanisms underlying the robust effect of probability on SSAs, this experiment did provide some evidence of multiplicative effects. A multiplicative or additive relation imposes different constraints on the products or sums above. According to a multiplicative relation the following *ratios* should be equated: $b/a = d/c$. According to an additive relation the following *differences* should be equated: $b - a = d - c$. We therefore calculated a predicted value of each product or sum from the remaining three known mean values. So for each topic we calculated, for example, b, from the mean values of a, c, and d, i.e., ad/c (multiplicative) or $d - c + a$ (additive). In one sample t-tests, we then compared the observed values of b with the predicted value of b acting as

[5] The anomaly for ID would appear to attributable to participants not uniformly regarding receiving a £1000 fine as more onerous than having to attend a police station within 28 days to show them your ID card.

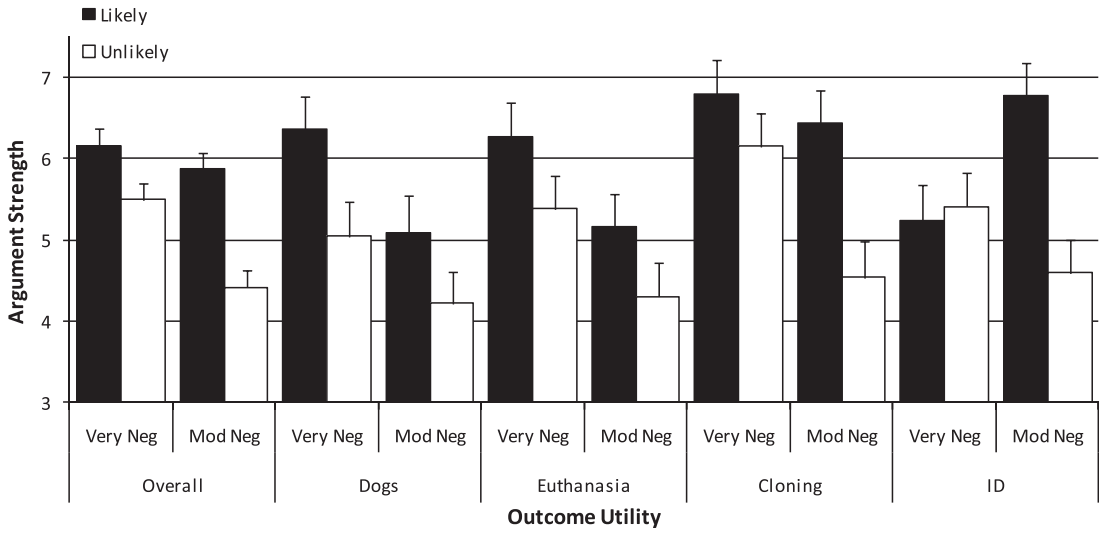

Fig. 2. Argument strengths by outcome utility (very vs. moderately [Mod] negative) and outcome probability (likely vs. unlikely) for the overall data and for each topic in Experiment 1.

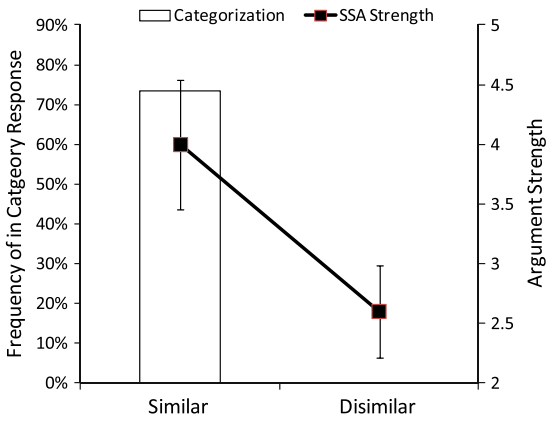

Fig. 3. SSA strength and proportion of categorisation judgements by similarity in Experiment 2. Note. Error Bars = Standard Errors.

the test value. If there is a significant difference from the test value then the additive or multiplicative relation can be rejected. Both the multiplicative and the additive relation could be rejected for all products and sums for the ID (multiplicative: $t(20–26) > 4.33$; additive: $t(20–26) > 5.36$)[6] and for the Cloning materials (multiplicative: $t(21–25) > 2.93$; additive: $t(20–26) > 2.84$). However, neither could be rejected for all products and sums for the Euthanasia (multiplicative: $t(22–25) < .61$; additive: $t(20–26) < .06$) or for the Dogs materials (multiplicative: $t(20–26) < .62$; additive: $t(20–26) < 1.02$). We compared these t-values for the multiplicative and the additive models in further paired t-tests for each topic.[7] For ID and for Cloning there were no significant differences, $t(3) = .07$ and $t(3) = 1.77$ ($p = .18$, two-tailed) respectively. For Euthanasia, the additive model gave a significantly better fit, i.e., lower t-values, $t(3) = 5.45$ ($p = .012$, two-tailed) but for the Dogs material the multiplicative model gave a significantly better fit, $t(3) = 11.80$ ($p = .001$, two-tailed). Of course, these last t-values shouldn't be taken too seriously given the very low degrees of freedom but they do provide a rough and ready comparative index of fit. In sum, these data do reveal some evidence for normative, multiplicative combination of probabilities and utilities but clearly this issue requires further research.

However, our focus in this paper is on the robust effect of probability, Pr(C|A). As we pointed out in the *Method* section, related social psychological work on persuasion has also examined the effects of probability and utility manipulations in persuasive messages has repeatedly found effects of outcome utility only (e.g., Areni & Lutz, 1988; Johnson et al., 2004; but see also Albarracin & Wyer, 2001), and this has led some authors to doubt whether probabilistic information is detectable by ordinary participants (van Enschot-Van Dijk et al., 2003). The clear effect of probability observed here demonstrates that this is not the case. Moreover, as we pointed out in the *Method* section, these robust effects of probability could not be attributed to the high probability cases also having higher utility, at least for the euthanasia, ID cards, and vicious dogs materials where our pre-tests explicitly ruled this out. More importantly, however, it is the probabilistic component that is subject to the defining mechanism underlying slippery slope arguments, and we focus on this mechanism—category boundary re-appraisal—in the remainder of the paper.

Experiment 2

The next three experiments investigate two related questions. First, does slippage occur, i.e., when exemplars *a* and *b* are similar are people more willing to classify *b*

[6] The DFs varied between the four comparisons; here we simply show the range.

[7] Although they are unlikely to be normally distributed, especially given the low sample size.

as *F* once *a* has been classified as *F*? Second, when slippage does occur, do people distinguish between SSAs in the appropriate way, i.e., are SSAs stronger when *a* and *b* are more similar?

If SSAs have an objective basis in category expansion driven by exemplar effects, there should be agreement between the perception of an SSA's strength and corresponding categorisation decisions, given identical data to evaluate. In this experiment we used a uni-dimensional, quantitative category and numerically defined exemplars. In classifying areas as of "outstanding natural beauty" people were first provided with exemplars, some of which were category members, and some of which were non-members, depending (solely) on the exact number of large animals found at that location. So for example, location *A* (114) and *B* (149) were "out", i.e., they were not classified as areas of outstanding natural beauty, and *C* (149) and *D* (224) were "in", i.e., they were classified as areas of outstanding natural beauty. In a classification task, participants were then told that two further locations, *I* (a mnemonic for "initial') and *X* (as in "experimental"), were up for consideration. Having been told that location *I* was designated as falling within the category they were asked to judge whether the target category *X* (179) likewise should be categorised as an area of outstanding natural beauty. In the dissimilar condition they are told that location *I* has 218 species of large animal whereas in the similar condition they were told it has 194. The category boundary re-appraisal mechanism should mean that location *X* is also categorised as an area of outstanding natural beauty when it is most similar to location *I*, i.e., 194, than when it is dissimilar, i.e., 218. Moreover, when given the following SSA:

> *Location I* (194/218) should not be categorised as an area outstanding natural beauty because it will lead to *location X* (179) also being categorised as an area outstanding natural beauty.

They should endorse it more when locations *I* and *X* are similar than when they are dissimilar.

