Title
Training Sensitivity to Biased Samples in Inductive Reasoning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Goswami, Spriha
Hayes, Brett

Publication Date
2023

Peer reviewed

Training Sensitivity to Biased Samples in Inductive Reasoning

Spriha Goswami (spriha.goswami@unsw.edu.au)
Brett K. Hayes (b.hayes@unsw.edu.au)

School of Psychology, University of New South Wales
Sydney, 2052, Australia

Abstract

Environmental restrictions often permit the sampling of some items while excluding others. Such restrictions are termed *sampling frames*, whereby items can be selected based on their category membership (*category frame*), or possession of a target property (*property frame*). According to Bayesian principles, narrower property generalization is expected when a sample is subject to property sampling than category sampling. The current work examined whether sensitivity to such sampling frames could be increased through training with worked examples and practice. Experiment 1 found that training in property or category sampling enhanced sensitivity to that frame relative to a no-training control. Experiment 2 employed a pre-post design where all participants received training in both frames. A positive training effect was found, but only for those with a poor understanding of sampling frames on the pre-test. This work indicates the viability of appropriate training for increasing understanding of the implications of sample selection mechanisms.

Keywords: inductive reasoning; sampling frames; training

Imagine that a new restaurant opened up in your neighborhood and you wanted to explore the quality of its cuisine. On five occasions, you tried and enjoyed different dishes. Given this new knowledge, you may infer that the chef is talented, and you will likely enjoy the rest of the menu as well. This is an example of *inductive reasoning*, in which we use prior knowledge to inform the extent to which we generalize information from a sample (e.g., the initial five dishes), to a novel population (e.g., the entire menu) (Denison & Xu, 2019; Hayes & Heit, 2018). Inductive reasoning is a crucial part of the human experience, as we make predictions every day based on what we already know, whether it pertains to trying a new restaurant, planning our commutes based on traffic, or taking an umbrella with us on a cloudy day. Indeed, our ability to make such intuitive statistical inferences emerges as early as 8 months of age without prior teaching, indicating that it may even be an innate human ability (Xu & Garcia, 2008).

Research on inductive reasoning typically uses property induction tasks, in which participants learn about items in a sample which possess a target property. Participants then use their experience with the sample to infer whether the property is likely to generalize to novel items (Hayes, Navarro et al., 2019). Early work identified a range of factors relating to the *contents* of an evidence sample which affect how the sample is used to make inferences about property generalization. For example, people are more likely to generalize a property to novel category members if they have observed a *large* sample of items from the same category that share the property, as compared to a smaller sample (Feeney, 2007; Hayes & Heit, 2018). Property generalization is also more likely when the sample contains a diverse range of category members that share the property (Feeney & Heit, 2011; Hayes & Heit, 2018; Rhodes & Brickman, 2010).

Sampling Assumptions and Inductive Inference

Although sample contents are important for inductive inference, recent work has also highlighted the key role of *sampling assumptions* – a learner's beliefs about *how the sample contents were selected* (Ransom et al., 2022; Tenenbaum & Griffiths, 2001). A crucial finding is that the same sample of evidence can lead to very different inductive inferences depending on one's assumptions about the sampling process.

One example of the role of sampling assumptions involves beliefs about the intentionality of sample selection. People draw very different inferences from a sample of evidence depending on whether they believe the sample was selected by a helpful agent, often termed *strong sampling*, or selected at random, a form of *weak sampling* (Navarro, Dry & Lee, 2012; Shafto & Goodman, 2008). For instance, while both adults and children are more likely to generalize from a diverse sample of items (Feeney, 2007; Rhodes & Brickman, 2010), this effect is attenuated when the sample is believed to have been selected at random (Hayes, Navarro et al., 2019; Ransom, Perfors & Navarro, 2016). Similar results have been found regarding sample size, in which the effect of a larger sample size is moderated by the intentionality of the sample selection (Hayes, Banner et al., 2019).

