Social categories vs. individuating information:

Re-examining the building blocks of impression formation

Thalia H. Vrantsidis and William A. Cunningham

Department of Psychology, University of Toronto

Preprint of submitted paper

Abstract

To understand how people form impressions and make inferences about others, theories have distinguished between two broad types of information – social categories (e.g. demographics, occupations) and individuating information (e.g. traits, behaviors). While some theories argue that categories are easier to process and use than individuating information, so that people overuse categories as a heuristic to simplify their inferences, other theories argue that these two types of information are processed and used similarly. The current research therefore re-examined two key claims of theories of categories as heuristics: 1) using this heuristic will lead categories to be overused; 2) this will occur more in conditions that reduce effortful processing (e.g. informational fit, cognitive load). The current research developed a Bayesian modelling approach to more clearly identify category overuse, and overcame several methodological challenges of previous work, thus providing some of the first clear and systematic tests of these claims. Eight studies tested these claims across a range of conditions, with studies 1-2 varying informational fit, and studies 3-8 varying cognitive load. Overall, these studies found little support for these claims. This suggests that categories and individuating information may not be processed or used differently, in line with alternative theories that treat these two types of information similarly. More broadly, it suggests the need to move beyond these broad content distinctions and focus more directly on underlying mechanisms, to more carefully examine claims about heuristics and biases in people's social inferences, and to further explore when and why categories might be overused.

Keywords: stereotyping and impression formation, social categories and individuating information, Bayesian modelling, continuum model, stereotypes as heuristics

Social categories vs. individuating information: Re-examining the building blocks of impression formation

To function in a social world, people need to form impressions and make inferences about others, and they can do this based on a variety of information they might have about a person. For example, one might know that someone is 29 years old, female, an accountant, got good marks in school, recently hosted a party, and dislikes spicy food. Depending on the context, this information could be used to infer the person's personality, their competence as a potential employee, or any number of other inferences. In trying to understand how people make these inferences, as well as how these inferences can go wrong, research has often distinguished between two broad types of information: sometimes referred to as 'social categories', which include a person's demographics (e.g. age, race, gender), occupation, or social roles (e.g. housewife); and 'individuating information', which includes a person's personality traits, behaviors, interests, or other attributes[1] (e.g. Bless et al., 1996; Bodenhausen et al., 1999; Fiske

[1] Previous work has not always clearly defined the distinction between these two types of information, and where it is defined, definitions can vary, as well as the terms used to refer to this distinction (e.g. compare Bless et al., 1996; Bodenhausen et al., 1999; Fiske & Neuberg, 1990; Rubinstein et al., 2018; Skorich & Mavor, 2013). We therefore define this distinction in terms of these prototypical examples of each type. This aligns with how most empirical research in this area operationalizes this distinction (see Fiske et al., 1999; Fiske & Neuberg, 1990; Kunda & Thagard, 1996 for reviews). It also aligns with the two key theories considered here, the continuum model (Fiske & Neuberg, 1990) and Bodenhausen and colleagues' view (1999), which highlight how demographics, occupations and social roles are especially likely to serve as

& Neuberg, 1990; Gawronski et al., 2003; Kunda & Sherman-Williams, 1993; Kunda & Thagard, 1996; Rubinstein et al., 2018). Major theories in this area have argued that these two types of information can be used in very different ways (particularly the continuum model: Fiske et al., 1999; Fiske & Neuberg, 1990; and Bodenhausen et al., 1999). Specifically, these theories argue that categories tend to be easier to process and use, so that people may over-rely on categories relative to individuating information as a heuristic to simplify their inferences, especially when more effortful processing is limited. This view, if correct, has important real-world applications in terms of how to address more problematic cases of category use. For example, it suggests that the influence of race or gender on hiring decisions might be reduced when people are more motivated or able to think carefully about their decisions.

Yet, while theories of categories as heuristics are well-established, and based on a wealth of empirical evidence, this is not the only perspective available. It has been argued elsewhere that categories and individuating information might be processed and used similarly, so that categories will not necessarily be overused in the way these theories predict (Köpetz & Kruglanski, 2008; Kunda & Thagard, 1996). Existing research has been largely unable to arbitrate between these views, mainly due to the challenges in how to conceptualize category

---

organizing frameworks for impressions. (In the continuum model's terms, these types of information are more likely to serve as the 'category', and thus be overused in this way.) This definition also has benefits over alternative views which define social category information simply as information about a group of people, because this definition could include most information treated as individuating information (e.g. one could consider 'extraverts' or 'people who dislike spicy food' to be groups of people; Andrews, 2017; Bodenhausen et al., 1999) overuse, as well as potential alternative explanations for existing results. The current research therefore re-examines theories of categories as heuristics, focusing on two of these theories' key claims: 1) that, when using this heuristic, categories will be overused in a way that provides a less accurate inference strategy, and 2) that this will occur more in conditions that reduce effortful processing. By overcoming key theoretical and methodological challenges faced by previous research, the current studies provide some of the first clear and systematic tests of these claims.

Theories of Categories as Heuristics

Theories of categories as heuristics propose that categories and individuating information tend to be represented, processed, and used differently (Bodenhausen et al., 1999; Fiske et al., 1999; Fiske & Neuberg, 1990). Specifically, these theories suggest that categories tend to involve richer, more multifaceted schemas, thus make better frameworks for organizing other information. This is thought to increase the chance that people will organize their impressions around categories rather than individuating information when both types of information are available, a tendency which may be further increased by the broad social meaning attached to demographic, occupation, and role categories. When these impressions are then used to make inferences, the dominant organizing schema (usually a category) is thought come to mind more easily, so that incorporating additional (usually individuating) information requires more effort. Therefore, when people are putting little effort into their inferences, they may not fully incorporate the individuating information, and end up overusing the category. While such a strategy might simplify one's inferences, it should tend to come at the cost of reduced accuracy, because one has not fully considered all available information. In this way, overusing categories relative to individuating information is thought to serve as a heuristic – a simpler but potentially less accurate strategy for making inferences about others.[2]

Theories of categories as heuristics propose that, while people might rely on this heuristic in general, if their default is to use relatively effortless processing strategies, people should rely on it more in conditions that tend to reduce effortful processing. We focus on two such conditions that have been considered in these theories. The first is when there is good informational fit between the category and individuating information – i.e. when they are consistent with each other, or fit well together (e.g. an accountant who is good with numbers). Good fit is thought to indicate that the dominant organizing schema works relatively well to organize the impression, and may serve as a fairly accurate heuristic, thus reducing people' motivation to more carefully process the other available information. The second such condition is when people are under cognitive load – i.e. when people have limited time, attention, working memory, or cognitive resources – which should reduce their ability to use more effortful processing strategies. By reducing effortful processing, both of these conditions are proposed to increase the use of categories as heuristics, and thus lead people to make relatively more category-based judgments.

These theories therefore imply two key claims, which the current work will re-examine. Claim 1 is that, when people rely on categories as heuristics, categories will be overused relative

---

[2] Other explanations for how categories might serve as heuristics also exist. For example, initially learned categories may ease the encoding or recall of later learned information, especially category-consistent information (Macrae et al., 1994; Sherman et al., 2000). However, this mechanism leads to somewhat different predictions. For example, it does not necessarily imply that categories will be overused compared to individuating information. Instead, it suggests that category-consistent information might be overused compared to category-inconsistent or category-irrelevant information. Evaluating evidence for these other mechanisms is beyond the scope of the current work.

to individuating information in a way that provides a less accurate strategy for inferences (this is referred to here simply as category overuse, for short) – and that this should occur either in general, or when effortful processing is limited (e.g. when there is good information fit and under cognitive load). Claim 2 is that conditions which tend to reduce effortful processing (e.g. greater information fit and greater cognitive load) will lead people to use this heuristic more, and therefore make relatively more category-based judgments – i.e. increase their relative use of categories compared to individuating information, by either increasing their use of categories, and/or by decreasing their use of individuating information.

Limitations of Previous Tests of these Claims

Despite the prevalence of theories of categories as heuristics (Monroe et al., 2018), this is not the only perspective available on how these two types of information are used. Alternative theories propose that categories and individuating information are instead represented, processed and used similarly, and that any apparent differences between them are due to other factors which could equally affect either type of information (e.g. differences in how salient or how relevant the information is; Köpetz & Kruglanski, 2008; Kunda & Thagard, 1996). According to this alternative view, the two claims above will not necessarily hold: i.e. categories will not necessarily be systematically overused compared to individuating information, nor will this necessarily tend to occur more when effortful processing is reduced.

Yet despite the large amount of research on this topic (see Fiske et al., 1999; Fiske & Neuberg, 1990; Kunda & Thagard, 1996 for reviews), existing research is not able to arbitrate between these different views. For example, though initial work suggested that categories are represented and processed differently than individuating information, this work was based on a relatively limited range of items, and may not reflect differences between these two types of information more broadly (Andersen et al., 1990; Andersen & Klatzky, 1987; Bond & Brockett, 1987; Bond & Sedikides, 1988). Furthermore, key to the current research, evidence about how these two types of information are used for inferences – i.e. evidence for the claims that categories will be overused, and that this will occur more when effortful processing is limited – is also unclear. Specifically, theoretical challenges in conceptualizing category overuse, as well the potential for observed effects to be explained by other factors (e.g. differences in how relevant the information is, or how the information was presented), mean that most existing research provides little clear evidence as to whether categories are used as heuristics, or whether both types of information are instead used similarly.

Challenges in Previous Tests of whether Categories are Overused in a way that provides a Less Accurate Inference Strategy

In testing the first of these two claims – i.e. that, when using categories as heuristics, categories will be overused in a way that provides a less accurate inference strategy – the main challenge in previous work has been how to define and identify the relevant sense of overuse. Most previous research has not directly addressed this issue, as it instead focused on how much categories and individuating information are *used*, and without considering whether this should count as overuse, or produces less accurate inferences. In particular, research has often focused on testing a whether categories ‘dominate’ impressions, as phrased in the continuum model (Fiske & Neuberg, 1990). Though never defined explicitly, claims about category dominance are often taken to mean that people’s judgments will be affected more by changes in the category compared to changes in the individuating information (Fiske et al., 1999; Jussim et al., 2009; Köpetz & Kruglanski, 2008; Kunda & Thagard, 1996; Monroe et al., 2018). Yet a category can dominate impressions without necessarily being overused, or used in a way that provides a less accurate inference strategy. For example, suppose someone was hiring for an accounting job, and knew that an applicant currently worked as an accountant, and disliked spicy food. Here, the category (that the person is an accountant) is more relevant to the judgment so it *should* dominate the impression, and this will likely produce the most accurate inferences. Similarly, categories might be overused in a way that provides less accurate inference strategy, without necessarily dominating the impression. For example, suppose someone was hiring for the same accounting job, but instead knew the person's gender and the fact that they had gotten high marks in their accounting courses. Further suppose that both of these pieces of information were equally for judgments. Here, even though neither type of information dominates the impression, this could still count as category overuse, since the most accurate inferences would likely result from relying only on the individuating information (their marks) and not at all on the category (their gender). Therefore, tests of category dominance cannot answer whether categories are overused in ways that provide a less accurate strategy for inferences, as required for this to count as a heuristic.