These kinds of similarity based exemplar effects are familiar from the inductive reasoning literature (e.g., Osherson, Smith, Wilkie, Lòpez, & Shafir, 1990; Rips, 1989; Sloman, 1993), for example,

> Rabbits have sesamoid bones
> Therefore Dogs (Bears) have sesamoid bones

People are more willing to conclude that dogs than bears have this property because dogs are more similar to rabbits than bears. Our studies of SSAs go beyond similarity effects like this because inductive generalisations like these involving natural categories cannot be used to formulate SSAs, i.e., it makes no sense to argue that:

> One should not categorise rabbits as having sesamoid bones because it will lead to (dogs) bears being classified as having sesamoid bones.

Whether an animal has a certain bone or not is a matter of objective discovery, not of human decision making. One can imagine related SSAs in the context of taxonomic judgements, e.g.,

> One should not categorise rabbits as X because they have sesamoid bones because this could lead to (dogs) bears being categorised as X because they have sesamoid bones.

But crucially, for the SSA to make sense (see *What is an SSA*), a *cost* for categorising an item as *X* would need to be specified. In the case of the materials we employed, categorising a location as of outstanding natural beauty imposes a cost, i.e., the land is set aside and cannot be used for industrial, agricultural or housing development. This cost is constant across the examples and so differences must be driven by Pr(C|A) alone. In the taxonomic case the nature of any implied cost is opaque.

In sum, while the core of our hypothesis in this experiment is that SSAs are underpinned by one of the mechanisms that underpin inductive generalisation, i.e., the willingness to extend a category to similar exemplars, it remains to be demonstrated that this mechanism affects both types of argument. This is what this and our subsequent experiments set out to test.[8]

Method

Participants. Sixty undergraduate psychology students from Cardiff University participated in the experiment in exchange for course credit.

Design, materials and procedure. All participants were presented with the following cover story:

> In Finland, some rural locations are designated by the Government as areas of Outstanding Natural Beauty. If an area is identified as being of Outstanding Natural Beauty, no development is permitted on it. At the same time as preserving unique natural habitats, however, the Finnish Government must provide housing for its growing population. Land is designated as being of Outstanding Natural Beauty if it contains an unusually high number of large animal species.

Having read the cover story, all participants were presented with the following information about the decisions already made by the Finnish Government:

Location A:
South Pernothea is home to 114 species of large animals.
Decision: Not eligible for Area of Outstanding Natural Beauty status.

Location B:
Reklan is home to 149 large animal species.
Decision: Not Eligible for Area of Outstanding Natural Beauty status.

Location C:
Grenadia is home to 259 types of large animals.
Decision: Eligible for Area of Outstanding Natural Beauty status.

[8] This experiment was reported in the *Proceedings of the Cognitive Science Society* (Corner et al., 2006).

Location D:
Scarathon is home to 224 species of large animals.
Decision: Eligible for Area of Outstanding Natural Beauty status.

All participants were then presented with two further locations (*Location I* and *Location X*), which had not yet been decided by the Finnish government. The similarity of *Location I* to *Location X* was manipulated by altering the value of the number of large species in *Location I*. In one group, *Location I* was similar to *Location X*:

Two further cases are currently being considered by the Finnish government and the Finnish Housing Association, the details of which are as follows:

Location I:
Aunskop is home to 194 species of large animals.

Location X:
Sellenfeld is home to 179 species of large animals.

In the second group, *Location I* was *dissimilar* to *Location X*:

Location I:
Aunskop is home to 218 species of large animals.

Location X:
Sellenfeld is home to 179 species of large animals.

The number of species of large animals was a between subjects manipulation; whether participants were asked for a categorisation decision or a rating of argument strength was also treated as a between subjects manipulation, creating a total of four experimental groups.

In the categorisation groups, participants were asked to make a categorisation decision of their own, based on the information they had just read, i.e. whether *Location X* was eligible for *Outstanding Natural Beauty* status. When *X* was more similar to *I* it should be more likely to be categorised as eligible because *I* should lead to the category boundary being re-appraised to be closer to *X*. In the argument conditions, in the similar condition participants should perceive *X* as sufficiently close to the category boundary defined by *I*, and therefore vulnerable to a slippery slope style re-appraisal (mirroring the exemplar effect predicted in the categorisation groups).

Results and discussion

Fig. 3 shows the results of Experiment 2. The yes/no data obtained from the categorisation groups were analysed using a ranked sign test (Siegel & Castellan, 1988). Participants in the similar condition, who had been told that *Location I* contained *194* animals, categorised the new location as deserving of *Outstanding Natural Beauty* status on 11 of 15 occasions. Participants in the dissimilar condition, who had been told that *Location I* contained *218* animals, categorised the new location as deserving of *Outstanding Natural Beauty* status on 0 of 15 occasions. This difference was significant at $p < .01$.

The argument rating data were analysed using a *t*-test. Participants who had been told that *Location I* contained *194* animals rated the arguments as significantly more convincing ($M = 4$, $SD = 2.1$) than participants who had been told that *Location I* contained *218* animals ($M = 2.6$, $SD = 1.5$), $t(28) = 2.08$, $p < .05$. These results provide empirical support for the philosophical analysis of slippery slope arguments by authors such as Govier (1982) and Volokh (2003) by demonstrating, in a tightly coupled design, how slippery slopes may rest on a category boundary extension process. That is, this is the mechanism by which people assess the probability that people will re-appraise the undesirable consequent, C, as falling under the target category, given that A has been categorised as falling under this category.

However, Experiment 2 did not demonstrate actual re-appraisal, i.e., it did not show participants actually changing their mind about location *X*. A paradigm to demonstrate this would require showing that when the same participants were told about *location I* (218) being classified as an area of outstanding natural beauty, they indicate that *X* should not be classified in this way but when they are subsequently told about *location V* (194) they change their mind and now indicate that *X* should be classified as an area of outstanding natural beauty. Demonstrating actual re-appraisal within individuals is important because, for example, anti-drugs campaigners suggest that the legalisation of cocaine might not seem such an abhorrent proposal if the legalisation of cannabis were to go ahead. This argument relies on at least some individuals being persuaded to change their mind about cocaine because of the decision about cannabis. We could find no demonstration of such within subjects effects in the category based induction literature or the work on exemplar models of categorisation, so in Experiment 2A, we constructed an experiment to ensure that re-appraisal does occur with similar materials to those used in Experiment 2.

Experiment 2A

Experiment 2A was designed to replicate the categorisation data obtained in Experiment 2 using a within-participants design in which participants are asked for two judgements about *Location X*, one after being told about *Location I* (218) then after being told about *Location V* (194, see above).

Method

Participants. Forty undergraduate psychology students from Cardiff University participated in the experiment in exchange for course credit. Experimental conditions were manipulated within-participants, and so all participants completed the same experimental tasks, in the same order.

Design. Experiment 2A was designed to provide a within-subjects replication of the between-participants categorisation effect observed in Experiment 2. Participants were presented with a list of locations that had already been assessed by the Finnish government (identical to the list used in Experiment 2). The independent variable was manipulated by presenting two exemplar locations (*Location I* and *Location V*) sequentially, differing in their

numerical value. The dependent measure was assessed across two categorisation decisions, following the presentation of each exemplar location. Effectively, participants in Experiment 3 completed both the experimental tasks that had been split between the groups of categorisation participants in Experiment 2. The first categorisation decision was made on the basis that *Location I* (the first exemplar), containing 218 species of large animal, was eligible for Outstanding Natural Beauty status. It was predicted that most participants would classify *Location X*, with179 animal species, as sufficiently dissimilar to *Location I* to be *undeserving* of Outstanding Natural Beauty status.

The second categorisation decision was made following information about *Location V* containing 194 species of large animal, which was also eligible for Outstanding Natural Beauty status. It was predicted a significant number of participants would perceive a similarity between *Location X* and *Location V*, re-appraise their categorisation decision and now classify *Location X* as *deserving* of Outstanding Natural Beauty status.

Materials and procedure. Each participant received an experimental booklet containing a brief description of the fictitious scenario, a list of locations that had already been adjudicated, and the dependent measure tasks (i.e. two categorisation decisions). The booklet was constructed so that participants would not see the additional information, or the second categorisation task, until they had made their first decision.