Sampling Frames and the Frames Effect

Another line of work demonstrating the impact of sampling assumptions on inductive inference, which is pertinent to the present study, concerns peoples' beliefs about *sampling frames* – environmental restrictions that permit the selection of some items in a sample while excluding others (Hayes, Banner & Navarro, 2017; Lawson & Kalish, 2009). A *category frame* limits sampling to instances that belong to a specific category while excluding members of another category. A *property frame* limits sampling to instances that share a specific property. A key idea is that the same sample of evidence can yield very different inferences depending on whether the sample was collected under a category or a property frame. To illustrate, say you wanted

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

to know whether a new fantasy television series is palatable for the general public. You have a sample of twenty people who liked the program. If this sample was collected under a category frame, (e.g., from attendees at a fantasy convention) then little can be concluded about the likely reactions to the program by other types of viewers. In this case, the absence of other viewers from the sample is explained by the category frame, rendering the sample relatively uninformative for generalization. In contrast, under a property frame, all types of items that have the property of interest can be observed. For example, we might have a sample of the *first* 20 people who declared they liked the fantasy program. If this sample was found to contain people from varied backgrounds, then you could infer the program is likely to have broad appeal. If, however, the sample only contained known fantasy fans (even though any type of viewer could have appeared in the sample), then we would infer that other types of viewers are unlikely to enjoy it.

This intuitive example illustrates general principles of property inference that can be derived from Bayesian models of induction (Hayes et al., 2019; Navarro et al., 2012; Ransom et al., 2022). According to such models, the breadth of property generalization from a given sample depends on the sampling frame. If all the sample members that share a novel property belong to a single category, and this sample was selected via property sampling, there should be little generalization of the property beyond the sample. In contrast, under category sampling, it remains possible that the target property generalizes beyond the sample. Research into the effects of sampling frames on inductive inference suggests that many people reason in line with these Bayesian predictions (e.g., Hayes, Banner et al., 2019; Lawson & Kalish, 2009). In these studies, learners are presented with a single sample of instances that share a novel target property (e.g., 10 small birds with "plaxium blood"). Different groups are told that the sample was collected under either category sampling (e.g., only small birds were sampled) or property sampling (the sample was made up of the first 10 animals found to have plaxium blood). In a subsequent test phase, learners are asked to infer whether the property generalizes to other categories that vary in similarity to the training sample.

Under property sampling, many people show relatively narrow generalization – only generalizing the target property to items that are highly similar to the sample (Hayes, Navarro et al., 2019). In contrast, when the training sample is said to have been collected via category sample, the target property is often generalized to a range of other categories. This pattern of narrower generalization under a property frame than a category frame is termed the *frames effect* and has been replicated across a range of stimuli and cover stories (e.g., Hayes, Navarro et al., 2019; Hayes et al., 2022; Lawson & Kalish, 2009; Ransom et al., 2022).

Individual Differences in the Frames Effect

Although previous research indicates that individuals are sensitive to the effect of sampling frames on property induction, we are only beginning to understand the cognitive mechanisms that are involved. For instance, recent research has investigated whether sampling frames moderate the *encoding* or *retrieval* of sample items (Ransom et al., 2022). When sampling frame mechanisms are made apparent prior to observation of the sample, the generalization tendencies of participants tend to be the same as observed in previous research (i.e., Hayes, Banner et al., 2019). However, sampling frames do *not* impact generalization when frames are introduced *after* sample items are observed (Ransom et al., 2022). These results suggest that individuals take selection biases into account *as they encode information* but cannot retrospectively apply their frames knowledge to what they have already taken in.

Just as there are specific cognitive mechanisms that are differentially susceptible to sampling frames, individuals also vary in their own sensitivity to these types of selection biases. Most sampling frames research conducted to date has been between-subjects, but a recent within-subjects study assessed *individual* sensitivity to the frames effect (Hayes et al., 2022). Participants in these studies were asked to make property inferences based on multiple training samples, some of which were presented with category framing instructions and others presented with property frames. The general frames effect was replicated, with broader generalization of target properties following category sampling as compared to property sampling. However, analysis of individual patterns of generalization found that this frames effect was driven by only around half of the participants. Most of the remaining participants demonstrated *no difference* in the pattern of their generalization judgments in response to category and property sampling frames. Either these individuals did not understand the difference between the frames or did not apply their understanding to their judgments. Regardless of the explanation, it is of interest to investigate whether these "frames-insensitive" individuals can be trained to understand and/or implement the frames effect in their future generalizations.