To test whether this type of overuse occurs, we propose that people's information use needs to be compared to some standard of how that information *should* be used, if one was using the most accurate possible inference strategy. Furthermore, we propose that Bayesian modelling can be used to provide just such a standard. Bayesian models define what counts as a valid inference based on a set of beliefs – i.e. the inference that logically follows from those beliefs, and is most likely to be true, assuming those beliefs are true (Eva & Hartmann, 2018; Hahn, 2014; Leitgeb & Pettigrew, 2010; Rosenkrantz, 1992). These models do this by extending the rules of logic to also apply to probabilistic beliefs (i.e. beliefs of varying certainty, or beliefs about the probability that something is true).[3] This is essential for modelling most social inferences, which typically do not guarantee 100% certainty in one's inferences. To use Bayesian modelling to test for category overuse, one must develop a Bayesian model to define valid inferences based on different beliefs people might hold, and then measure people's actual beliefs, as well as their resulting inferences. People's actual information use can then be compared to how the model says that information should validly be used, given the person's beliefs. This allows for whether people overuse categories in a way that provides a sub-optimal inference strategy, i.e. one that deviates from valid inference given their beliefs.

Exactly this approach was taken by Locksley and colleagues (Locksley et al., 1980, 1982) and follow up work (Beckett & Park, 1995; Rasinski et al., 1985). Yet, despite the benefits of this approach, the few studies using it have so far been inconclusive, for several reasons. First, the results of this work are mixed, with different studies showing either underuse, overuse, or valid use of categories vs. individuating information. Second, potential confounds in many of these studies make it difficult to fully rule out alternative explanations. For example, category underuse in some cases was likely caused by presenting the category more subtly (as shown by Beckett & Park, 1995). Third, this work has examined only a fairly limited range of information content and processing conditions. Specifically, most of these studies only examined gender-assertiveness stereotypes (except Locksley et al., 1982, study 1), and in all studies, participants

---

[3] We use the term 'valid' as in logic, to refer to forms of inference where the conclusion logically follows from the premises. In standard formal logic, if the premises are true, and the inference is valid, then the conclusion is guaranteed to be true. By analogy, in the probabilistic cases considered here, true beliefs about the probabilities of something, used to make inferences in valid ways (as defined by the mathematical rules of probability), guarantees that the resulting inference will be a true belief about other probabilities, which also provides true beliefs about which outcomes is most likely to occur (Eva & Hartmann, 2018; Hahn, 2014).

could devote their full time and attention to the task, so may not have needed to rely on heuristics. This makes it unclear whether these results would generalize to other content or conditions.

To summarize, one key idea of theories of categories as heuristics is that relying on this heuristic will lead categories to overused in a way that provides a less accurate strategy for inferences (claim 1). However, most existing research cannot fully test this claim, because it does not clearly define valid inferences, nor test whether people overuse categories in a way that deviates from this. The few studies that do these things are inconclusive due to mixed results, possible alternative explanations, and using a limited range of content and processing conditions. Thus, it largely remains unclear whether people will tend to overuse categories in the ways proposed by theories of categories as heuristics, or whether they might instead use both categories and individuating information similarly – perhaps using both types of information validly, based on how relevant it is thought to be for the judgment.

Challenges in Previous Tests of whether People make Relatively More Category-based Judgments in Conditions that Reduce Effortful Processing

Theories of categories as heuristics not only imply that categories will be overused in a particular way, but also that this will occur more when people have a greater need to simplify their inferences. Specifically, these theories claim that conditions that tend to reduce effortful processing (e.g. informational fit and cognitive load) will increase the use of this heuristics, and thus lead to relatively more category-based judgments (claim 2). Though this claim has apparently widespread empirical support, potential alternative explanations for most existing results mean that there is actually little clear evidence as to whether or not this occurs.

Informational Fit. To understand why most research on informational fit has been unable to clearly test this claim, it is important to distinguish two different forms of informational fit, which are conceptually distinct and can vary at least somewhat independently. In the continuum model, these two forms are referred to as category-consistency, and judgment-consistency (Fiske & Neuberg, 1990). Category-consistency describes the extent to which individuating information is, *in general*, typical of the category (e.g. being good with numbers may be typical of accountants), rather than being unrelated to the category (category-irrelevant), or atypical of the category (category-inconsistent). Judgment-consistency describes how consistent the category and individuating information are *in terms of the specific trait being judged*. For continuous traits, we take this to mean that both pieces of information imply the person is above (/below) average on the to-be-inferred trait. For example, if inferring someone's extraversion from the fact that they are an accountant and dislike parties, both pieces of information would be judgment-consistent, as they both imply the person is below average extraversion. This contrasts to judgment-inconsistent cases, where one piece of information implies the person is above average and the other implies the person is below average (e.g. an accountant who enjoys parties), as well as to cases when one or both items are judgment-irrelevant (e.g. an accountant who dislikes spicy food). Research on both types of informational fit appears to show that greater fit leads to relatively more category-based judgments, in line with claim 2 (Bless et al., 2001; Chasteen, 2000; Craig & Bodenhausen, 2018; Fiske et al., 1987; Lambert, 1995; Lambert et al., 2004; Lord et al., 1984; Madon et al., 2006; Peters & Rothbart, 2000; Wilson, 2016; but see Monroe et al., 2018). However, these results are not as clear-cut as they may first appear.

One of the major challenges in interpreting this work is that most previous results could have been caused not by actual changes in information use, but instead an unrecognized type of statistical artifact. This artifact can occur whenever two conditions hold. First, the study must include varying levels of judgment-consistency. This likely occurred in many existing studies, which either varied judgment-consistency intentionally, or else may have varied it while trying to vary category-consistency. Importantly, when information is judgment-*inconsistent*, the traits implied by the category and individuating information should be anticorrelated (since, by definition, when one piece of information implies the person is above average on the trait, the other should imply they are below average). The second condition required for this artifact is that the trait implied by the individuating information (e.g. how extraverted people who enjoy parties are) is not statistically controlled for. This condition occurs in most studies, as the main focus of the research is often on category use, so that trait implications of the individuating information are not even measured. When these two conditions occur together, failing to control for this anti-correlated individuating information can lead to the false appearance that category use decreases, or even reverses its effect (as in Chasteen, 2000) when informational fit is lower. This is simply due to the un-accounted for effect of the individuating information, which pulls judgments in the opposite direction of the category when judgment-consistency is low. As the conditions to create this artifact were likely present in many existing studies, these studies cannot provide clear tests of the effect of informational fit.

Only a small number of studies exist without these conditions, all of which examined category-consistency (Fiske et al., 1987; Monroe et al., 2018; Peters & Rothbart, 2000). These few studies provide some, though mixed, support for the claim that informational fit leads to relatively more category-based judgments: two studies supported this claim (Craig &

Bodenhausen, 2018; Peters & Rothbart, 2000), one study may have supported this claim, though did not report the relevant statistical tests (Fiske et al., 1987), and one study failed to support it, showing instead that both types of information were used more with greater fit (Monroe et al., 2018).

Putting this work together, it is still unclear whether informational fit leads to relatively more category-based judgments, as implied by theories of categories as heuristics. In particular, there is no clear evidence for this claim in terms of judgment-consistency, due to potential statistical artifacts that may have driven existing tests of this, while there is some, though mixed, support for this claim in terms of category-consistency.

Cognitive Load. In terms of whether increasing cognitive load leads to relatively more category-based judgments, most published studies on this topic appear to support this claim, with 16 out of 18 studies supporting it in at least some conditions (see Supplementary Table 1). However, these results may also have other explanations. Specifically, the changes in information use may have been driven other factors that were confounded with information type, such as differences in presentation format. In particular, all published studies in this area appear to contain one or more of the following confounds: 1) always presenting the category first, 2) never presenting the category under load (i.e. only manipulating load while learning the individuating information), 3) presenting categories in simpler, shorter, or otherwise easier to process formats, compared to the judgment-relevant individuating information, or 4) choosing categories that are less socially or morally acceptable to use for the judgment, compared to the individuating information (see Supplementary Table 1). Furthermore, each of these confounding factors has been found to lead to cognitive load effects either on its own or in combination (Blair et al., 2004; Chun & Kruglanski, 2006; Skorich & Mavor, 2013). Therefore, it is still unclear whether cognitive load leads to relatively more category-based judgments based only on inherent differences in how easy-to-use these two types of information are (rather than these other factors), as proposed by theories of categories as heuristics.

The Current Research

To summarize, despite the wealth of research on the topic, it is still unclear how people combine categories and individuating information to make inferences about others. In particular, it is still unclear if people will overuse categories and make relatively more category-based judgments in the ways predicted by theories of categories as heuristics. Alternatively, both types of information might be used similarly, when more extraneous differences between these two types of information are removed (e.g. differences in presentation format). The current work therefore aims to overcome the theoretical and methodological challenges faced by previous work, to provide some of the first clear and systematic tests of these two claims.

The Current Approach to Testing for Category Overuse

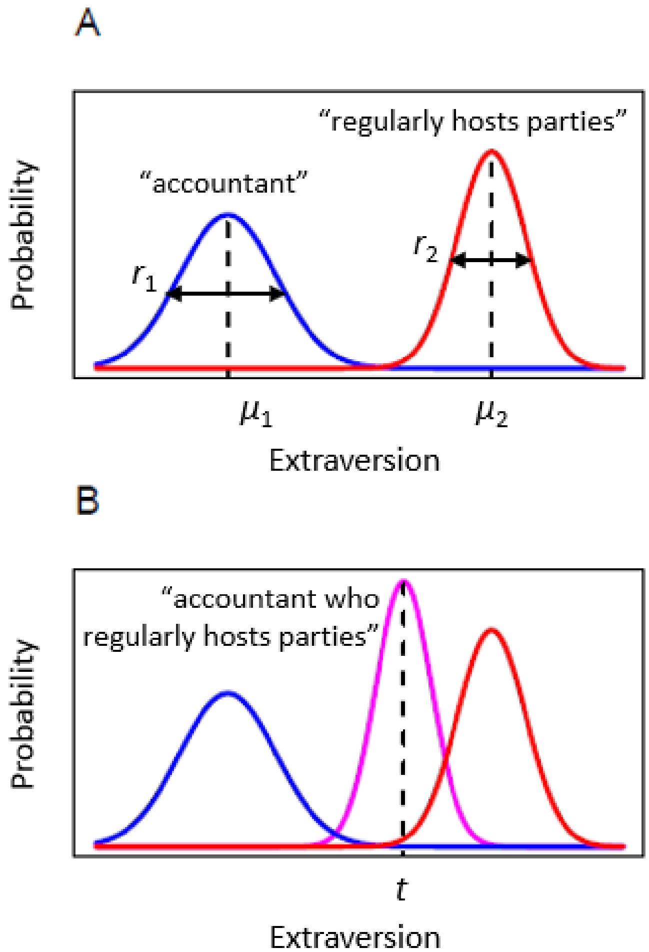
To more thoroughly test whether categories are overused in the ways predicted by categories as heuristics (as per claim 1), the current work used a Bayesian modelling approach, which built off a similar approach used in previous work (e.g. Locksley et al., 1982). As in this previous work, this approach provides a powerful way to test whether people are overusing categories, by using Bayesian modelling to precisely define what counts as valid information use, so that category overuse can be identified in contrast. Equally important, this approach provides guidance for how to conceptualize and measure the relevant aspects of the beliefs used to make these inferences. While previous work focused on inferences about binary traits, the current work extended this approach to apply to inferences about continuous traits, which may more naturally capture how people think about many inferences (e.g. inferences about extraversion, job competence). We therefore drew on models of perceptual cue integration (e.g. Angelaki et al., 2009; Oruç et al., 2003; Rohde et al., 2016) to develop a Bayesian model that applied to these continuous cases, and used this to define category overuse for these types of inferences, and to guide conceptualizing and measuring people's underlying beliefs. This model will be explained here for a simplified case where prior beliefs – i.e. beliefs about people in general – are uninformative. Full model details are provided in the online supplement.