Results and discussion

The distribution of categorisation responses is presented in Fig. 4. A McNemar Change test (Siegel & Castellan, 1988) revealed a significant shift in categorisation behaviour following the presentation of the second exemplar location, $\chi^2(1, N = 40) = 11.02$, $p < .01$. 37.5% of participants who initially excluded *location X* from the being classified as an area of outstanding natural beauty (Fig. 4, Decision 1) moved to including *location X* under this classification after encountering *location V* (Fig. 4, Decision 2). The minority of participants who responded "yes" on the first categorisation task all maintained this decision on the second categorisation task. This demonstrates that people were not simply changing their mind in response to being asked to make the same decision twice. Excluding these participants from the statistical analysis (as it is not possible to extend a category to incorporate data it already includes), did not alter the finding of a significant change across the two decisions, $\chi^2(1, N = 32) = 11.02$, $p < .01$.

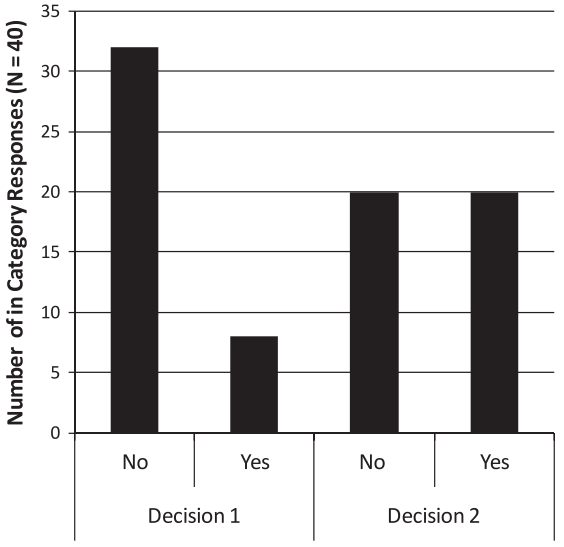

Fig. 4. Number of in category judgements when dissimilar (Decision 1) and then when similar (Decision 2) in Experiment 2A.

The distribution of YES/NO responses on the second classification task was equal, i.e., half the participants perceived the dependent location, *X*, as *not* sufficiently similar to the exemplar location, *V*, to warrant classifying them together. The focus of the experiment was, of course, on the relative change in response distribution, rather than the absolute magnitudes of the distributions themselves. However, finding that re-evaluation is probabilistic, rather than deterministic (i.e., it will occur for some individuals but not others), fits well with both a probabilistic account of SSA strength and an intuitive understanding of how slippery slopes are realised in practical contexts. Typically, SSAs warn of threats, dangers and risks, rather than inevitable outcomes. And typically, *some* change in attitudes towards a particular outcome is sufficient to warrant the use of an SSA. In order for cocaine to be legalised in the future, for example, it is not necessary that *all* individuals re-evaluate its status following the legalisation of cannabis. Rather, a *sufficient* number of people must change their mind in order for a qualitative shift in legislation to be brought about. That this 'sufficient' amount is typically undefined only serves to reinforce the idea that SSAs trade on the uncertainty of the future.

Taken together, the results of Experiments 2 and 2A suggest that the re-appraisal of category boundaries may provide an empirical mechanism underpinning the occurrence of an important class of SSAs. Most natural categories lack clearly defined boundaries and are susceptible to exemplar-based modifications as documented here, and extensively elsewhere (e.g., Lamberts, 1995; Nosofsky, 1986, 1988a, 1988b). The probability of a predicted outcome following from an initial proposal appears to be directly related to the empirical similarity between the initial proposal and the predicted outcome. Furthermore, category boundary re-appraisal can occur not only across groups of people, but within the same individual or group over time (for related work on the flexible re-appraisal of information in social judgements, see Wegener & Petty, 1995).

Experiment 3

Experiments 2 and 2A raised a couple of questions. First, these experiments demonstrated that the same materials can lead separate groups of participants to make categorisation judgements and evaluate SSAs consistent with a category boundary re-appraisal account. However, it remains to be demonstrated that participants asked to

simultaneously make both categorisation judgements and argument evaluations will behave in the same way. Second, these experiments used uni-dimensional, numerical stimuli. This quantitative approach allowed the objective similarity between the exemplar and existing category members to be controlled with some accuracy. However, real world categories are neither defined using explicit numerical criteria, nor based on a single, quantifiable dimension. SSAs typically relate to notions such as personal freedoms (the legalisation of voluntary euthanasia) or civil rights (the introduction of ID cards). Accordingly, we sought to replicate the results of the exemplar-based slippery slope experiments using materials closer to real world SSAs.

Our final experiment pursued both these questions. First, it attempted to replicate Experiment 2 and to demonstrate a correlation between ratings of category inclusion and argument strength. Second, to extend to qualitative categories, participants were presented with information about the legal system of a fictional location 'Sotherby Island.' They were informed that all crimes committed on Sotherby Island would now be punished by either less than 20 years in prison (SUB 20 offences), or by more than 20 years in prison (20 PLUS offences). After being presented with two exemplars of serious offences categorised as 20 PLUS and two exemplars of less serious offence categorised as SUB 20, one group of participants in the similar condition were presented with the following SSA:

> One should not categorise assault in possession of a knife (Offence *A*) as SUB20 because this will lead to assault in possession of a gun (Offence *X*) being classified as SUB 20.

In the dissimilar condition, the SSA was:

> One should not categorise assault without a weapon (Offence *B*) as SUB20 because this will lead to assault in possession of a gun (Offence *X*) being classified as SUB 20.

Participants were asked to rate how, convincing (*convincingness*), persuasive (*persuasiveness*), and strong (*strength*) they found the argument on a 0–10 rating scale. These measures were intended to measure the same underlying construct, i.e. argument evaluation, although any differences that emerged would permit a more stringent analysis of how participants were interpreting the argument evaluation task.

For the categorisation judgement participants were then told that the islanders did actually classify Offence *A* (similar) or Offence *B* (dissimilar) as SUB 20 and were then asked to rate on a continuous scale from 0 (Definitely Not) to 10 (Definitely) whether they thought Offence *X* would be made a SUB 20 crime (*decision*). Participants were also asked to rate how confident they were in this judgement (*confidence*). The reason for using these continuous scales was to directly establish the relationship between category boundary re-appraisal in correlational analyses.

The nature of the relationship between these continuous variables requires some elaboration. We look at the similar case first. For the categorisation judgement, assault with a knife (Offence *A*) is initially classified as SUB20 and participants must state whether, as a consequence, they believe that assault with a gun (Offence *X*) will be also classified as SUB20. A *decision* rating above 5 indicates YES and a rating below 5 indicates NO; "5" indicates indecision or uncertainty. High confidence in *either* direction (i.e., values of *decision* towards 10 for YES and towards 0 for NO) will correspond to high *convincingness*, *persuasiveness*, or *strength* in the SSA argument. This is because if a participant indicates YES (*decision* > 5), then this means that she thinks that assault with a gun (*X*) would be classified as SUB 20. This outcome is so undesirable that in the SSA task she should be highly convinced that assault with a knife (*A*) should not be classified as SUB 20. If a participant indicates NO, then this means that she thinks that assault with a gun (*X*) would not be classified as SUB 20 because it is so undesirable. Consequently, in the SSA task they should still be highly convinced that assault with a knife (*A*) should not be classified as SUB 20. This is just category boundary re-appraisal working in both directions. Someone who indicates NO (*decision* < 5) believes that assault with a gun (*X*) would not be classified as SUB20 and so given how similar they are, they should also believe that assault with a knife (*A*) *should* also not be classified as SUB20 (regardless of the islanders initial decision). These considerations imply that the relationship between *decision* and the SSA measures will be curvilinear, with a minimum at 5.