The Current Studies

The main aim of the current studies was to examine whether appreciation of the inferential implications of category and property sampling frames can be increased through training. Several previous studies have attempted to enhance the quality of inductive reasoning through relevant training. In classic work, Fong, Krantz, and Nisbett (1986) provided brief training to participants about the formal properties of the law of large numbers. This training improved participants' performance in solving inferential problems that involved an appreciation of the impact of sample size or regression to the mean.

Training via the practice of various inductive reasoning problems and subsequent feedback has been successful in enhancing the reasoning performance of seven-year-old children on analogous tasks (Tomic, 1995). Likewise, young children who do not initially appreciate the importance of sample diversity for inductive inference can be trained to do so (Rhodes & Brickman, 2010). Further, it has been shown that the provision of feedback on correct and incorrect responses effectively increases the accuracy of adults' deductive reasoning (Khemlani & Moore, 2012).

Such studies show that appropriate application of statistical or logical principles can be enhanced through training. However, to our knowledge, no previous studies have sought to improve people's understanding of the implications of different types of sampling assumptions for inductive inference.

In the two current studies, we provided participants with a brief training regime that comprised worked examples of category and/or property frames, together with feedback identifying what sorts of inferences could be drawn from each type of frame. Following the general approach used in previous studies (e.g., Fong et al., 1986), the training examples used relatively abstract materials (sampling frames versions of balls-and-urns problems). Property inference tests involving more familiar animal or social categories were used to assess whether people exhibited the frames effect of broader property generalization following category than property sampling. In Experiment 1, the effects of training were investigated in a between-subjects design where the inferences of participants who received frames training were compared to those of a baseline no-training group. Experiment 2 examined the effects of training at an individual level in a pre-post design, where all participants completed property induction tasks before and after training.

Experiment 1: Between-Subjects Training in Sampling Frames

Method

Participants 201 participants (121 male, 73 female, 7 non-binary; M_{age} = 25.75, SD_{age} = 7.53) were recruited via Prolific Academic in exchange for payment at a rate of 7 GBP per hour.

Design The experiment used a 2 (sampling frame: category vs. property) × 2 (training: no training baseline vs. training) × (4) (test items) design, with repeated measures on the last factor. Participants were randomly allocated to either the training or baseline condition. In each condition, participants completed generalization judgments for four scenarios with either a category or property sampling frame (cell n's: baseline category frame, n = 58; baseline property frame n = 44; training category frame, n = 49; training property frame, n = 50). When making generalization judgments for each scenario, all participants responded to four types of test items based on their similarity to the observed sample (familiar, near transfer, medium transfer, and far transfer).

Materials and Procedure Induction problems and training examples were presented with text and cartoon images obtained from Google Images and Microsoft Word Icons, and the experimental procedure was created using the JS Psych programming language.

All participants completed four property induction problems, presented in random order. In the training groups, this was preceded by either category or property training. In each of the four induction problems, participants were asked to use an observed sample containing six instances from a single category that shared a property (e.g., six dogs owned by children in Town X), to make inductive inferences (e.g., whether other children in the town are likely to own dogs, cats, or fish). In all problems the six sample items belonged to the same category, with minor variations in their visual appearance to convey that they were discrete sample items (see Figure 1). Other induction problems included stimuli/cover stories concerning which sports children played, what transport was used to go to work, and whether aliens were friendly.

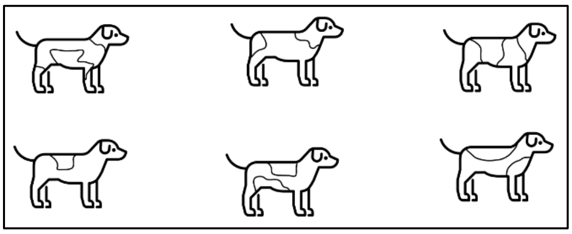

Figure 1: Example of Sample Items

In the category frame condition for each problem, an environmental restriction was described which restricted sampling to a single category (e.g., observations were made at a dog park, which prevented sampling of cats and fish). In the property frame condition, items were said to be sampled because they had the target property (e.g., observations were made at Bring Your Pet to School Day, whereby any type of pet owned by children in the town could have been observed). The six sample items were presented sequentially on a computer screen.