A Bayesian Model of Valid Inferences. The current Bayesian model specifies how to make valid inferences about continuous traits, based on combining both category and individuating information about a person. An example inference might be inferring how extraverted someone is (say, on a scale from -50 to 50, with 0 being 'average') based on knowing that they are an accountant and regularly host parties. This model captures the intuition that, to make a valid inference, people should use each piece of information based on how informative (relevant, diagnostic, etc.) they think it is for the inference, relative to the other information available.

When this intuition is applied to continuous trait inferences, it is important to break down informativeness into two aspects (see Figure 1) The first aspect is referred to here as the *implied trait* of the information (μ, in Figure 1), which corresponds to the degree of the trait that someone with this characteristic is seen as most likely to have. For example, one might think that an accountant is most likely to be -30 extraverted (i.e. moderately introverted), and that someone who regularly hosts parties is most likely to be +30 extraverted (i.e. moderately extraverted). This aspect of informativeness has been considered most often in previous work (e.g. Krueger & Rothbart, 1988; Monroe et al., 2018), but it cannot fully capture informativeness when continuous traits are involved. For example, consider the difference between someone who got a B in one course, vs. got Bs in multiple courses. Even if these two people are both seen as 'average' intelligence, the second case could still be seen as more informative, and should have more influence on judgments. This difference is captured by the second aspect of informativeness: the *reliability* of the information (*r*, in Figure 1). This is reflected in one's certainty that a random person with that characteristic will be very close to the most likely degree of the trait (e.g. that a particular accountant will be exactly -30 extraversion), rather than having a wide range of plausible trait values. Intuitively, reliability should depend on factors like how similar people with that characteristic (e.g. accountants) are thought to be, and how much knowledge one has about their traits (Vrantsidis & Cunningham, in press). For the current example, we will assume the individuating information is seen as more reliable than the category, and so assign them assigned reliability values of 2 and 1, respectively. Note that, since the focus here is on defining valid inferences based on people's beliefs, the reliability and implied trait both refer to people's beliefs about the information's informativeness, which may or may not be accurate.

Figure 1

Bayesian model predictions



Note. Figures display posterior probability distributions, representing the inferred probability that someone will have each possible degree of extraversion, given either one or two pieces of information about them. Prior beliefs about people in general are assumed to be uninformative here. **(a)** Inferences based on single pieces of information. $\mu_1$and μ_2 correspond to the implied trait of the category and individuating information. *r*1 and r2 correspond to the reliability of the category and individuating information, with narrower distributions indicating greater reliability. **(b)** When combining two pieces of information, the Bayesian model predicts that the inferred trait (*t*) can be computed through reliability-weighted averaging: averaging the implied trait of each piece of information, with greater weight given to information that is relatively more reliable.

According to the current model, these two aspects of informativeness together determine how information should be combined to make a valid inference. Specifically, given certain assumptions (see online supplement for details), the model says that information should be combined through *reliability-weighted averaging*. This means that the resulting trait inference should be a weighted average of the implied trait of each piece of information, where the weight of each piece of information should depend on its relative reliability compared to the other information available (i.e. the reliability of that piece of information, as a proportion of the total reliability of all known information). Applying this the previous example of the accountant who regularly hosts parties, the inferred extraversion for this person would be +10 (i.e. $\text{weight}_{\text{category}}$ * $\text{implied trait}_{\text{category}}$ + $\text{weight}_{\text{individuating}}$ * $\text{implied trait}_{\text{individuating}}$ = $(\frac{1}{(1+2)}) * (-30) + (\frac{2}{(1+2)}) * (30)$). In this case, despite the two implied traits having equal and opposite magnitudes (-30 and +30 extraversion), the person is inferred to be above average extraversion (above 0). This is because the individuating information was seen as more reliable than the category, so the implied trait of the individuating information (+30 extraversion) had more influence on the judgment.

Defining Category Overuse in Contrast to Valid Inferences. By defining valid inference in this way, this model can then be used to define category overuse in contrast, thus providing a more precise, testable, version of claims about category overuse in these cases. Specifically, for cases where this model applies, claim 1 implies that using categories as heuristics will lead to overusing categories compared to individuating information, in a way that deviates from reliability-weighted averaging given people’s beliefs. This seems like a fair interpretation of theories of categories as heuristics, as, for example, the continuum model is explicitly framed as arguing that people often deviate from weighted averaging (Fiske et al., 1999; Fiske & Neuberg, 1990).

In order to test whether categories are overused in this way, we focus on specific cases where this overuse can be easily identified. In particular, we focus on cases where the model says that both types of information should be used equally: i.e. when the relative reliabilities (or weights) of the two types of information are the same. In these cases, to make a valid inference, the implied trait of the category and the individuating information should have equal impact on judgments. Furthermore, the same sized change in each implied trait should lead to the same sized change in judgments. This is analogous to a regression, where if two predictors have the same slope, the same sized change in either predictor should lead to the same sized change in the outcome. This is the key feature of the model that is used to identify category overuse in the current studies. Specifically, category overuse can be identified through cases where both types of information have the same relative reliability, but they do *not* have the same impact; instead, the category has greater impact. (That is, changes in the category's implied trait have a greater impact on judgments, compared to the same sized change in the individuating information's implied trait.) If observed, this would mean that, despite in effect reporting that both types of information should be used equally, people end up using the categories more. Observing this would provide support for the claim that people overuse categories relative to individuating information in a way that provides a less accurate strategy for inferences (claim 1), and thus provide support for the broader idea of categories as heuristics.

Conceptualizing and Measuring Perceived Informativeness. The current Bayesian model also provides guidance for how to conceptualize and measure the beliefs that go into people's inferences – specifically, beliefs about the perceived informativeness of the information

– which is essential for testing whether people overuse categories beyond what is justified by their beliefs. At a conceptual level, the model helps do this by distinguishing two key aspects of informativeness that must be considered whenever inferences about continuous traits are involved, i.e. the implied trait and the reliability of the information. Most previous work in this area has not clearly distinguished these two aspects, instead continuing to treat informativeness as a single construct even when continuous traits are involved (typically treating it as similar to the implied trait; e.g. Krueger & Rothbart, 1988). The main exception to this is previous work using non-Bayesian weighted-averaging models, which have been used to study other aspects of social inferences (Anderson, 1965, 2014). In line with the current model, this work conceptually separated the two aspects of informativeness considered here. However, it did not provide a way to directly measure these two aspects, instead computing them indirectly from the pattern of people's inferences across various combinations of information (Vidotto & Vincentini, 2007; Zalinski & Anderson, 1991). This indirect approach cannot be used in the current work, because it does not allow for measuring the beliefs underlying people's inferences separate from how information is combined to make those inferences. Separating these is crucial to test whether something in the process of combining categories and individuating information (e.g. using categories as heuristics) leads people to make inferences that do not follow from their underlying beliefs.

Building off the current Bayesian model, we therefore developed a more direct approach to measuring beliefs about these two aspects of informativeness. To measure the implied trait of each piece of information, we used a method common to previous research (e.g. Fiske et al., 1987; Krueger & Rothbart, 1988; Monroe et al., 2018) – asking people to make a trait rating based on only that piece of information (e.g. how extraverted are accountants?) To measure reliability, we drew on a key prediction unique to the Bayesian version of weighted-averaging models: that reliability should be related to people's *certainty* when making trait inferences based on only that piece of information. Certainty ratings were therefore used as a more direct measure of reliability (e.g. how certain one is that a random accountant would be exactly that extraverted).

For the main analyses presented in this paper, these two types of ratings – trait ratings and certainty ratings based on single items – were used as the primary measures of perceived informativeness. However, according to the current model, these single-item ratings should only provide pure measures of the implied trait and reliability in the case where prior beliefs (beliefs about people in general) are uninformative, and are expected to contain additional influences when prior beliefs are informative. To address this, the main statistical analyses were designed to be largely unaffected by these additional influences (see online supplement for details). A secondary analytic approach was also used, which involved directly measuring and accounting for the influence of prior beliefs. Results of this approach are mentioned briefly in the main text for comparison, and details are presented in the online supplement and on OSF.

Addressing Other Limitations of Previous Tests of Category Overuse. The core of the current approach to testing for category overuse addresses key challenges faced by most previous tests of this. Specifically, the current approach provides a way to conceptualize and identify category overuse, extending the Bayesian approach of Locksley and colleagues (1980, 1982) to apply to inferences about continuous traits. The current studies were also designed to overcome other limitations in the previous work using this approach. Specifically, the current studies presented both types of information in similar formats, to rule out differences in salience as a potential alternative explanation for differences in information use, and they used a wide range of information content and processing conditions (i.e. varying informational fit and cognitive load) to increase the generalizability of the results. This allowed for testing whether categories are in fact overused compared to individuating information, either in general or when effortful processing is limited, as proposed by theories of categories as heuristics, or if, instead, categories might be used similarly to individuating information (i.e. both used based on their relative reliability).

The Current Approach to Testing whether People make More Category-based Judgments in Conditions that Reduce Effortful Processing

The current studies also aimed to provide clearer tests of whether people make relatively more category-based judgments in conditions thought to reduce effortful processing (i.e. better informational fit and greater cognitive load), as proposed by theories of categories as heuristics. To do this, the current studies aimed to rule out alternative explanations for previous results. In terms of informational fit, the current studies were designed to avoid the statistical artifacts that may have driven previous results, by measuring and statistically controlling for the implied trait of both types of information. In terms of cognitive load, the current studies examined effects of load after removing the four confounds identified in previous work. (In addition, certain studies intentionally re-introduced some confounds to more closely replicate previous research.) Overall, by removing other factors that might have accounted for previous results, the current studies can better test whether these conditions in fact change the relative reliance on categories vs. individuating information, and whether this occurs due to inherent differences in how these two types of information are represented or processed. In contrast, if these two types of information are represented and processed similarly, then it might turn out that both types of information are affected similarly by these conditions, or perhaps not even affected at all.

Overview of Current Studies

Using the approach just outlined, a series of eight studies re-examined these two key claims about how categories and individuating information are combined to make inferences. This allowed for testing whether, in line with theories of categories as heuristics, people might overuse or increase their use of categories relative to individuating information in certain conditions, or whether people might instead use both types of information similarly. Two preliminary studies validated the current Bayesian model and measures of informativeness. Then studies 1 and 2 examined these two claims in conditions of varying informational fit, with study 1 focused on category-consistency, and study 2 focused on judgment-consistency. Studies 3 to 8 then examined these claims in conditions of varying cognitive load. Of these studies, studies 3-5 removed the confounds identified in previous research to more clearly identify differences between categories and individuating information, while studies 6-8 included some of these confounds to more closely replicate previous research.