This relationship should be moderated by similarity, being far weaker in a dissimilar condition. For example, someone who indicates NO (*decision* < 5) believes that assault with a gun (*X*) would not be classified as SUB20 but given how dissimilar they are, may still believe that assault without a weapon (*B*) should be classified as SUB20. We included a *confidence* measure to establish this moderating effect in a moderated linear regression analysis. *Confidence* should be high when *decision* is low (high confident NO) and when it is high (high confident YES) and so should be linearly related to the SSA measures of *convincingness*, *persuasiveness*, or *strength*. Consequently, if this relationship can be established, then the moderating effect of similarity can be tested in a moderated linear regression analysis regressing confidence on the SSA measures (Aguinis, 2004).

Method

Participants. Fifty-two undergraduate students from Cardiff University participated in Experiment 4 in return for course credit.

Design. All participants were presented with the following information:

> The inhabitants of Sotherby Island have decided to clear up their complex legal system once and for all, by creating a two-tier sentencing structure. They have voted for a strict division of sentences, such that some crimes are automatically punished by 20 years (or more) in prison (the 20 PLUS category), and all other crimes are automatically punished by less than 20 years in prison (the SUB 20 category).

The islanders have already voted on the categorisation of many crimes, some of which are shown below.

Offence: Burglary
Decision: SUB 20 category

Offence: Manslaughter
Decision: 20 PLUS category

Offence: Rape
Decision: 20 PLUS category

Offence: Arson
Decision: SUB 20 category

Offence: Murder
Decision: 20 PLUS category

All participants in the experiment completed both an argumentation *and* a categorisation measure, which were combined in the same set of materials. Participants were presented with two further crimes, which had not yet been voted on by the Sotherby Islanders. The qualitative exemplar manipulation was implemented by altering the similarity of these two crimes to each other, creating two experimental groups. Participants in the *similar* experimental condition ($n = 27$) were presented with these two crimes:

Offence A: Assault in Possession of a Knife
Decision: ?

Offence X: Assault in Possession of a Gun
Decision: ?

Participants in the *dissimilar* experimental condition ($n = 25$) were presented with these two crimes:

Offence B: Assault without a Weapon
Decision: ?

Offence X: Assault in Possession of a Gun
Decision: ?

As in previous experiments, an SSA was constructed stating that *if Offence A* (or *B*) is made a SUB 20 crime, *then Offence X* will also be made a SUB 20 crime, and that *Offence A* (or *B*) should therefore be made a 20 PLUS crime. Participants were asked to evaluate the strength of this argument, but in this experiment three measures of argument evaluation were obtained, *convincingness*, 0 (Unconvincing) to 10 (Very Convincing), *persuasiveness*, 0 (Unpersuasive) to 10 (Very Persuasive), and *strength*, 0 (Very Weak) to10 (Very Strong).

Participants were then informed that the Sotherby Islanders *did* decide to make *Offence A* (or *B*) a SUB 20 crime. Their next task was then to decide whether *Offence X* would also be made a SUB 20 offence. Participants were asked to indicate on a continuous scale from 0 (Definitely Not) to 10 (Definitely) whether they thought *Offence X* would be made a SUB 20 crime. They also rated how confident they were in this judgement, 0 (Not Confident at all) to 10 (Completely Confident).

In an attempt to avoid the potential effects of prior beliefs about which crimes should and should not receive lengthy jail sentences in the 'real world', participants were explicitly instructed to make their ratings of argument strength and categorisation decisions based *only* on the information they had seen in the experiment. Specifically, participants were requested not to consider whether (in 'real life') it would be morally right or wrong to place any particular crime in any particular category, but simply to make their assessments based on the information in the experiment.

Results and discussion

We first computed a reliability coefficient (Cronbach's Alpha) for the three measures of SSA evaluation, which was .89. Such a high reliability suggests that the three variables can be interpreted as measuring the same underlying construct (i.e., SSA evaluation). Responses on the three measures of SSA evaluation were therefore averaged to create a single measure of argument evaluation (*combinedmeasure*). The mean values of *decision* and *combinedmeasure* are shown in Fig. 5 by similarity (Similar, Dissimilar). *T*-tests were conducted to examine the effect of the exemplar manipulation. The combined measure was significantly higher in the similar condition ($M = 5.02$, $SD = 1.8$) than in the dissimilar condition ($M = 2.66$, $SD = 1.7$), $t(50) = 4.77$, $p < .001$. That is, people were significantly more convinced by the SSA that *Offence A* should not be should not be categorised as SUB 20 than they were that *Offence B* should not be categorised as SUB 20. Participants in the similar condition also gave significantly higher ratings ($M = 5.18$, $SD = 2.68$) of whether *Offence X* should be placed in the same category of crimes as *Offence A* than participants in the dissimilar condition gave for whether *Offence X* should be placed in the same category of crimes as *Offence B* ($M = 3.8$, $SD = 2.25$), $t(50) = 2.01$, $p < .05$. The results replicate Experiment 2 using a qualitative exemplar manipulation, showing that exemplar-based category re-appraisal can be demonstrated equally in the contexts where SSAs are typically found.

However, the main purpose of Experiment 3 was to demonstrate a direct link between category boundary reappraisal and SSA evaluation. We therefore conducted correlational analyses which are shown in Fig. 6. The upper row of Fig. 6 shows the results for the overall data, the middle row shows the similar condition and the bottom row

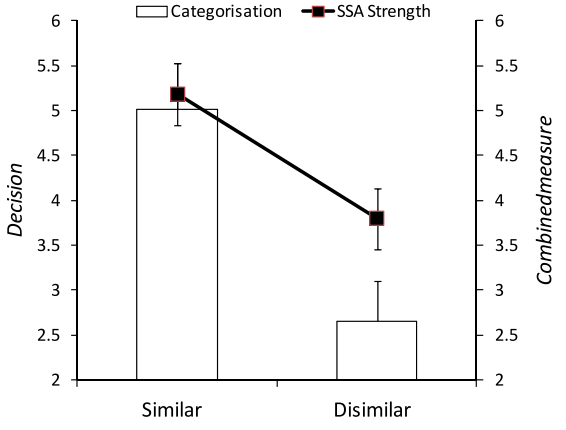

Fig. 5. SSA strength (*combinedmeasure*) and strength of categorisation judgements (*decision*) by similarity in Experiment 3. Note. Error Bars = Standard Errors.