After observing all sample items, participants made generalization judgments about four test items. Familiar test items were identical to one of the previously observed sample items, while near, medium, and far transfer items were three categories varying in similarity to the sample (e.g., a dog that was visually different to those in the sample, a cat, and a fish), as shown in Figure 2. The familiar test item was presented first, followed by the near, medium, and far items in a randomized order. For each test item participants had to infer whether it shared the target property with the sample items (e.g., whether it was likely to be owned by children from Town X). Responses were made by pressing a button that said either "Most Likely No" (coded as 0 for analysis) "Equally Likely Yes or No" (coded as 1), or "Most Likely Yes" (coded as 2).

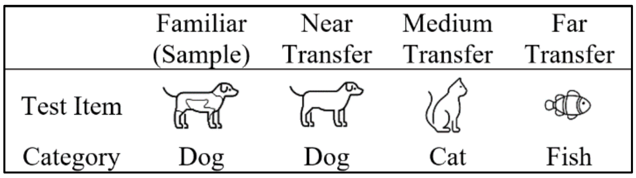

Figure 2: Example of Test Items

Prior to completion of the induction problems, participants in the training conditions observed a worked example of their assigned sampling frame (either category or property). In the training example, 30 colored balls (10 each of three colors) were shown in an opaque bag. Participants were told that some balls may have prizes inside, and they had to determine which ones possessed this property. In *category frame training*, participants were told that two of the three sets of colored balls were stuck inside the bag and could not be sampled. As such, all six sample items were of the same color, and possessed the target property (i.e., had prizes). Participants were subsequently informed that based on this sample, they could conclude that balls with the sampled color would most likely contain prizes, but that little could be inferred about the other colored balls – for these, the most appropriate response to the inference question was "equally likely yes or no".

In *property frame training*, the general set-up was similar with the crucial exception that the sample was said to be made up of balls that were found to contain prizes in a previous game. A comparison between the category and property frame set-up is shown in Figure 3. All six sample items were again of the same color. As such, participants were informed that other balls of the sampled color would most likely contain prizes, but that balls with the non-sampled colors most likely would not. All participants then observed another ball scenario (with different colors) in their allocated frame and made generalization judgments. With each judgment, participants were given feedback informing them of the correct response. Participants in the training conditions then proceeded to complete the four induction test problems.

Figure 3: Category vs. Property Frame in Training Task

Results and Discussion

Generalization judgments for each of the four types of test items (familiar, near, medium, far) were averaged across the four induction problems to give mean generalization scores between 0 and 2. As previous research suggests that the greatest evidence of the frames effect is observed in items more dissimilar to the observed sample (e.g., Hayes, Navarro et al., 2019), and since there was no significant difference between generalization for medium and far transfer items ($F_{1, 197}$ = 2.556, p = .111, η^2 = 5.281e^{-4}), we collapsed mean generalization scores across the medium and far items into a single *novel* transfer category. Group generalization data for familiar, near, and novel test items are shown in Figure 4. These data were analyzed using a frequentist repeated measures analysis of variance (ANOVA).

There was a significant effect of test item, averaged across frame and condition, whereby generalization was greatest towards familiar items ($M_{familiar}$ = 1.928), then near transfer items (M_{near} = 1.591), and lowest towards novel transfer items (M_{novel} = 0.412; $F_{2, 394}$ = 1070.881, p < .001, η^2 = .714). There was a significant interaction between test items and sampling frame, whereby generalization to novel items was narrower under a property frame than a category frame ($F_{1, 394}$ = 57.021, p < .001, η^2 = .038). Crucially, there was also a three-way interaction between test items, sampling frame, and training, whereby this frames effect was greater in the training condition compared to the baseline condition ($F_{1, 394}$ = 4.681, p = .01, η^2 = .003). As shown in Figure 4, the difference in generalization to novel items under category and property framing was larger in the training than in the baseline condition.