Transparency and Openness

For all studies, we report how we determined sample sizes and all data exclusions. All manipulations and measures that are central to the main questions of this paper are reported in the main text, and additional (non-central) measures and manipulations are noted in either the main text or supplementary materials. The hypotheses, design, and analysis plans for study 7 and 8 were pre-registered. Other studies were not pre-registered.[4] Experimental materials, raw data, analysis code, additional results, and preregistrations are available online [link to be inserted

[4] Study 4 was also pre-registered, but several deviations from the preregistration were made due to changes in the research focus and analytic approach, so we treat this study as equivalent to a non-preregistered study (see online supplement for details).

here]. (Note that raw data are only able to be shared publicly for studies 3-8, due to restrictions in the research ethics protocol. Data for other studies are available upon request.) Data were analyzed using R, v. 3.6.0 (R Core Team, 2018), and the following packages: lmerTest v. 3.1-2 (Kuznetsova et al., 2017), ggplot v. 3.3.0 (Wickham, 2016), boot v. 1.3-22 (Canty & Ripley, 2019; Davison & Hinkley, 1997), emmeans v. 1.3.5 (Lenth, 2019), and metafor v. 2.1-0 (Viechtbauer, 2010).

Validating the Measures and Model: Preliminary Studies

Before testing these claims, two preliminary studies aimed to provide initial validation of the current measures of perceived informativeness, and the current Bayesian model of information integration (see online supplement for details) Specifically, these studies aimed to confirm that the current measures of the implied trait and reliability tracked factors expected to affect these measures. They also examined whether reliability-weighted averaging is a reasonable model of how people integrate information about others, by testing whether the implied trait of a piece of information has more influence when that information is seen as relatively more reliable. Observing reliability-weighted averaging using the current measures of informativeness would also further validate these measures.

To examine this, these studies used simple, highly controlled social inference tasks, where only one type of information was involved, and where heuristics were unlikely to be used. In each study, participants learned about the personality traits of five members of various artificial social groups (e.g. five values from 0-100, to represent the extraversion levels of five members of the red group). These groups could differ in the average and variability of the group members' traits. Participants then had to infer the same personality trait for a member of two of these groups (e.g. someone in the red and blue group), thus requiring them to combine two different pieces of information. The perceived informativeness of each piece of information was also measured by asking people to rate the same personality trait for a member of each single group (measuring the implied trait), and asking people how certain they were that this person would have exactly the degree of the trait just rated (measuring reliability). In the first study, participants made all responses while the information about the known group members was visible, while in the second study, all responses were made from memory, with participants also reporting the strength of their memories.

The results of these studies provided validation for both aspects of the current approach. The measures of informativeness tracked factors expected to affect them. Specifically, the measure of the implied trait tracked the average of the known group members' traits, and the measure of reliability tracked the (lack of) variability of the known group members' traits. In the second study, reliability also increased with the strength of people's memories. Results also confirmed that people combined multiple sources of information at least broadly in line with reliability-weighted averaging. Specifically, in both studies, the trait ratings for members of two groups were more strongly influenced by each group's implied trait when that group was seen as relatively more reliable, compared to the other group. Thus, these two studies provided validation for the current measurement methods and Bayesian model that form the basis of the subsequent studies.

Integrating Categories and Individuating Information under Varying Informational Fit: Studies 1 and 2

Studies 1 and 2 used this approach to test whether people integrate categories and individuating information in line with theories of categories as heuristics. Because good informational fit is thought to promote the use of heuristics (Craig & Bodenhausen, 2018; Fiske

& Neuberg, 1990), these two studies varied informational fit, and tested versions of claim 1 and 2 that apply in these conditions. Specifically, these studies examined whether people would overuse categories compared to individuating information, either in general, or especially when there is good informational fit (as per claim 1). These studies also examined whether better informational fit would increase the relative use of categories compared to individuating information (as per claim 2). Providing support for these claims would support theories of categories as heuristics, as well as the idea that categories and individuating information are represented, processed, and used differently. In contrast, failing to support these claims might indicate that these two types of information are represented, processed, and used similarly, or at least not used in the way proposed by theories of categories as heuristics. The two studies here each examined a different kind of informational fit, with study 1 focusing on category-consistency, and study 2 focusing on judgment-consistency. Though previous theories make similar predictions for both types of fit (Fiske & Neuberg, 1990), the limited evidence supporting these predictions, especially for judgment-consistency, and the fact that these two types of fit can potentially vary somewhat independently, makes it important to examine both types of fit separately.

Study 1: Varying Category-Consistency

Study 1 examined these claims in cases where people combined categories and individuating information with varying degrees of category-consistency. Participants made trait judgments about people described with two pieces of information: one category, and individuating. Pairs of items varied in how category-consistent, vs. category-irrelevant, they were, and participants rated how category-consistent they thought each pair of items was. To assess the perceived informativeness of each piece of information, participants made trait and certainty ratings about each piece of information on its own. Study 1 consisted of two sub-studies (study 1a and 1b) collected in different participant populations, and with slightly different stimuli. As these studies were largely the same, the methods and results of both studies are presented together.

Methods

Participants. Study 1a included 115 participants[5,6] from University of Toronto undergraduate participant pools who took part in the study in exchange for course credit (age: *M* = 18.73, *SD* = 1.35; gender: 76 female, 24 male, 15 not recorded; ethnicity: 5 Black/African-Canadian/African-American, 39 East Asian, 2 Middle Eastern, 12 South Asian, 31 White/Caucasian, 26 mixed/other/prefer not to answer/not recorded). Study 1b included 447 participants from the United States and Canada who were recruited online through Amazon Mechanical Turk (age: *M* = 38.95, *SD* = 12.49; gender: 278 female, 166 male, 3 other/prefer not to answer; ethnicity: 27 Black/African-Canadian/African-American, 18 East Asian, 12

[5] Sample sizes for all studies except study 7 and 8 were determined based on time and resource constraints, and independently of the data within each study. Power analyses were not used for these studies, since limitations in previous research made it unclear what effect sizes to expect, or if effects would exist at all in the current work. For study 1 and the two preliminary studies, which were collected as two sub-studies, this applied within each sub-study. However, some preliminary analyses were done on the first sub-studies before determining the sample sizes for the second. Thus, while the full sample sizes for the combined studies are not strictly independent of the full dataset, the focal analyses had not been performed yet, meaning that the full sample size should be largely independent of the main results, reducing concerns about optional stopping.

[6] Sample sizes for all studies are reported after excluding participants based on the following criteria. For all studies, participants were excluded if they did not complete the task. For studies run through Mechanical Turk, participants were excluded if they did not correctly pass all attention check questions. For studies run in person, participants were excluded if the experimenter noted that they should be excluded (e.g. due to obvious inattention). If these criteria meant that a participant would be included in the study more than once (i.e. if they completed the task twice), only the first set of responses for that participant was included.

Hispanic/Latin, 1 Indigenous/Native/Aboriginal/First Nations, 9 South Asian, 359 White/Caucasian, 21 mixed/other/prefer not to answer/no response).

Materials. Both sub-studies used short, one sentence descriptions of various people, where each description was designed to be judged on a particular trait (e.g. see Table 1). The main two-item descriptions involved one category and one piece of individuating information (e.g. "Someone is an aerospace engineer and can easily solve difficult physics problems."). To measure the informativeness of each item, single-item descriptions were created using each of the items from the two-item descriptions (e.g. "Someone is an aerospace engineer."). The categories used in the descriptions covered a broad range of demographic and occupation-based groups (e.g. Buddhist, aerospace engineer, female, teenager), and individuating information consisted of various behaviors, interests, and other attributes (e.g. always forgets where their keys are, likes to take naps, owns a private jet). Judgments covered a range of personality traits and other traits someone might have (e.g. intelligent, rich, good memory).

The key manipulation here involved varying category-consistency. To do this, two-item descriptions were created where the individuating information was category-consistent for half the descriptions (i.e. typical of the category it was paired with), and category-irrelevant for the other half of the descriptions (i.e. neither typical nor atypical of the category). As a secondary manipulation, both the category and individuating items also varied in whether they were relevant or irrelevant to the trait to be judged (most of the relevant items implied the person was above average on the trait). This manipulation was done to ensure sufficient independent variability in the perceived informativeness for the two types of information; effects of this manipulation were not directly of interest. This created a 2 (category-consistent/category-irrelevant) x 2 (category: judgment-relevant/judgment-irrelevant) x 2 (individuating information:

judgment-relevant/judgment-irrelevant) design. 32 two-item descriptions were created, so that there were four descriptions in each condition of the design. See Table 1 for example descriptions in each condition. Stimuli in each condition were informally pilot tested by research assistants in the lab. However, the accuracy of the condition assignment for each description was not essential, as participants' individual ratings of the implied trait, reliability, and category-consistency of the information were used for the main analyses.[7] The full list of descriptions and traits that were judged are listed on OSF, including slight changes between study 1a and 1b.

Table 1

Study 1 Example Descriptions (Judgment: How intelligent is this person?)

Individuating Information	Category	
	Judgment-Relevant	Judgment-Irrelevant
	Category-Consistent	
Judgment-Relevant	Someone is an aerospace engineer and can easily solve difficult physics	Someone is a Buddhist and can explain meanings of very difficult
Judgment-Irrelevant	Someone is an aerospace engineer and likes watching airplanes.	Someone is a Buddhist and just meditated for an hour.
	Category- Irrelevant	
Judgment-Relevant	Someone is an aerospace engineer and learned five languages on their own.	Someone is a Buddhist and just solved a very difficult math question.

[7] Secondary analyses also examined effects of the consistency manipulation, which was found to be successful: on average items in category-consistent pairs were rated as more similar than category-irrelevant pairs ($M_{consistent} - M_{irrelevant} = 22.04$; $p < .001$)

Judgment-Irrelevant	Someone is an aerospace engineer and just went grocery shopping.	Someone is Buddhist and just went grocery shopping.

Procedure.

Two-Item Ratings. In the first section of the study, participants read about 32 target people, each described with one of two-item descriptions, and rated each target on a single trait corresponding to that description.[8] Trait ratings were made on a 100-point scale: 0 = "Extremely below average", 50 = "Average", and 100 = "Extremely above average".

Single-Item Ratings. In a second section, participants rated the perceived informativeness of each piece of information. Specifically, they read about individuals described with each single-item description, and rated each individual on the same trait as in the corresponding two-item description. They also rated how certain they were that the individual would have exactly the level of the trait just rated (rather than having a very different level of the trait). Trait ratings were made on the same scale as for the two-item ratings. Certainty was also rated on a 100-point scale, where 0 = "Not at all", and 100 = "Extremely".[9]

Category-Consistency Ratings. A third section measured the perceived category-consistency of the two-item descriptions. Specifically, for each item pair, participants rated the similarity between someone described by the category and someone described by the

[8] All studies aside from studies 5, 7, and 8 also included ratings of people's certainty in these trait ratings based descriptions (equivalent to the single-item certainty ratings). These were not related to the main questions of this paper and therefore are not discussed further.