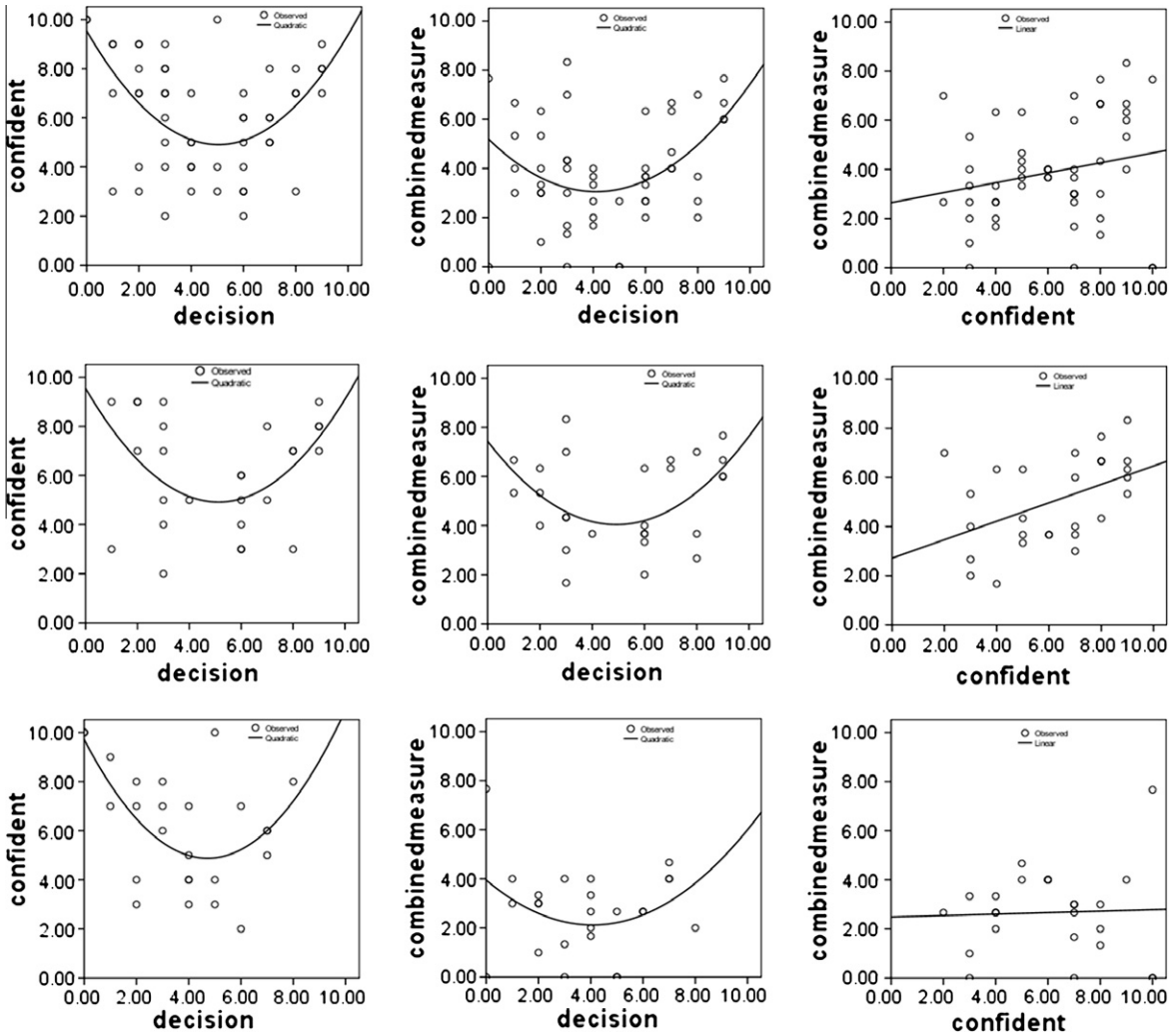

Fig. 6. Correlational analyses for Experiment 3. Row 1 shows the analyses for the Overall data, Row 2 for the Similar condition, and Row 3 for the Dissimilar condition. From left to right, the first graph is the regression of *confidence* (confident) on *decision*; the second is the regression of the *combinedmeasure* of SSA evaluation on *decision*, and the third graph is the regression for *combinedmeasure* of SSA evaluation on *confidence*.

shows the dissimilar condition. For the overall data, as predicted there was a significant quadratic, curvilinear relationship between *decision* (d) and *confidence* (c), $r(51) = .52$, $p < .0001$. The Beta coefficients in the regression equation, $c = 9.54 - 1.84d + .18d^2$, were both highly significant (β_1: $t = 4.25$, $p < .0001$; β_2: $t = 4.10$, $p < .0001$), and the minimum was almost exactly at $d = 5$ (5.055).[9] As predicted, there was a similar relationship between *decision* and the combined SSA evaluation measure (*combinedmeasure*, "*m*"), $r(51) = .41$, $p < .01$. The Beta coefficients in the regression equation, $m = 5.19 - 1.04d + .13d^2$, were both significant (β_1: $t = 2.42$, $p < .01$; β_2: $t = 2.86$, $p < .01$). Finally, the linear regression of *combinedmeasure* on *confidence* was approaching significance in the overall data, $r(51) = .22$, $p = .062$.

Our principle predictions arise for the disaggregated data for the similar condition (middle row, Fig. 6). As predicted there was a significant quadratic, curvilinear relationship between *decision* (d) and *confidence* (c), $r(26) = .44$, $p < .05$. The Beta coefficients in the regression equation, $c = 9.55 - 1.81d + .18d^2$, were both significant (β_1: $t = 2.34$, $p < .025$; β_2: $t = 2.42$, $p < .025$), and the minimum was close to $d = 5$ (5.131). As predicted, there was a similar relationship between *decision* and the combined SSA evaluation measure (*combinedmeasure*, "*m*"), $r(26) = .45$, $p < .05$. The Beta coefficients in the regression equation, $m = 7.43 - 1.38d + .14d^2$, were both significant (β_1: $t = 2.21$, $p < .025$; β_2: $t = 2.37$, $p < .025$). Finally, the linear regression of *combinedmeasure* on *confidence* was significant for the similar condition, $r(25) = .47$, $p < .01$.

These relationships between the categorisation variables and the SSA evaluation measures were weaker in the dissimilar condition (bottom row, Fig. 6), as predicted. There was a significant quadratic, curvilinear relationship

[9] For all the quadratic relationships reported here, neither the linear nor the cubic relationship was significant or offered any improvement on the quadratic relationship.

between *decision* (d) and *confidence* (c), $r(24) = .59$, $p < .005$. The Beta coefficients in the regression equation, $c = 9.72 - 2.05d + .22d^2$, were both significant (β_1: $t = 3.27$, $p < .005$; β_2: $t = 2.81$, $p < .01$), and the minimum was close to $d = 5$ (4.72). However, as predicted, the relationship between *decision* and the combined SSA evaluation measure (*combinedmeasure*, "*m*") was not significant, $r(24) = .33$, $p = .27$. The Beta coefficients in the regression equation, $m = 3.95 - .90d + .11d^2$, were also not significant (β_1: $t = 1.65$, $p = .11$; β_2: $t = 1.64$, $p = .11$). Finally, the linear regression of *combinedmeasure* on *confidence* was not significant for the dissimilar condition, $r(24) = .04$, $p = .85$.

The rightmost column of Fig. 6 reveals a clear moderating effect of similarity, i.e., the correlation between *combinedmeasure* and *confidence* reduced from .47 in the similar condition to .04 in the dissimilar condition. To test this result we performed a moderated multiple linear regression analysis with similarity as a categorical moderator (Aguinis, 2004). This involves coding similarity as a dummy variable ("1" = similar; "0" = dissimilar) and testing two models, the first with *confidence* and *similarity* as predictors to test simple effects, and the second with *confidence*, *similarity* and the product *confidence* × *similarity* as predictors. This allows a test of whether the interaction leads to a significant increase in the proportion of variance accounted for, i.e., ΔR^2. For the first model $r(49) = .60$, $p < .0001$, and for the second, model $r(48) = .63$, $p < .0001$. $\Delta R^2 = .034$, which was significant, $p = .05$ (one-tailed). Moreover, at 3.4% of additional variance explained, this is more than the 1–2% normally considered as practically as well as statistically significant (Aguinis, 2004, pp. 140–141). For the similar condition (*similarity* = 1) the regression equation was $m = 2.72 + .38c$ and for the dissimilar conditions (*similarity* = 0) the regression equation was $m = 2.49 + .03c$.

In sum, the effect of *decision* on SSA evaluation (*combinedmeasure*) is partially mediated[10] by *confidence* and is moderated by similarity. The more distant in conceptual space the more decoupled confidence in categorisation judgments and SSA evaluation becomes as one would expect if similarity in conceptual space and so category boundary re-appraisal was the prime driver of Pr(C|A) in these arguments. The primary effect is to radically reduce the effect of *confidence* on SSA evaluation (*combinedmeasure*) from $\beta = .38$ in the similar condition to $\beta = .03$ in the dissimilar condition.

General discussion

The goal of this research was to uncover the mechanism underlying the effect of probability on SSA evaluation. Experiment 1 established that there was indeed a robust effect of Pr(C|A) which was significant across the materials used and also within each set of materials. There was also some evidence for multiplicative combination of utilities and probabilities but only for one set of materials (Dogs). These results for SSAs largely replicated the related effects for conditional warnings, in particular for the ordinal predictions, found by Evans et al. (2008). What distinguishes SSAs from conditional warnings is the mechanism by which Pr(C|A) is assessed. Evans et al. (2008) only considered the generic Ramsey test without identifying an underlying psychological mechanism. But for SSAs there is an implied mechanism such that classifying one item *a* under a category *F*, i.e., *Fa*, increases the probability the further items, *b*, will be classified under the same category, *Fb*. Such category boundary re-appraisal depends on similarity or closeness in conceptual space between *a* and *b* and so relies on similar mechanisms to those that underlie category based induction. The difference hinges on the implied costs involved in SSAs which lead to the inference that one should not classify item *a* under a category *F* in the first place. This issue does not arise in category based induction which relies on factual discovery rather than human decision. Experiments 2 and 3, set out to demonstrate that SSA evaluation is directly related to category boundary re-appraisal. Experiment 2, showed that when *a* and *b* are similar, identical material led different groups of participants to evaluate SSAs as strong and to categorise new items, *b* as *F*, when *a* had been categorised as *F*. But this did not happen when *a* and *b* were dissimilar. Experiment 2A showed the same participants who had initially refused to categorise *b* as *F*, when *a* had been categorised as *F* and *a* and *b* are dissimilar, re-appraised this decision on being told about intermediate item *c* that was similar to *b* and which was also categorised as *F*. Experiment 3 directly related category boundary re-appraisal to SSA evaluation by showing in correlational analyses that confidence in categorising *b* as *F*, when *a* had been categorised as *F*, predicted the strength of SSA evaluation but only when *b* and *a* were similar.