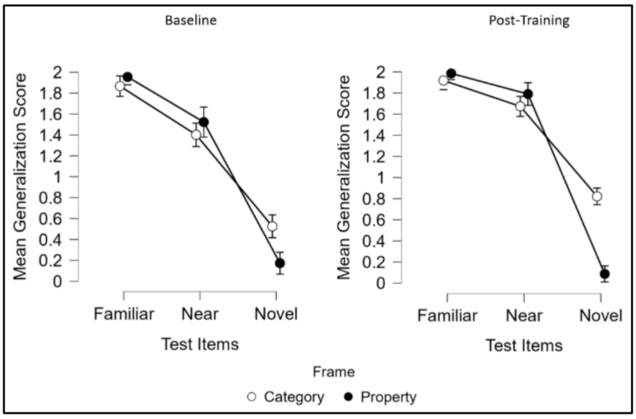

Figure 4: Generalization Responses in Experiment 1. Error bars denote 95% confidence intervals.

These results reflect a successful replication of the frames effect observed in previous research (e.g., Hayes et al., 2017). Importantly, this study is the first to also demonstrate a successful *training effect* in the application of sampling frames to property generalization. Participants who observed an analogous but more abstract worked example of sampling frames, and were given feedback on their practice judgments, subsequently showed a greater frames effect than those who received no training.

It is important to note, however, that the results of most previous sampling frames research, including the present study, are based on aggregate data that do not reflect individual variability in sensitivity to sampling frames. Such variability is clear in the results of Hayes et al. (2022), whereby only around half of the participants demonstrated clear evidence of the frames effect. As such, Experiment 2 examined the effect of our training paradigm using a within-subjects design to assess shifts in individual understanding of the implications of each sampling frame.

Experiment 2: Within-Subjects Training in Sampling Frames

In this study, individuals' use of category and property frames for property generalization was examined before and after training. Based on the Experiment 1 results, we predicted that an overall frames effect would be replicated and that this effect would be augmented post-training. A particularly interesting exploratory question was whether participants who initially showed little understanding of the implications of different frames would show a reliable frames effect after training.

Method

Participants 170 participants (103 male, 65 female, 2 non-binary; M_{age} = 26.98, SD_{age} = 7.81) were recruited via Prolific Academic in exchange for payment at a rate of 7 GBP per hour.

Design and Procedure The experiment used a fully within-subjects design: (2) (sampling frame: category vs. property) × (2) (pre-test vs. post-test) × (4) (test item), whereby each participant made generalization judgments to four test items across four pairs of induction problems (four involving a category frame and four involving a property frame). Two problems involving category frames and two problems involving property frames were presented before training. Four problems with an analogous structure were presented after training. The structure of these induction problems was identical to Experiment 1: each involved presentation of a sample of six instances that had a target property, selected via a category or property frame. Participants then inferred how likely it was that the target property was shared by four test items (familiar, near, medium, and far transfer). Because participants in this study completed more induction problems than in Experiment 1 (a total of eight as compared to four in the previous study), new problem contents were generated. The additional problems contained cover stories about which fish ate specific seaweed, which birds ate specific grubs, children's preferred ice cream flavors, and which sea creatures had colored blood. Allocation of contents to category/property roles and to pre-test vs. post-test presentation was counterbalanced across participants.

The study commenced with the presentation of the four pre-test problems. Training in category and property frames was administered immediately after the pre-test. The training regime was the same as in Experiment 1 except that, in this case, all participants received training about both category and property frames. The four post-test problems were administered immediately after training. Participants were randomly allocated to one of two presentation order conditions for induction problems and training examples, whereby approximately half were first presented with category problems at both pre- and post-tests and received category frame training before property frame training. The remainder were presented with problems and training in the opposite frame order. This served as a control on the time interval between the completion of test problems and training involving the same type of frame.

Results and Discussion

Generalization judgments for each of the test items in each induction problem were scored on the same three-point scale as in Experiment 1. Generalization scores were again collapsed over the two novel (medium and far) test items. Responses were averaged across problems involving the same type of frame to give pre-test and post-test scores for each type of test item – which are shown in Figure 5.