[9] Studies 1-8 also included ratings of the perceived informativeness of people's prior beliefs (i.e. similar ratings about a person one knows nothing about), which were used in the secondary analyses that accounted for prior beliefs.

individuating information (e.g. “How similar are someone who is an aerospace engineer and someone who can easily solve difficult physics problems?”). This was rated on a 100 point scale (0 = “Very dissimilar”, 50 = “No relationship”, 100 = “Very similar”).

Results

Unless otherwise specified, all analyses in this work were conducted using multilevel modelling with random intercepts for each participant.

Data Preparation. All trait ratings (based on either single items or two items) were centered at 0, so that possible responses ranged from -50 to 50.

Are Categories used More than Individuating Information when their Relative Reliability is the Same?

Analytic approach. With this task, the first goal was to examine whether categories are overused relative to individuating information, either in general, or specifically when information is category-consistent (in line with claim 1). Conceptually, this was examined by testing if categories were used more than individuating information when their relative reliability was the same. Specifically, this involved testing whether changes in the implied trait of the category have more impact on judgments, compared to the same sized changes in the implied trait of the individuating information. This was considered specifically for cases where both pieces of information have the same relative reliability, so that, if people are making valid inferences, they should use both types of information equally, while if they are overusing categories, the category should be used more than the individuating information.

These cases of equal relative reliability were examined in multiple ways. The first and more straightforward way is when both the category and individuating information known about a person have the same reliability, and therefore the same relative reliability. To test for category overuse in these cases, the trait rating of each two-item description was predicted from the single-item trait ratings for the corresponding category and individuating information, with both of these interacted with the relative certainty of the category compared to the individuating information (i.e. $(\text{certainty}_{\text{category}} / (\text{certainty}_{\text{category}} + \text{certainty}_{\text{individuating}})) - 0.5$). Relative certainty ranged from -0.5 to 0.5, with 0 indicating equal reliability, higher values indicating greater certainty for the category than individuating information, and lower values the reverse. A Wald's test compared the slope of the two trait ratings (category minus individuating), thus testing for differences in information use at equal reliability.

Category overuse was also examined when one piece of information was more reliable than the other, by comparing two different cases where the relative reliabilities were swapped. So, for example, if one case had a reliability of 3 for the category and 2 for the individuating information – i.e. the category had higher relative certainty – this could be compared to a second case with reliabilities of 2 and 3, respectively – i.e. where the individuating information had higher relative certainty – which also swaps the relative reliabilities ($\frac{3}{5}$ and $\frac{2}{5}$ vs. $\frac{2}{5}$ and $\frac{3}{5}$). If people make valid inferences, the influence of the category's implied trait in the first case should be equal to the influence of the individuating information's implied trait in the second case, as they both have equally high relative certainties (and vice versa, for equally low relative certainties). To test if this occurs, two regression models were fit like the one for equal certainty, but with relative category certainty re-centered at either +2 or -2 SD, so that the trait rating slopes reflect information use at this new 0 value. Then participant-stratified bootstrapping was used to compare the slopes of the trait ratings across the two models (category in one model minus individuating in the other), thus testing for differences in information use at equally high or equally low relative reliabilities.

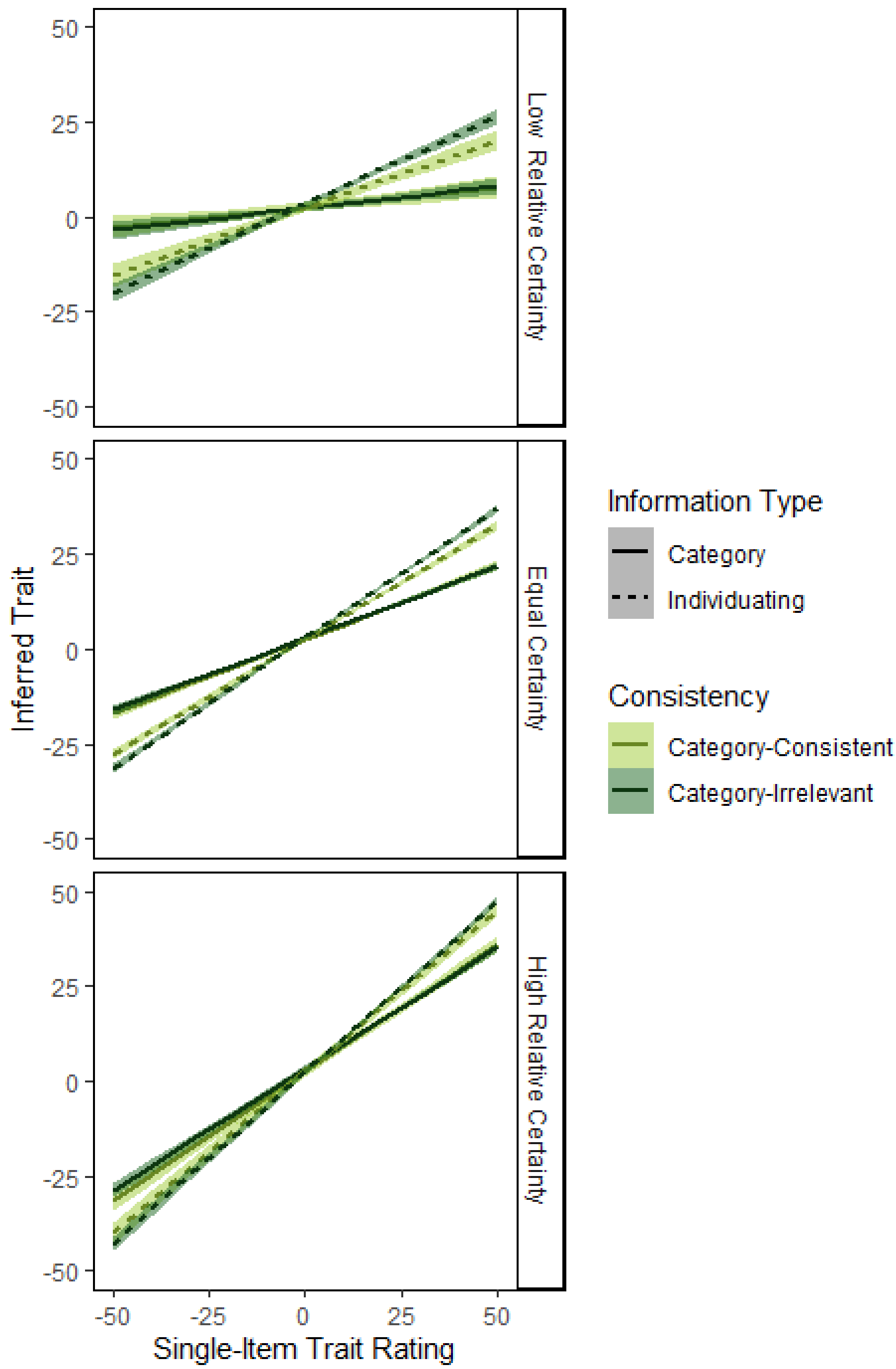
To make the effect size of these comparisons more interpretable, we computed what will be referred to as the judgment bias. This conceptually reflects the bias in trait judgments caused by a given amount of category overuse/underuse. Specifically, it computes how much two-item judgments would differ for a particular case, compared to if both items were used equally, on the 100-point trait rating scale used by participants. If single-item trait ratings for a study were on average, say, 25 points away from the scale midpoint, this judgment bias would be computed for a hypothetical case in which the category was rated at +25 and the individuating information was rated at -25. A judgment bias of, say, 5 (reported as bias(25) = 5) would indicate that the two-item judgment was 5 points more in the direction of the category than it would have been if both items were used equally. (Negative values would indicate bias in the direction of the individuating information.) Though there are no standard guidelines for interpreting these effect sizes, we refer to anything less than a 5 point bias (or 5 point change in bias, when testing claim 2) as a small effect. See supplementary materials for details on computing judgment bias.

Evaluating whether Categories are used More than Individuating Information when their Relative Reliability is the Same. This analytic approach was used to examine whether categories were overused (as per claim 1), either in general, or when there was good informational fit – i.e. whether, in these cases, categories were used more than individuating information of equal relative reliability. This was examined first in general (i.e. across all trials in the task). In general, all three slope comparisons showed that categories were *not* used more than individuating information: instead, individuating information was used more than categories at each value of relative certainty considered (See Figure 2; Equal Certainty: Slope Diff. = -0.28, 95% CI [-0.30, -0.26], $p < .001$, Bias(15) = -4.22, 95% CI [-4.51, -3.93]; High Relative Certainty: Slope Diff. = -0.24, 95% CI [-0.27, -0.21], Bias(15) = -3.60, 95% CI [-4.07, -3.11];

Low Relative Certainty: Slope Diff. = -0.32, 95% CI [-0.38, -0.27], Bias(15) = -4.85, 95% CI [-5.69, -4.02]). This held even when examining only cases that were seen as category-consistent (similarity ratings above 70, as in Fiske et al., 1987 and Monroe et al., 2018), going against the idea that categories are overused particularly when there is good informational fit (Equal Certainty: Slope Diff. = -0.22, 95% CI [-0.25, -0.18], $p < .001$, Bias(15) = -3.25, 95% CI [-3.74, -2.77]; High Relative Certainty: Slope Diff. = -0.19, 95% CI [-0.25, -0.14], Bias(15) = -2.92, 95% CI [-3.70, -2.15]; Low Relative Certainty: Slope Diff. = -0.24, 95% CI [-0.34, -0.14], Bias(15) = -3.58, 95% CI [-5.04, -2.15]). Analyses accounting for prior beliefs, as well as those defining consistency based on condition, produced similar results (see online supplement for details of analytic approach accounting for prior beliefs, and see OSF for results of all secondary analyses). These results therefore do not support claim 1, as categories were not used more than individuating information, either in general, or when information was category-consistent, thus going against the idea that categories are overused in the ways predicted by theories of categories as heuristics.

Figure 2

Individuating Information was used More than Category Information with the Same Relative Certainty, and Less So when it was Category-Consistent



Note. As indicated by the greater slope for trait ratings of individuating information compared to category information, individuating information was used more than categories of equivalent relative certainty (going against claim 1). This held for all cases of relative certainty considered, and for both category-consistent and category-irrelevant information. Category-consistency slightly decreased the use of individuating information compared to when information was category-irrelevant, leading to relatively more category-based judgments (consistent with claim 2). Error bands represent 95% confidence intervals. Category-consistency is plotted at the maximum and minimum values of category-consistency considered (+/- 1), and relative certainty is plotted at equal certainty (0 relative certainty), and high and low relative certainty (0 +/- 2 SD relative certainty; i.e. +/- .34).

Does the Relative use of Categories compared to Individuating Information Increase with Informational Fit? Though categories were not overused in any condition examined here, it is still possible that people made relatively more category-based judgments (i.e. increased the relative use of categories compared to individuating information) when there was greater (vs less) category-consistency, in line with claim 2. This was tested by taking the regression models used to test claim 1, and interacting all predictors with the degree of perceived category-consistency, operationalized as similarity ratings. Cases where items were considered category-inconsistent, i.e. had similarity ratings below 50, were excluded due to their small number (12% of responses) and to avoid these outlier values skewing results. The remaining similarity ratings were rescaled to range between -1 (category-irrelevant) and 1 (maximum category-consistency). This was used to test whether the degree of category-consistency significantly altered the difference in slopes between categories and individuating information – i.e. the relative use of categories compared to individuating information – for the three cases of relative certainty considered. Significance tests were done using Wald's test for equal certainty, and bootstrapping for high and low relative certainty.