These are the first empirical investigations of the mechanism underlying the slippery slope argument. By drawing on the normative framework of decision theory, as proposed by Hahn and Oaksford (2007), and well-established empirical phenomena from the categorisation literature, we have shown that people do not view all SSAs as inherently bad. Despite their philosophical notoriety, SSAs are used (and seem to be accepted) in a wide variety of practical contexts. The experimental evidence reported in this paper suggests that in some circumstances, their practical acceptability can be justified, not just because the decision-theoretic framework renders them subjectively rational, but also because it is demonstrated how, objectively, the slippery slopes they claim do in fact exist.

Experiment 1 is the first empirical demonstration that SSAs vary predictably in their acceptability. Although this variation was not perfectly captured by the decision-theoretic approach to consequential argument strength, this is of interest with regard to argumentation theory and the study of the fallacies in general, variation in strength for arguments of identical structure has typically been problematic for existing theories of fallacy (e.g., van Eemeren & Grootendorst, 2004). However, the idea that argument strength is a graded concept is a central tenet of the Bayesian approach to informal fallacies (Hahn & Oaksford, 2007).

[10] Establishing the mediation effect is complicated by the curvilinear relationships observed (Edwards & Lambert, 2007).

In this respect the results mirror those recently obtained for other supposed fallacies such as the 'argument from ignorance' (Corner & Hahn, 2009; Hahn & Oaksford, 2007; Oaksford & Hahn, 2004). With regard to theorising about SSAs specifically, our results provide a concrete rebuttal of the position that the use of SSAs is never acceptable in rational debate (Enoch, 2001; see also Corner & Hahn, 2007).

The principles of decision theory have been used on previous occasion to model the acceptability of deontic conditionals (Over et al., 2004; Perham & Oaksford, 2005; Thompson et al., 2005). What sets SSAs apart is that the rationality of their content can also be assessed *objectively*. Experiments 2–3 demonstrate how and why there can be an objective, non-zero probability that the re-evaluation or 'slippage' on which SSAs are predicated can, in fact, occur. In other words, some slopes really are slippery, because their beginning and end are similar.

Crucially, these experiments not only demonstrate exemplar effects (which are well known), they suggest that people understand and naturally take into account such effects when making judgments about SSAs in precisely the way that theories of assimilation and contrast effects (Schwarz & Bless, 1992) and the Social Judgement Involvement theory of persuasion (Sherif, Sherif, & Nebergall, 1965) suggest that they should. Not only do exemplar effects in the context of conceptual vagueness provide a widely applicable underlying mechanism for real-world slippery slopes that underwrites their objective rationality, people seem naturally sensitive to this mechanism in their own subjective evaluations of SSAs: Their ratings of argument strength mirror their categorisation judgements, both within and between participants. That SSA evaluation corresponds so closely to predictions derived from well-established cognitive theory suggests that SSAs are not simply fallacies that are 'wrong but persuasive'. Many of the most famous examples of SSAs seem to involve a category boundary re-appraisal of some kind, including arguments relating to drug classification, civil liberties, abortion and euthanasia. Thus, the linking of category judgements and argument strength seems an appropriate way with which to conduct an empirical examination of SSA evaluation.

Experiment 1 seemed to provide further evidence that people do not combine probabilities and utilities multiplicatively. This is consistent with Evans et al. (2008) who also failed to find such effects. Consequently, our results are consistent with the hypothetical thinking account (Evans, 2007) of the utility conditionals investigated by Evans et al. (2008) but using a different argument form. Our category boundary re-appraisal account provides the psychological mechanism, at least for SSAs, from which the conditional probability is derived in epistemic mental models. However, as we observed, failing to observe an interaction in an ANOVA does not rule out multiplicative effects and so this possibility merits continued research.

On a more general theoretical level, our empirical investigation of classical reasoning fallacies like the slippery slope is part of the *New Paradigm* in reasoning research (Over, 2009) that seeks to investigate inferential relations between degrees of belief and how these interact with our values to arrive at decisions. This move away from the *Deduction Paradigm* (Evans, 2002), has lead to the investigation of a much broader class of arguments than previously investigated, arguments that have practical import like the SSA. Moreover, in assessing the algorithms or mechanisms that underlie these patterns of inference, these experiments show that special purpose reasoning mechanisms may not be required. Existing psychological mechanisms, like category boundary re-appraisal or constraint satisfaction (Oaksford & Chater, 2010b), can be shown to adequately implement conditional probability evaluation, i.e., the Ramsey test. At the computational level, the formal tools of the new paradigm are Bayesian probability and decision theory. While some evidence of normative evaluation of SSAs (consistent with subjective utility theory) was found in Experiment 1, as Evans et al. (2008) also concluded for utility conditionals, the question of whether probabilities and utilities combine as SEU theory predicts remains to be answered.

Appendix A. Materials used in Experiment 1

Topic (i): Euthanasia

(Likely/Very Negative Outcome)

Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to an increase in cases of involuntary euthanasia—or "medical murder". The British Medical Association has warned that once voluntary euthanasia is permitted in some cases, it will be difficult to formulate clear guidelines about when doctors can euthanize patients. The article ends with the view of the author about the future of euthanasia legislation;

> "We should oppose the legalisation of euthanasia in the UK, as it will lead to an increase in the number of instances of 'medical murder'".

(Unlikely/Very Negative Outcome)

Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to an increase in cases of involuntary euthanasia – or "medical murder". The British Medical Association has indicated, however, that there will be extremely clear and strict guidelines about if and when doctors may euthanize patients, and those who break them will be removed from the medical register. The article ends with the view of the author about the future of euthanasia legislation;

> "We should oppose the legalisation of euthanasia in the UK, as it will lead to an increase in the number of instances of 'medical murder'".

(Likely/Less Negative Outcome)

Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to other hospital patients feeling that their lives are not as worthwhile. The British Medical Association has warned that once voluntary euthanasia is permitted, terminally ill patients may start to view their lives as of less worth than healthy individuals. The article ends with the view of the author about the future of euthanasia legislation;

> "We should oppose the legalisation of euthanasia in the UK, as it will lead to other terminally ill patients feeling psychologically damaged by the process".

(Unlikely/Less Negative Outcome)

Whilst flicking through a copy of Ethics magazine, you come across an article on the thorny issue of euthanasia. Despite almost unanimous agreement (from both the medical profession and terminally ill individuals) on the acceptability of helping some patients to end their suffering, opponents claim that the legalisation of voluntary euthanasia will lead to other hospital patients feeling that their lives are not as worthwhile. The British Medical Association has indicated, however, that most hospital patients are unconcerned by the thought of voluntary euthanasia in hospitals. The article ends with the view of the author about the future of euthanasia legislation;

> "We should oppose the legalisation of euthanasia in the UK, as it will lead to other terminally ill patients feeling psychologically damaged by the process".

Topic (ii): developments in cloning technology

(Likely/Very Negative Outcome)

Imagine you have just come back from a lecture about developments in cloning technology. The news that a sheep was successfully cloned for research purposes was received with great praise from the scientific community, but some critics believe that once farm animals are cloned for research purposes, it will only be a matter of time before people will want to 'play god' and clone desirable humans – a form of ethnic cleansing. A cloning expert at the lecture stated that once the technology had been developed, it would be easy to apply the technology to the human genetic code. The lecture ended with an argument for the audience to consider:

> "We should not permit farm animals such as sheep to be cloned for research purposes, as in the future people will want to 'play god' and clone desirable humans – a form of ethnic cleansing."