A preliminary analysis confirmed there were no significant effects of frame order ($F_{1, 168}$ = 2.517, p = .115, η^2 = .001), so this variable was excluded from subsequent analyses. There was a significant effect of test item, whereby generalization was greatest for familiar items ($M_{familiar}$ = 1.92), then near transfer items (M_{near} = 1.68), and lowest towards novel transfer items (M_{novel} = 0.454; $F_{2, 338}$ = 1600.414, p < .001, η^2 = .737). There was a significant interaction between test items and sampling frame – generalization to novel items was narrower under a property frame than a category frame ($F_{2, 338}$ = 74.882, p < .001, η^2 = .016). However, the three-way interaction between test items, sampling frame, and training was not-significant ($F_{2, 338}$ = 2.182, p = .114, η^2 = 3.516e^{-4}).

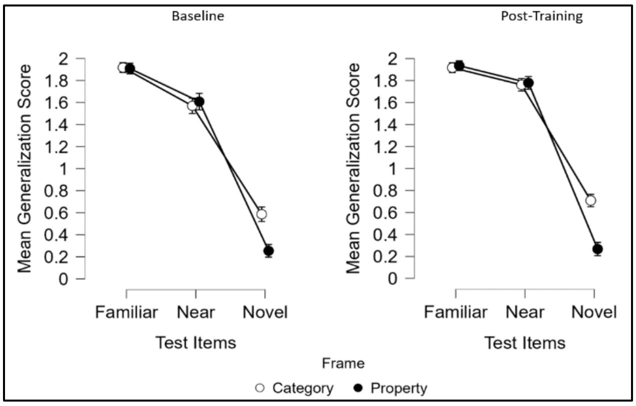

Figure 5: Generalization Responses in Experiment 2. Error bars denote 95% confidence intervals.

Further analysis was conducted on the subset of data from participants who did not demonstrate sensitivity to the difference between sampling frames at baseline (i.e., those whose average property frame generalizations were equal to or greater than their average category frame generalizations; n = 84, 49%). Mean responses from these "frames-insensitive" participants are shown in Figure 6.

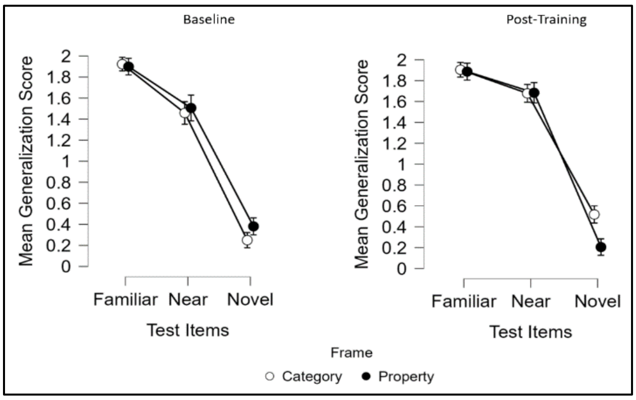

Figure 6: Generalization Responses from "Frames-Insensitive" Participants in Experiment 2. Error bars denote 95% confidence intervals.

The figure shows that the general pattern of results for this subset of participants was similar to the complete sample. Notably, however, this group showed a significant three-way interaction between test items, sampling frame, and training ($F_{2, 166}$ = 13.267, p < .001, η^2 = .004). The figure shows that, for these participants, there was an increase in the frames effect after training.

This within-participants study replicated the sampling frames effect from Experiment 1 and many previous studies such that overall, people showed narrower property generalization to novel transfer items from a sample selected via property sampling, as compared to a sample selected via category sampling. Training tended to increase this effect of sampling frames, but the result was not statistically reliable.

The current study also allowed us to examine individual sensitivity to the implications of different sampling frames. As in Hayes et al. (2022), in the pre-test we found that around half the participants showed little sensitivity to sampling frames. Training in category and property framing did, however, benefit this sub-group – increasing the size of the frames effect in the post-test.