Providing some support for the idea that category-consistency leads to relatively more category-based judgments, when certainty was equal, category-consistency led to a significant though small increase in the relative use of categories (Equal Certainty: Change in Slope Diff. = 0.09, 95% CI [0.03, 0.15], $p = 0.003$, Change in Bias(15) = 2.76, 95% CI [0.94, 4.58]; High Relative Certainty: Change in Slope Diff. = 0.09, 95% CI [-0.01, 0.18], Change in bias(15) = 1.28, 95% CI [-0.21, 2.75]; Low Relative Certainty: Change in Slope Diff. = 0.1, 95% CI [-0.08, 0.29], Change in bias(15) = 2.96, 95% CI [-2.47, 8.7]). A secondary analysis defining consistency based on experimental condition further supported this idea, with relative category use significantly increasing in all three cases considered (Equal Certainty: Change in Slope Diff. = 0.13, 95% CI [0.09, 0.17], $p < .001$, Change in Bias(15) = 3.91, 95% CI [2.77, 5.06]; High Relative Certainty: Change in Slope Diff. = 0.10, 95% CI [0.04, 0.17], Change in bias(15) = 1.54, 95% CI [0.60, 2.49]; Low Relative Certainty: Change in Slope Diff. = 0.16, 95% CI [0.04, 0.27], Change in bias(15) = 4.75, 95% CI [1.26, 8.16]). Accounting for prior beliefs produced similar results, which were significant when consistency was defined based on condition, but not self-reported similarity. Therefore, while these effects were small and not always significant, the current study provides some, albeit weak, support for the idea that category-consistency leads to relatively more category-based judgments.

Examining the regression coefficients from the main analysis suggests that this change was driven primarily by decreased use of individuating information when information was more category-consistent (Category: $b = 0.01$, 95% CI [-0.01, 0.02], $p = 0.51$; Indi.: $b = -0.04$, 95% CI [-0.06, -0.02], $p < .001$). There may also have been some increased use of categories, since, when consistency was coded based on experimental condition, there was also significant increases in category use (Category: $b = 0.04$, 95% CI [0.03, 0.05], $p < .001$; Indi.: $b = -0.02$, 95% CI [-0.03,

-0.01], $p < .001$). However, results were not entirely reliable: in analyses accounting for prior beliefs here, there were no significant effects when consistency was defined based on similarity ratings, and only the increase in category use was significant when defined based on condition.

Overall, these results provide some, though weak, support for the idea that category-consistency leads to relatively more category-based judgments (as per claim 2). Specifically, these results suggest that category-consistency may lead to slight increases in the relative reliance on categories compared to individuating information, potentially through both increasing the use of categories and decreasing the use of individuating information. However, this was only significant in some analyses and effects were overall quite small, so these results should be interpreted with caution.

Discussion

Study 1 examined two key claims that follow from theories of categories as heuristics, focusing on conditions of varying category-consistency. The results of this study provided evidence against claims about category overuse (i.e. against claim 1): categories were *not* used more than individuating information of the same relative reliability, either in general, or when information was category-consistent. Instead, individuating information was used more than categories in all cases examined. This provides evidence against theories of categories as heuristics. It also provides evidence against theories that categories and individuating information are used similarly, though it is possible that some other factor led to these differences, rather than inherent, generalizable, differences in how these types of information are processed and used. While claims about category overuse were clearly not supported, there was some, though weak, support for the claim that, in conditions of greater category-consistency, people would increase their relative use of categories compared to individuating information.

This result provides tentative support for at least this one aspect of theories of categories as heuristics.

Study 2: Varying Judgment-Consistency

As study 1 did not provide clear support for the predictions of theories of categories as heuristics, nor for alternative theories which propose these two types of information are used similarly, study 2 examined whether there might be clearer evidence found when varying a different form of informational fit: judgment-consistency. The fact that judgment-consistency is at least somewhat separable from category-consistency, as well as the potential statistical artifacts in previous studies that varied informational fit, meant that either theory, or perhaps an alternative theory, could be supported here. Study 2 therefore varied judgment-consistency by including both judgment-consistent and judgment-inconsistent pairs of information. It then examined these two key claims that follow from theories of categories as heuristics: i.e. whether categories would be used more than individuating information, either in general, or when information was judgment-consistent (as per claim 1), and whether judgment-consistency would increase the relative reliance on categories vs. individuating information (as per claim 2).

Methods

Participants. Study 2 included 494 participants who were recruited online through Amazon’s Mechanical Turk from the United States and Canada (age: M = 37.27, SD = 12.47; gender: 290 female, 203 male, 1 other/prefer not to answer; ethnicity: 43 Black/African-Canadian/African-American, 26 East Asian, 30 Hispanic/Latin, 1 Indigenous/Native/Aboriginal/First Nations, 7 South Asian, 364 White/Caucasian, 23 mixed/other/prefer not to answer).

Materials. The descriptions in study 2 were designed to vary in judgment-consistency, by including both judgment-consistent and judgment-inconsistent pairs of information. These were formed using information that could imply that someone was either above average or below average on the judged trait, so that pairs of information could be either judgment-consistent (both above/below average) or judgment-inconsistent (one above average, and one below). To investigate information use at different levels of relative reliability, each piece of information was also manipulated to imply either higher or lower subjective reliability while always being potentially judgment-relevant (e.g. by saying that the person regularly vs. recently did a behavior). This design meant that each two-item description fell into one cell of a 2 (category above or below average on trait) x 2 (individuating information above or below average on trait) x 2 (category high or low reliability) x 2 (individuating information high or low reliability) design. The stimuli and manipulations used here were modified from those used by (Krueger & Rothbart, 1988).

These descriptions were formed from two sets of items, one set related to the trait ‘aggressiveness’, and one set related to the trait ‘extraversion’ (see Table 2 for full set of items). From each set of items, two-item descriptions were generated by pairing a category and individuating item. For example, a two-item description related to aggressiveness might be: “Someone is a football player, and wants to continue this career. This person yelled at a cashier recently.” To measure the informativeness of each item, single-item descriptions were also created from each set by describing a person with only one piece of information (e.g. “Someone yelled at a cashier recently”). Stimuli in each condition were informally pilot tested by research assistants in the lab. As in study 1, the condition assignment of the stimuli was not essential, as participants’ ratings of the stimuli were used for all of the main analyses. Instead, these conditions were primarily used to generate descriptions that covered a range of implied traits, reliability, and category-consistency, and create independent variability in participant ratings of these factors.

Table 2

Study 2 Descriptions

Implied Trait	Category		Individuating Information	
	High reliability	Low reliability	High reliability	Low reliability
Trait: Aggressiveness				
Above average aggression	Is a football player, and wants to continue this career.	Is a football player, but wants a different career.	Often yells at cashiers.	Yelled at a cashier recently.
Below average aggression	Is a kindergarten teacher, and wants to continue this career.	Is a kindergarten teacher, but wants a different career.	Reacts calmly when criticized.	Recently reacted calmly when criticized
Trait: Extraversion				
Above average extraversion	Is an accountant, and wants to continue this career.	Is an accountant, but wants a different career.	Usually does not attend parties they are invited to.	Recently did not attend a party they were invited to.

Below average extraversion	Is a salesperson, and wants to continue this career.	Is a salesperson, but wants a different career.	Regularly hosts parties.	Recently hosted a party.

Procedure. Participants read about and rated two target people. Each target person was described with a two-item description, where one description was generated from each set of items. Each target was rated on the corresponding trait for that set (aggressiveness or extraversion). To measure the perceived informativeness of each item, participants rated people described with each of the single-item descriptions used in the two-item descriptions seen previously. In each case, participants rated the person on the corresponding trait for that item, and rated their certainty that the person would be exactly like this.

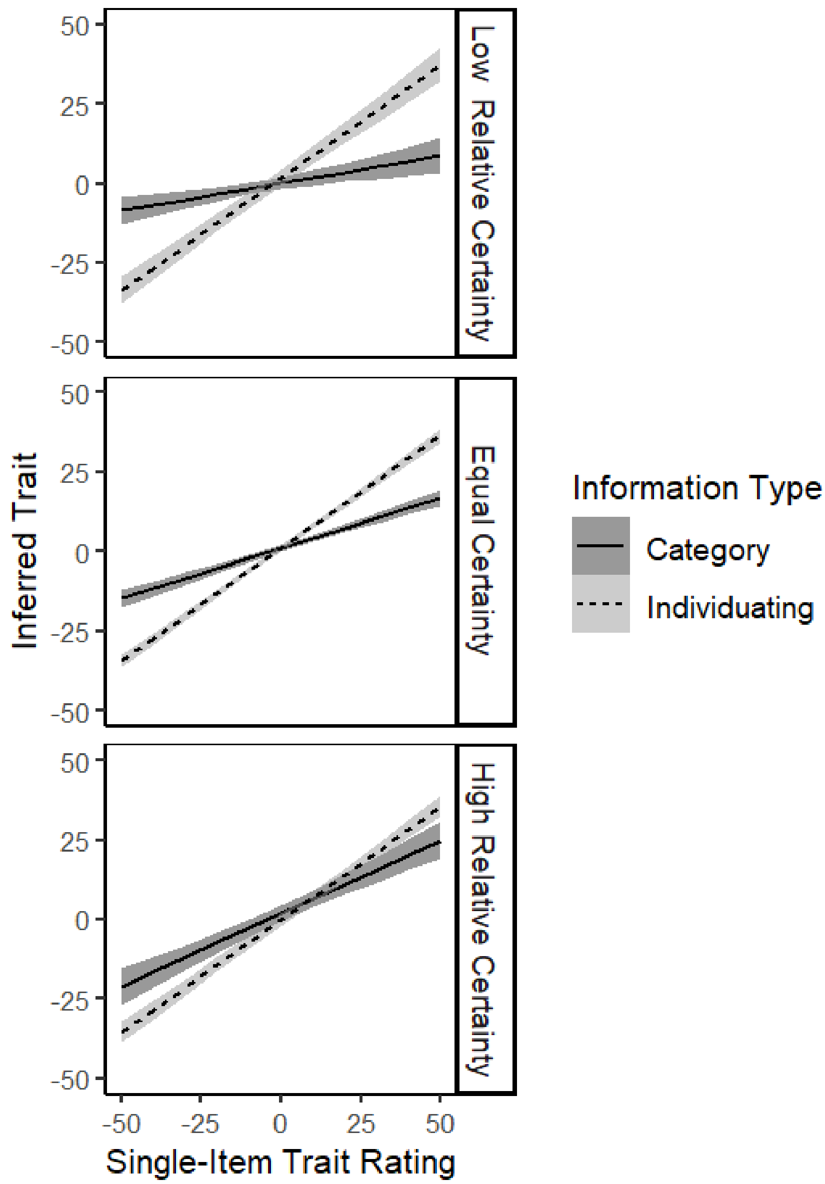
Results

Data Preparation. All trait ratings (based on either single items or two items) were centered at 0, so that possible responses ranged from -50 to 50.