(Unlikely/Very Negative Outcome)

Imagine you have just come back from a lecture about developments in cloning technology. The news that a sheep was successfully cloned for research purposes was received with great praise from the scientific community, but some critics believe that once farm animals are cloned for research purposes, it will only be a matter of time before people will want to 'play god' and clone desirable humans – a form of ethnic cleansing. A cloning expert at the lecture stated, however, that the technology for cloning the human genetic code, which is much more complex, did not exist and that scientists were uncomfortable about working on human genetics. The lecture ended with an argument for the audience to consider:

> "We should not permit farm animals such as sheep to be cloned for research purposes, as in the future people will want to 'play god' and clone desirable humans – a form of ethnic cleansing."

(Likely/Less Negative Outcome)

Imagine you have just come back from a lecture about developments in cloning technology. The news that a sheep was successfully cloned for research purposes was received with great praise from the scientific community, but some critics believe that once farm animals are cloned for research purposes, it will only be a matter of time before they are cloned for food to feed people, and this could be dangerous. A cloning expert at the lecture stated that once the technology had been developed, it would be easy to apply it over and over again in order to clone farm animals for food. The lecture ended with an argument for the audience to consider:

> "We should not permit farm animals such as sheep to be cloned for research purposes, as in the future people will want to clone farm animals for food and this could be dangerous".

(Unlikely/Less Negative Outcome)

Imagine you have just come back from a lecture about developments in cloning technology. The news that a sheep was successfully cloned for research purposes was received with great praise from the scientific community, but some critics believe that once farm animals are cloned for research purposes, it will only be a matter of time before they are cloned for food to feed people, and this could be dangerous. A cloning expert at the lecture stated, however, that just because the technology had been developed to clone one sheep, the technology does not exist to clone multiple copies of farm animals for food. The lecture ended with an argument for the audience to consider:

> "We should not permit farm animals such as sheep to be cloned for research purposes, as in the future people will want to clone farm animals for food and this could be dangerous".

Topic (iii): the introduction of ID cards

(Likely/Very Negative Outcome)

You attend a discussion session on ID cards and civil rights. You notice that while some people are in favour of voluntary ID cards, there is strong opposition to them nonetheless. The main problem appears to be that further legislation has been suggested whereby the cards would be compulsory and it would be an offence not to carry your ID. The offence would be punishable by a £1000 fine. The

government's legal advisor has indicated that further legislation would be difficult to oppose once voluntary ID cards had been introduced, as the House of Lords support the idea of compulsory ID. The meeting is concluded with the following statement:

> "We oppose voluntary ID cards out of the concern that once they have been introduced, compulsory ID cards will be introduced making it a criminal offence (punishable by a £1000 fine) to not produce I.D. when requested by a Police Officer."

(Unlikely/Very Negative Outcome)

You attend a discussion session on ID cards and civil rights. You notice that while some people are in favour of voluntary ID cards, there is strong opposition to them nonetheless. The main problem appears to be that further legislation has been suggested whereby the cards would be compulsory and it would be an offence not to carry your ID. The offence would be punishable by a £1000 fine. The government's legal advisor has indicated, however, that any further legislation would meet with strong opposition from the House of Lords, who do not support the idea of compulsory ID. The meeting is concluded with the following statement:

> "We oppose voluntary ID cards out of the concern that once they have been introduced, compulsory ID cards will be introduced making it a criminal offence (punishable by a £1000 fine) to not produce I.D. when requested by a Police Officer."

(Likely/Less Negative Outcome)

You attend a discussion session on ID cards and civil rights. You notice that while some people are in favour of voluntary ID cards, there is strong opposition to them nonetheless. The main problem appears to be that further legislation has been suggested whereby the cards would be compulsory and it would be an offence not to carry your ID. If requested to produce your ID. by a Police Officer, you would have to do so within 28 days at your local police station. The government's legal advisor has indicated that further legislation would be difficult to oppose once voluntary ID cards had been introduced, as the House of Lords support the idea of compulsory ID. The meeting is concluded with the following statement:

> "We oppose voluntary ID cards out of the concern that once they have been introduced, compulsory ID cards will be introduced making it a criminal offence to not produce I.D. when requested by a Police Officer."

(Unlikely/Less Negative Outcome)

You attend a discussion session on ID cards and civil rights. You notice that while some people are in favour of voluntary ID cards, there is strong opposition to them nonetheless. The main problem appears to be that further legislation has been suggested whereby the cards would be compulsory and it would be an offence not to carry your ID. If requested to produce your ID. by a Police Officer, you would have to do so within 28 days at your local police station. The government's legal advisor has indicated, however, that any further legislation would meet with strong opposition from the House of Lords, who do not support the idea of compulsory ID. The meeting is concluded with the following statement:

> "We oppose voluntary ID cards out of the concern that once they have been introduced, compulsory ID cards will be introduced making it a criminal offence to not produce I.D. when requested by a Police Officer."

Topic (iv): the cessation of postal deliveries to houses inhabited by vicious dogs

(Likely/Very Negative Outcome)

The County Council of Derbyshire are considering whether to allow postal workers to refuse to deliver mail to houses where vicious dogs are likely to attack them. Opponents of the scheme have claimed that although no-one wants to deny postal workers the right to deliver mail without fear of attack, postal workers will gradually stop delivering mail to any household where there is a dog, as it will be up to them to decide whether the dog is vicious or not. The postal workers' union POSTEC have indicated that the safety of its members must take precedence and accept that the definition of a "vicious dog" is likely to be interpreted widely. The council have issued the following statement:

> "We cannot allow postal workers refuse to deliver to houses where there are 'vicious dogs', in case they stop delivering post to all dog owners whether they are a threat or not"

(Unlikely/Very Negative Outcome)

The County Council of Derbyshire are considering whether to allow postal workers to refuse to deliver mail to houses where vicious dogs are likely to attack them. Opponents of the scheme have claimed that although no-one wants to deny postal workers the right to deliver mail without fear of attack, postal workers will gradually stop delivering mail to any household where there is a dog, as it will be up to them to decide whether the dog is vicious or not. The postal workers' union POSTEC have indicated that the safety of its members must take precedence but deny that the definition of a "vicious dog" will be interpreted so carelessly, adding that there would be severe disciplinary measures for any postal worker refusing to deliver mail where there was no actual threat to their safety. The council have issued the following statement:

> "We cannot allow postal workers to refuse to deliver to houses where there are 'vicious dogs', in case they stop delivering to any dog owners whether they are a threat or not"

(Likely/Less Negative Outcome)

The County Council of Derbyshire are considering whether to allow postal workers to refuse to deliver mail to houses where vicious dogs are likely to attack them. Opponents of the scheme have claimed that although no-one wants to deny postal workers the right to deliver mail without fear of attack, postal workers may decide to not deliver to houses where they know certain aggressive breeds of dog are kept, whether they are actually dangerous or not. The postal workers' union POSTEC have

indicated that the safety of its members must take precedence and accept that the definition of a "vicious dog" is likely to be interpreted widely. The council have issued the following statement:

> "We cannot allow postal workers to refuse to deliver to houses where there are 'vicious dogs', in case they stop delivering to the owners of certain breeds of dogs whether they are a threat or not".

(Unlikely/Less Negative Outcome)

The County Council of Derbyshire are considering whether to allow postal workers to refuse to deliver mail to houses where vicious dogs are likely to attack them. Opponents of the scheme have claimed that although no-one wants to deny postal workers the right to deliver mail without fear of attack, postal workers may decide to not deliver to houses where they know certain aggressive breeds of dog are kept, whether they are actually dangerous or not. The postal workers' union POSTEC have indicated that the safety of its members must take precedence but deny that the definition of a "vicious dog" will be interpreted so carelessly, adding that there would be severe disciplinary measures for any postal worker refusing to deliver mail where there was no actual threat to their safety. The council have issued the following statement:

> "We cannot allow postal workers to refuse to deliver to houses where there are 'vicious dogs', in case they stop delivering to the owners of certain breeds of dogs whether they are a threat or not".