The modest training effects obtained in this study may seem surprising when compared with the robust training effect found in Experiment 1. However, demonstrating an effect of training on post-test performance was likely to be more difficult in this study compared to the previous experiment. For an individual to show an effect of training, they would not only have to understand the respective category and property frame training examples but would *also* have to correctly map the principles from these examples to the appropriate types of post-test problems. Given these complexities, we see the success of training for participants who showed little initial understanding of frame implications, as encouraging.

General Discussion

A considerable body of previous work has shown that many but not all people grasp the implications of different sampling frames for property generalization – showing narrower generalization to novel transfer items following exposure to samples selected via property as compared with category sampling (e.g., Hayes, Banner et al., 2019). Although such frame effects are usually robust in aggregate data, there is also evidence of considerable individual variability in sensitivity to sampling frames.

The current studies, therefore, attempted to enhance sensitivity to the implications of sampling frames via training with worked examples and practice with feedback. Experiment 1 found that training separate groups on the implications of category or property frames led to an enhanced sampling frames effect. In a more complex, within-subjects design, Experiment 2 found that training enhanced sensitivity to the implications of sampling frames, but only for those participants who showed little initial sensitivity in the pre-test. Inspection of Figures 4 and 6 suggests that the augmentation of the frames effect was primarily driven by an upwards shift in generalization judgments towards novel items under a category frame. While it is possible that generalization judgments under a property frame also experienced a downward shift, this is difficult to evaluate due to floor effects.

The core principle underlying sampling frames effects concerns the reason *why* the observed sample only contains members of a single category that have the target property. In category sampling, the absence of instances from other categories is attributable to the frame. This renders the sample uninformative about whether the target property generalizes to other categories. By contrast, in property sampling, instances from any category *could* be observed, as long as they have the target property. In this case, the fact that only members of a single category are observed is highly informative – indicating that the property does not generalize beyond that category. Our training regime aimed to promote understanding of these principles via the use of balls-and-urns examples that had little surface similarity to any of the test induction problems. The modest training effects that we observed suggest that people are capable of a) grasping some of the abstract principles that underlie inductive reasoning based on samples of evidence, and b) applying these principles to problems involving different stimuli and cover stories.

In this respect, the success of our training approach is in line with previous demonstrations of the positive effects of training in other types of inductive (Fong et al., 1986; Tomic, 1995) and deductive reasoning (Khemlani & Moore, 2012). The current work, however, is novel in that our training targeted people’s sampling assumptions – their understanding of the way that sample instances are selected and the implications for property generalization. This work also shows that even though many people may show “myopia” about the implications of sample selection mechanisms for inductive reasoning (cf. Fiedler, 2012) this can be overcome through appropriate training.

That said, the modest training results of Experiment 2 suggest that there is room for further investigation of frames training, with a view to increasing its effectiveness. Previous work (Hayes et al., 2022) has shown that sensitivity to an understanding of sampling frames (in the absence of training) is positively correlated with working memory capacity and level of cognitive reflection (Frederick, 2002). Such individual differences may also mediate the effectiveness of frames training.

A further issue for future research is to undertake a more fine-grained examination of how frames training changes the way that people process sample instances. Previous empirical work and Bayesian modeling (Ransom et al., 2022) suggest that information about sample selection mechanisms affects the encoding rather than the retrieval of sample instances. In property sampling, each new sample instance provides further evidence about whether the properties of sample instances generalize to other items. This may prompt people to pay more attention to sample instances in property sampling than in category sampling. Future work could examine whether patterns of attention to sample instances change as a result of training in the principles that underlie property or category sampling.

In sum, the current results provide encouraging evidence that peoples’ appreciation of the implications of different types of sample selection mechanisms can be improved through brief training. Many if not most of the samples we encounter and use as a basis for inference outside the laboratory are subject to some form of selection bias (Hogarth, Lejarraga & Soyer, 2015). An important implication of our results is that appropriate forms of training can help people to understand the implications of such selection biases for their inferences and judgments.

Acknowledgements

This work was supported by an Australian Research Council Discovery Grant DP190101224 to BKH, and an Australian Government Research Training Program (RTP) Scholarship awarded to SG.