Are Categories used More than Individuating Information when their Relative Reliability is the Same? As in study 1, evidence for category overuse was examined in general, and when there was good informational fit (as per claim 1). To do this the use of each type of information was again compared at three levels of equivalent relative certainty: when both items were equally certain (0 relative certainty), or at equivalently high or low relative certainties (relative certainty of +/- 2 SD: +/- .40 in this case). See study 1 results for full details of the analysis. Looking in general across all trials, results showed that, as in study 1, categories were not used more than individuating information. Instead, individuating information was used more than categories (see Figure 3; Equal Certainty: Slope Diff. = -0.39, 95% CI [-0.46, -0.32], p < .001, Bias(21) = -8.22, 95% CI [-9.62, -6.82]; High Relative Certainty: Slope Diff. = -0.25, 95% CI [-0.32, -0.11], Bias(21) = -5.22, 95% CI [-6.76, -2.38]; Low Relative Certainty: Slope Diff. = -0.53, 95% CI [-0.71, -0.47], Bias(21) = -11.23, 95% CI [-14.81, -9.97]). Furthermore, looking only at cases with good informational fit produced similar results: i.e. when only including trials that were seen as judgment-consistent, in that both items had trait ratings above or below the scale midpoint in the same direction (42% of trials; Equal Certainty: Slope Diff. = -0.45, 95% CI [-0.64, -0.26], p < .001, Bias(21) = -9.46, 95% CI [-13.44, -5.48]; High Relative Certainty: Slope Diff. = -0.23, 95% CI [-0.33, -0.08], Bias(21) = -4.85, 95% CI [-6.83, -1.73]; Low Relative Certainty: Slope Diff. = -0.67, 95% CI [-0.84, -0.58], Bias(21) = -14.07, 95% CI [-17.65, -12.27]). Versions of these two analyses that accounted for prior beliefs, and versions that defined consistency based on the experimental manipulations, also produced similar results. This provided clear evidence against claims about category overuse that stem from categories as heuristics (i.e. against claim 1), since categories were not used more, and instead were used less than individuating information of the same relative reliability.

Figure 3

Individuating Information was used more than Category Information with the Same Relative Certainty



Note. As indicated by the greater slope for trait ratings of individuating information compared to categories, individuating information was used more than categories of equivalent relative certainty when making judgments based on both types of information, going against claim 1. This held for all cases of relative certainty considered. Error bands represent 95% confidence intervals. Relative certainty is plotted at equal certainty (0 relative certainty), and high and low relative certainty (0 +/- 2 SD relative certainty).

Does the Relative use of Categories compared to Individuating Information Increase with Informational Fit? According to theories of categories as heuristics, conditions of greater informational fit should increase the use of categories as heuristics, thus increase the relative use of categories compared to individuating information (as per claim 2). This was tested as in study 1, but using judgment-consistency instead of category-consistency here. The degree of judgment-consistency was computed by multiplying the two items' trait ratings, after centering them at the scale midpoint. This value was then scaled by its standard deviation. Higher positive values here mean that the two items were rated more extremely in the same direction (both above or below average), so are more judgment-consistent; lower negative values mean the two items were rated more extremely in opposite directions, so are more judgment-inconsistent; zero values mean one or both items were rated as 'average' and therefore the items are neither consistent nor inconsistent. This was used to examine whether judgment-consistency altered the relative use of categories compared to individuating information, which was tested at the three levels of relative certainty considered previously.

Results showed that judgment-consistency did not significantly alter the use of categories relative to individuating information, at any degree of relative certainty examined (Equal Certainty: Change in Slope Diff. = 0.05, 95% CI [-0.09, 0.19], $p = 0.45$, Change in Bias(21) = 2.24, 95% CI [-3.61, 8.09]; High Relative Certainty: Change in Slope Diff. = -0.03, 95% CI [-0.19, 0.23], Change in bias(21) = -0.63, 95% CI [-4.01, 4.75]; Low Relative Certainty: Change in Slope Diff. = 0.14, 95% CI [-0.21, 0.24], Change in bias(21) = 5.74, 95% CI [-8.62, 10.09]). Secondary analyses defining consistency based on condition found no significant change at equal certainty, but significant increases in relative category use at high relative certainty, and significant decreases at low relative certainty. On the other hand, accounting for prior beliefs led to no significant effect in either version of the analysis. Therefore, overall, these results do not provide clear evidence that judgment-consistency alters relative information use.

The regression results were further examined to see if perhaps the use of both items was similarly affected by judgment-consistency, which would lead the relative information use to stay the same. The use of both types of information was found to decrease with greater judgment-consistency (Category: $b = -0.04$, 95% CI [-0.08, 0], $p = 0.034$; Indi.: $b = -0.07$, 95% CI [-0.11, -0.03], $p = 0.002$). However, this result was not robust across other versions of this analysis, which showed either no significant change, or significant increases in both types of information use.[10] These results therefore do not show any clear patterns of change in information use with varying judgment-consistency.

Altogether, there was little to no support for the idea that judgment-consistency would lead to relatively more category-based judgments; instead, it did not clearly lead to any changes in information use.

Discussion

Study 2 examined conditions of varying judgment-consistency to provide another test of the two key claims that stem from theories of categories as heuristics. This study again found that categories were *not* overused – i.e. not used more than individuating information of the same relative certainty – either in general or when information was judgment-consistent, thus going against claim 1. Instead, like in study 1, categories were used less than individuating information

---

[10] When consistency was defined based on experimental condition, no significant change was found for either type of information use. When accounting for prior beliefs, the opposite results were found from those reported in the main text: the use of both types of information increased (rather than decreased) with greater judgment-consistency, and this held for both ways of defining consistency.

in all cases examined. Study 2 also found that judgment-consistency had no clear effect on how categories and individuating information were used, contrary to claim 2, which proposes that greater informational fit should reduce effortful processing, and thus lead to relatively more category-based judgments.[11]

Putting study 1 and 2 together, so far these results provide little support for theories of categories as heuristics, and their predictions about how categories and individuating information will be combined to make inferences. Across varying degrees and types of informational fit, there was no evidence that categories were overused compared to individuating information (instead, categories were underused in both studies), and there was little to no evidence that greater informational fit would lead to relatively more category-based judgments. These results could imply that categories are not any easier to process and use than individuating information, and thus not used as heuristics. On the other hand, people may simply have not relied on heuristics in the conditions examined here (e.g. perhaps even with good informational fit, people had sufficient motivation and ability to use more effortful processing strategies).

The results so far also do clearly support alternative theories proposing that categories and individuating information are used similarly, since, in both studies, the individuating information was overused compared to the category. If this pattern of results holds more generally, it could indicate that categories and individuating information are in fact processed or represented in reliably different ways – just not the ways predicted by theories of categories as heuristics. On the other hand, if this result does not hold more generally, it could indicate that

[11] This lack of effect was replicated in data from studies 4 and 5, where judgment-consistency again had no effect on relative information use (see supplementary materials).

some other factor led individuating information to be overused in these studies. (Other factors which may explain these results are discussed in the general discussion).

Integrating Categories and Individuating Information under Cognitive Load: Studies 3-8

As mentioned, one explanation for the lack of support for theories of categories as heuristics is that people in the previous studies may not have needed to rely on heuristics in any of the conditions examined, if they had sufficient ability and motivation to use more effortful processing strategies. Therefore, the next set of studies aimed to increase the chance that people would rely on categories as heuristics, put putting them under cognitive load – i.e. limiting the amount of attention and working memory they can devote to their impressions and inferences, which should limit their ability to use more effortful processing strategies. Specifically, studies 3-8 manipulated whether or not people were under cognitive load while forming their impressions, and tested predictions of categories as heuristics that apply in these conditions.

One goal of these studies was to test whether, under cognitive load, categories would be overused compared to individuating information (as per claim 1), in line with the idea that category overuse serves as a heuristic. These studies provide the first tests of this claim, since previous studies involving cognitive load did not test whether people deviated from valid information use. These studies could also provide further tests of whether categories would be overused in normal processing conditions (as in studies 1 and 2), by testing whether categories would be overused in the no load condition.

A second goal of these studies was to test whether cognitive load would lead to relatively more category-based judgments (i.e. increase the use of categories and/or decrease the use of individuating information) compared to no load conditions, as per claim 2. All previous tests of this question confounded information type with at least one of four other factors (i.e. presenting categories first, never under load, more concisely, or using socially unacceptable categories). Therefore the first three studies, studies 3-5, examined effects of load after removing these confounding factors.

To preview the results, studies 3-5 did not find support for either prediction of theories of categories as heuristics. Therefore, studies 6-8 aimed to use methods more similar to previous studies which found cognitive load effects on relative category use, to provide the greatest chance that categories would be used as heuristics. To do this, studies 6-8 included several confounds that may have been at least partly (if not fully) responsible for these load effects in previous work (i.e. presenting categories first, never under load, and, at times, more concisely). Studies 7 and 8 also attempted to directly replicate a previous study in this area (Pendry, 1998), in case some other aspect of these methods was required for categories to be used as heuristics. While results in line with claims 1 or 2 in these studies would only provide weak support for theories of categories as heuristics, since they could be attributed to the other confounding factors, results *not* in line with these claims would provide even stronger evidence against these theories, as these studies should provide the greatest chance for categories to be used as heuristics.

In order to convey the overall results across these 6 studies, key features of each study's methods are described below, followed by a meta-analysis of all results.

Methods

Overview

Several core aspects of the methods were shared across all 6 studies. Specifically, as in studies 1 and 2, the main impression formation task involved reading about a target who was described with both category and individuating information, and then rating the target on a trait, as well as rating the informativeness of the information. Extending studies 1 and 2, cognitive load was manipulated during the impression formation task: targets could either be learned about while participants were distracted by a secondary task (load condition), or without any distracting secondary task (no load condition). The no load condition was akin to the normal processing conditions used in studies 1-2. The current studies also attempted to make the impression formation task more difficult than in studies 1 and 2, to further increase the chance that people would use simplifying heuristics. This was done by making the target descriptions longer (using either 8 or 21 items per description), and only telling participants the traits to be judged once the target descriptions were no longer visible (so that participants could not simplify the task by attending to only judgment-relevant information).

While these core aspects of the methods were shared across all 6 studies, other aspects differed across studies (described below; see Table 3 for summary). The methods of studies 3 and 7 are described in detail. As the other studies contained only minor changes, they are described more briefly. Additional methods details are listed in the supplementary materials, including several exploratory aspects of the designs that were not relevant to the main questions or analyses, and full materials are available on OSF.