References

Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: Guilford Press.

Albarracin, D., & Wyer, R. S. (2001). Elaborative and nonelaborative processing of a behavior-related communication. *Personality and Social Psychology Bulletin, 27*, 691–705.

American Family Research Council (2004). *The slippery slope of same-sex marriage*.

Areni, C. S., & Lutz, R. J. (1988). The role of argument quality in the elaboration likelihood model. *Advances in Consumer Research, 15*, 197–203.

Boger, G. (2005). Subordinating truth—Is acceptability acceptable? *Argumentation, 19*, 187–238.

Bonnefon, J. F. (2009). A theory of utility conditionals: Paralogical reasoning from decision theoretic leakage. *Psychological Review, 116*, 888–907.

Bonnefon, J. F., & Hilton, D. J. (2004). Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology; Learning, Memory & Cognition, 30*, 28–39.

Corner, A., & Hahn, U. (2007). Evaluating the meta-slope: Is there a slippery slope argument against slippery slope arguments? *Argumentation, 21*, 349–359.

Corner, A., & Hahn, U. (2009). Evaluating science arguments: Evidence, uncertainty & argument strength. *Journal of Experimental Psychology: Applied, 15*, 199–212.

Corner, A., Hahn, U., & Oaksford, M. (2006). The slippery slope argument: Probability, utility and category boundary re-appraisal. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 1145–1151). Austin, TX: Cognitive Science Society.

Edgington, D. (1995). On conditionals. *Mind, 104*, 235–329.

Enoch, D. (2001). Once you start using slippery slope arguments, you're on a very slippery slope. *Oxford Journal of Legal Studies, 21*, 629–647.

Estes, Z. (2003). Domain differences in the structure of artifactual and natural categories. *Memory & Cognition, 31*, 199–214.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978–996.

Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove, UK: Psychology Press.

Evans, J. St. B. T., Neilens, H., Handley, S., & Over, D. (2008). When can we say 'if'? *Cognition, 108*, 100–116.

Govier, T. (1982). What's wrong with slippery slope arguments. *Canadian Journal of Philosophy, 12*, 303–316.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological Review, 114*, 704–732.

Hamblin, C. L. (1970). *Fallacies*. London: Methuen.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition*. Oxford: Oxford University Press.

Heysse, T. (1997). Why logic doesn't matter in the (philosophical) study of argumentation. *Argumentation, 11*, 211–224.

Holtug, N. (1993). Human gene therapy: Down the slippery slope. *Bioethics, 7*, 402–419.

Johnson, R. H. (2000). *Manifest rationality: A pragmatic theory of argument*. Mahwah, NJ: Hillsdale.

Johnson, B. T., Smith-McLallen, A., Killeya, L. A., & Levin, K. D. (2004). Truth or consequences: Overcoming resistance to persuasion with positive thinking. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 215–233). New Jersey: Laurence Erlbaum.

Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition, 30*, 340–352.

Kalish, C. W. (2002). Essentialist to some degree: Beliefs about the structure of natural kind categories. *Memory & Cognition, 30*, 340–352.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116*, 20–58.

Kirk, R. E. (1995). *Experimental design – Procedures for the behavioral sciences*. London: Brooks/Cole.

Labov, W. (1973). The boundaries of words and their meanings. In C. J. Bailey & W. W. Shuy (Eds.), *New ways of analysing variation in English* (pp. 340–373). Washington, DC: Georgetown University Press.

Lakoff, G. (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors*. Cambridge: Cambridge University Press.

Lamb, D. (1988). *Down the slippery slope – arguing in applied ethics*. London: Croom Helm.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General, 124*, 161–180.

Landsheer, J., van den Wittenboer, G., & Maassen, G. (2006). Additive and multiplicative effects in a fixed 2×2 design using ANOVA can be difficult to differentiate: Demonstration and mathematical reasons. *Social Science Research, 35*(1), 279–294.

Launis, V. (2002). Human gene therapy & the slippery slope argument. *Medicine, Health Care and Philosophy, 5*, 169–179.

Lode, E. (1999). Slippery slope arguments and legal reasoning. *California Law Review, 87*, 1468–1543.

Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*, 462–472.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M. (1988a). Exemplar-based accounts of the relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 700–708.

Nosofsky, R. M. (1988b). Similarity, frequency and category representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 54–65.

Oakley, J., & Cocking, D. (2005). Consequentialism, complacency and slippery slope arguments. *Theoretical Medicine & Bioethics, 26*, 227–239.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*(8), 349–357.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (Eds.). (2010a). *Cognition and conditionals: Probability and logic in human thinking*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2010b). Conditionals and constraint satisfaction: Reconciling mental models and the probabilistic approach? In M. Oaksford & N. Chater (Eds.), *Cognition and*

conditionals: Probability and logic in human thinking (pp. 309–334). Oxford: Oxford University Press.

Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology, 58*, 75–85.

Osherson, D., Smith, E. E., Wilkie, O., Lòpez, A., & Shafir, E. (1990). Category based induction. *Psychological Review, 97*, 185–200.

Over, D. (2009). New paradigm psychology of reasoning. *Thinking and Reasoning, 15*, 431–438.

Over, D. E., Manktelow, K. I., & Hadjichristidis, C. (2004). Conditions for the acceptance of deontic conditionals. *Canadian Journal of Experimental Psychology, 58*, 96–105.

Perham, N. R., & Oaksford, M. (2005). Deontic reasoning with emotional content: Evolutionary psychology or decision theory? *Cognitive Science, 29*, 681–718.

Ramsey, F. P. (1931). *The foundations of mathematics and other logical essays*. London: Routledge and Kegan Paul.

Ricco, R. B. (2007). Individual differences in the analysis of informal reasoning fallacies. *Contemporary Educational Psychology, 32*, 459–484.

Rips, L. J. (1989). Similarity, typicality and categorisation. In S. Vosinadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press.

Rizzo, M. J., & Whitman, D. G. (2003). The camel's nose is in the tent: Rules, theories and slippery slopes. *UCLA Law Review, 51*, 539–592.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 574–605.

Savage, L. J. (1954). *The foundations of statistics*. John Wiley: New York.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: Assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 217–245). Hillsdale, NJ: Erlbaum.

Sherif, C. W., Sherif, M., & Nebergall, R. E. (1965). *Attitude and attitude change: The social judgment–involvement approach*. Philadelphia: W.B. Saunders.

Siegel, S., & Castellan, N. J. Jr., (1988). *Non-parametric statistics for the behavioral sciences*. New York: McGraw Hill.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231–280.

Thompson, V. A., Evans, J. St. B. T., & Handley, S. J. (2005). Persuading and dissuading by conditional argument. *Journal of Memory & Language, 53*, 238–257.

van der Burg, W. (1991). The slippery slope argument. *Ethics, 102*, 42–45.

van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation – The pragma-dialectical approach*. Cambridge: Cambridge University Press.

van Enschot-Van Dijk, R., Hustinx, L., & Hoeken, H. (2003). The concept of argument quality in the elaboration likelihood model. In F. H. van Eemeren, J. Anthony Blair, C. A. Willard, & A. Francisca Snoeck Henkemans (Eds.), *Anyone who has a view: Theoretical contributions to the study of argumentation* (pp. 319–333). Dordrecht: Kluwer Academic Publishers.

Volokh, E. (2003). The mechanisms of the slippery slope. *Harvard Law Review, 116*, 1026–1137.

Wahlsten, D. (1991). Sample size to detect a planned contrast and a one degree-of-freedom interaction effect. *Psychological Bulletin, 110*, 587–595.

Walton, D. (1992). *Slippery slope arguments*. Oxford: Clarendon Press.

Wegener, D. T., & Petty, R. F. (1995). Flexible correction processes in social judgment: The role of naïve theories in corrections for perceived bias. *Journal of Personality & Social Psychology, 68*, 36–51.

Woods, J., Irvine, A., & Walton, D. (2004). *Critical thinking, logic & the fallacies*. Toronto: Prentice Hall.