References

Denison, S. & Xu, F. (2019). Infant statisticians: The origins of reasoning under uncertainty. *Perspectives on Psychological Science, 14*(4), 499-509. https://doi.org/10.1177/1745691619847201

Feeney, A. & Heit, E. (2011). Properties of the diversity effect in category-based inductive reasoning. *Thinking & Reasoning, 17*(2), 156-181. https://doi.org/10.1080/13546783.2011.566703

Feeney, A. (2007). How many processes underlie category-based induction? Effects of conclusion specificity and cognitive ability. *Memory & Cognition, 35*(7), 1830-1839. https://doi.org/10.3758/BF03193513

Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 1–55). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-394293-7.00001-7

Fong, G. T., Krantz, D. H. & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18*(3), 253-292. https://doi.org/10.1016/0010-0285(86)90001-0

Frederick, S. (2002). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42. https://doi.org/10.1257/089533005775196732

Hayes, B. K. & Heit, E. (2018). Inductive reasoning 2.0. *Cognitive Science, 9*(3), e1459. https://doi.org/10.1002/wcs.1459

Hayes, B. K., Banner, S. & Navarro, D. J. (2017, July 26-29). Sampling frames, Bayesian inference, and inductive reasoning. In G. Gunzelmann, A. Howes, T. Tenbrink & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 488-493). Austin, TX: Cognitive Science Society.

Hayes, B. K., Banner, S., Forrester, S. & Navarro, D. J. (2019). Selective sampling and inductive inference: Drawing inferences based on observed and missing evidence. *Cognitive Psychology, 113*, 101221. https://doi.org/10.1016/j.cogpsych.2019.05.003

Hayes, B. K., Liew, S. X., Desai, S. C., Navarro, D. J. & Wen, Y. (2022). Who is sensitive to selection biases in inductive reasoning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0001171

Hayes, B. K., Navarro, D. J., Stephens, R. G., Ransom, K. & Dilevski, N. (2019). The diversity effect in inductive reasoning depends on sampling assumptions. *Psychonomic Bulletin & Review, 26*(3), 1043-1050. https://doi.org/10.3758/s13423-018-1562-2

Hogarth, R. M., Lejarraga, T. & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science, 24*(5), 379-385. https://doi.org/10.1177/0963721415591878

Khemlani, S. & Moore, A. (2012). Evaluative feedback can improve deductive reasoning. In N. Miyake, D. Peebles & R. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1780-1785). Austin, TX: Cognitive Science Society.

Lawson, C. A. & Kalish, C. W. (2009). Sample selection and inductive generalization. *Memory & Cognition, 37*(5), 596-607. https://doi.org/10.3758/MC.37.5.596

Navarro, D. J., Dry, M. J. & Lee, M. D. (2012). Sampling assumptions in inductive generalization. *Cognitive Science, 36*(2), 187-223. https://doi.org/10.1111/j.1551-6709.2011.01212.x

Ransom, K. J., Perfors, A. & Navarro, D. J. (2016). Leaping to conclusions: Why premise relevance affects argument strength. *Cognitive Science, 40*(7), 1775-1796. https://doi.org/10.1111/cogs.12308

Ransom, K., Perfors, A., Hayes, B. K. & Desai, S. C. (2022). What do our sampling assumptions affect: How we encode data or how we reason from it? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0001149

Rhodes, M. & Brickman, D. (2010). The role of within-category variability in category-based induction: A developmental study. *Cognitive Science, 34*(8), 1561-1573. https://doi.org/10.1111/j.1551-6709.2010.01137.x

Shafto, P. & Goodman, N. D. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1632-1637). Austin, TX: Cognitive Science Society.

Tenenbaum, J. B. & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629-640. https://doi.org/10.1017/S0140525X01000061

Tomic, W. (1995). Training in Inductive Reasoning and Problem Solving. *Contemporary Educational Psychology, 20*, 483-490. https://doi.org/10.1006/ceps.1995.1036

Xu, F. & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *PNAS 105*(13), 5012-5015. https://doi.org/10.1073/pnas.0704450105