Table 3

Comparison of Methods for Studies 3-8

Study	Confounds with information type?	Direct replication of previous research?	Load task	Number and Judgment-Relevance of Category and Individuating Items	Judgment-Consistency of Relevant Items	Study Format & Participants
3			Verbal working memory (8-digit number)	1 category, 1 relevant indi., 6 filler indi.	Inconsistent	In person Undergrads N = 117
4			Verbal working memory (8-digit number)	1 category, 1 relevant indi., 6 filler indi.	Inconsistent/ Consistent	In person Undergrads N = 124
5			Visual working memory	1 category, 1 relevant indi.,	Inconsistent/ Consistent	Online MTurk N = 192

				6 filler indi.		
6	Category first, never under load		Visual working memory	1 category, 1 relevant indi., 6 filler indi.	Inconsistent	Online MTurk N = 94
7	Category first, never under load, more concise	Yes, Pendry (1998)	Distracting audio	1 category, 20 potentially relevant indi.	Varied	Online MTurk N = 101
8	Category first, never under load, more concise	Yes, Pendry (1998)	Distracting audio	1 category 20 potentially relevant indi.	Varied	Online Undergrads N = 174

Removing Confounds with Information Type: Studies 3-5

Studies 3-5 aimed to remove other factors often confounded with information type in previous research, so that any differences in the use of categories and individuating information could be attributed more directly to inherent differences in how these two types of information are processed or represented. To do this, first, all information was presented in a randomized order (rather than the category always first). Second, in the load condition, all information was presented while under load (rather than only the individuating information). Third, both the category and the judgment-relevant individuating information were presented similarly, by using a single, short statement for each (rather than presenting judgment-relevant individuating information in a longer, more complex, or more detailed format). Fourth, all information was designed to be socially acceptable to use for the judgments being made (rather than the category being socially unacceptable).

Study 3.

Participants. Study 3 included 117 participants from University of Toronto undergraduate participant pools who participated for course credit (age: M = 19.30, SD = 1.56; gender: 67 female, 47 male, 1 other/prefer not to answer; ethnicity: 3 Black/African-Canadian/African-American, 63 East Asian, 1 Hispanic/Latin, 4 Middle Eastern, 21 South Asian, 19 White/Caucasian, 6 mixed/other).

Materials. Descriptions of eight target people were used for the main impression formation task, with description corresponding to a different trait that the target would be judged on. See Table 4 for an example. Each description contained 8 items: 1 category (in most cases, occupations) and 7 pieces of individuating information (traits, behaviors, hobbies, etc.). The category and one piece of individuating information were designed to be relevant to the judged trait, but to have judgment-inconsistent implications (e.g. one implying the person was extraverted, the other, introverted). The rest of the items were used as fillers, and were designed to be irrelevant to the judged trait, but potentially relevant to an overall impression of the person. Pretesting was used to select relevant and irrelevant items.

Table 4

Study 3 Example Description (Judgment: How Talkative is this Person?)

Item (“Someone who…”)	Information Type
is a bartender	Category
doesn't socialize much	Judgment-Relevant Individuating
sometimes goes to sleep late	Filler Individuating
likes to take long walks	Filler Individuating
does watercolour painting	Filler Individuating
bikes to work	Filler Individuating
only eats organic food	Filler Individuating
likes trying new recipes	Filler Individuating

Procedure.

Impression Formation Task. For each target person, participants were asked to form an impression of someone described with the eight-item description. All items were presented on screen at once, in a random order. Participants could read the description for as long as they wanted for a minimum of four seconds, then advance to the next screen, where they rated this target on the corresponding trait.

Load Manipulation. Load was manipulated within-participants using a number recall task that has been used successfully in several previous studies in this area (Blessum et al., 1998; Gilbert & Hixon, 1991; Macrae et al., 1993; Pendry & Macrae, 1994; Wigboldus et al., 2004). For targets in the load condition, participants were given an eight-digit number to remember before learning about each target. After making their trait judgment for that target, participants were asked to type the number into the computer, and were told they could forget the number. For targets in the no load condition, participants were not asked to remember any numbers.

Single-Item Ratings. After rating all eight of the main target people, participants then made trait ratings and certainty ratings about people described with each of the single items used in the original descriptions, with ratings made about the same trait as in the original description.

Study 4. Study 4 was largely similar to study 3. Changes involved including judgment-consistent descriptions in addition to inconsistent ones, asking participants to type the number in the load condition before, rather than after, making their trait judgments, and including a free recall task where participants listed the items they recalled from the target descriptions.

Study 5. Study 5 differed from study 4 in two main ways. First, the impression formation task was altered to control how long participants had to view each piece of information, by presenting each item separately for a fixed amount of time. Second, the load manipulation was modified to use a visual working memory task (Bethell-Fox & Shepard, 1988; Miyake et al., 2001), where participants had to remember a pattern of 5 dots on a 4x4 grid (equivalent to the 'very high load' condition in Trémolière et al., 2012). This visual working memory task has been used successfully to manipulate cognitive load times in a variety of domains (Neys, 2006; Trémolière et al., 2012). This manipulation was used to prevent participants from cheating, as this study was done online, where participants could more easily cheat by writing down the numbers in the number recall task used previously.

Increasing Similarity with Previous Research: Studies 6-8

Studies 6-8 aimed to extend these studies by using methods more similar to previous research, to try to increase the chance that people would rely on categories as heuristics (assuming they actually relied on categories as heuristics in these previous studies). Therefore, studies 6-8 included several factors that have been frequently confounded with information type in previous work, and which could have worked in combination with more inherent differences between categories and individuating information in causing relatively more category-based judgments under load (see Table 3). Studies 7 and 8 took this one step further, by attempting to directly replicate all aspects of a previous study in this area (Pendry, 1998), in case some of these other aspects of the design were required for categories to be used as heuristics.

Study 6. Study 6 was similar to study 5, but added in two confounding factors: the category was always presented first, and only the individuating information was presented under load in the load condition, never the category. As in study 3, this study used only judgment-inconsistent descriptions.

Study 7. Study 7 attempted to replicate a previous study which found that cognitive load led to relatively more category-based judgments (Pendry, 1998). This study was chosen to replicate because these cognitive load effects were observed in three studies using very similar methods (Pendry, 1998; Pendry & Macrae, 1994, study 1 and 2), as well as several other work using related methods (Macrae et al., 1993). Like study 6, this study included several confounding factors which could have led to these load effects. As in study 6, the category was presented first, and never under cognitive load. In addition, the category was presented in a potentially simpler, more concise format compared to the judgment-relevant individuating information (i.e. the category was always presented as a single statement, while the relevant individuating information could be spread out over multiple statements). While the current study attempted to reproduce the methods of (Pendry, 1998) as closely as possible, small modifications were made due to having to generate new stimuli and due to the constraints of an online testing environment; these differences are noted below. A preregistration for this study specified sample size, exclusion criteria, hypotheses and analysis plans.[12]

Participants. 101 participants were included. Participants were recruited online from the United States through Amazon Mechanical Turk and the Cloud Research platform (age: *M* = 41.93, *SD* = 12.55; gender: 42 female, 57 male, 2 other/prefer not to answer; ethnicity: 10 Black/African-Canadian/African-American, 4 East Asian, 3 Hispanic/Latin, 2 South Asian, 74 White/Caucasian, 7 mixed). Due to a technical error, one extra participant completed the task beyond the preregistered sample size of 100; this person was included in the dataset. The current sample was estimated to provide 96% power to find an effect of the same size as Pendry's study (see online supplement for details), and this sample was over 4 times larger than that of Pendry's original (1998) study (n=24).

Materials. The main impression formation task involved learning about a single target: Hilda, a 75-year-old woman. The target's age and gender served as the category information, and a description of the target contained 20 pieces of individuating information (Pendry used 24 pieces). The exact individuating items and the traits to be rated were generated through pre-testing in the same way as in Pendry's study. Nine of the selected items were category-consistent (average ratings of 7 or above, on a 9 point scale for how typical of 75-year-old women it was) and nine items were category-inconsistent (average ratings of 3 or below; Pendry & Macrae (1994) used a cutoff of 2). An additional two category-irrelevant items were added (average

[12] In studies 7 and 8, the only minor deviation from the analysis plan was in the value of relative certainty used for comparing information use at high and low relative certainty, which was changed to ensure consistency across studies in the meta-analysis. (The equivalent change was made in the weight used in the analysis accounting for the priors, again to ensure consistency across studies.) Only the analyses performed on the individual studies were pre-registered, not the use of a meta-analysis across studies.

typicality ratings between 5 and 6) for exploratory purposes. Examples of the individuating information included: “Hilda likes knitting” (category-consistent), “Hilda dyed her hair purple” (category-inconsistent), and “Hilda likes watercolour painting” (category-irrelevant). Seven traits which Hilda was to be rated on were also selected. These were either typical of elderly women (“forgetful”, “loving to her family”) or atypical (“physically fit”, “good with technology”, “promiscuous”); two category-irrelevant items were also included for exploratory purposes (“artistic”, “introspective”). One or more pieces of individuating information could potentially be seen as relevant to each trait.

Procedure.

Impression Formation Task. Participants were told that they would be asked to form an impression of someone, and that they would need to justify their impression of this person later. (The online nature of this task made it implausible to say that they would have to justify their impression ‘to a member of the psychology department’, as in Pendry, 1998.) Participants were then told they would be learning about Hilda, a 75-year-old woman, and proceeded to learn the 20 pieces of individuating information about her, in random order. Each piece of individuating information was presented for a fixed time of 5 seconds. Though Pendry’s (1998) task was self-paced, fixed timing was used here so that the audio in the load condition could be matched to the length of the impression formation task. The timing of 5 seconds per statement was chosen to approximate the average reading times in Pendry and Macrae (1994), adjusted for the average number of characters per statement in the current study.

Load Manipulation. Load was manipulated between-participant by giving half of the participants a distracting audio to listen to while learning the individuating information (load condition), while the other half of participants did not listen to any audio (no load condition).

The audio consisted of 12 trivia facts about chocolate. Participants in the load condition were told they would be tested on the information in the audio later, and to attend to both the audio and the impression formation task, as both were equally important. The audio played throughout the time that participants saw the individuating information. This manipulation differed slightly from Pendry's study, where all participants overheard a conversation that varied in how personally relevant the topic was. The current version was better suited to an online testing environment, and was expected to produce a larger difference between conditions due to having no audio in the no load condition.

Filler Task. After learning about Hilda, all participants answered 12 multiple choice trivia questions based on the facts presented in the load condition audio, along with 6 easy general knowledge questions. This was designed to take about 3 minutes. (Pendry's study used a different filler task which took about 5 minutes.)

Target Ratings. Participants then rated Hilda on the 7 personality traits. For the last trait, they were asked to justify their answer with a short statement.

Category Recall. Participants were asked how old Hilda was, to ensure they remembered the category information.

Item Ratings. For each trait that was rated, participants saw a list of all 20 individuating items and selected those that they thought were relevant to that trait. Then, for each trait, participants made trait ratings and certainty ratings about: 1) a person described with only the category (i.e. a 75 year old woman), 2) a person described by all the individuating information that had been selected as relevant to that trait (this was skipped if nothing was selected as relevant), and 3) a person they know nothing about.

Manipulation Checks. At the end of the task, participants rated how much attention they paid to the impression formation task, and how difficult it was to form their impression of Hilda.

Study 8. Study 8 again attempted to replicate Pendry (1998), with a few minor changes from study 7. Study 8 used a larger sample size than study 7 (n = 174) to further increase power. As per the preregistration, this sample was over the minimum of 100 participants (determined using the same power analysis as in study 9), and the decision to stop data collection was made independent of the study results. Three changes were also made to the study design. First, in the impression formation task, the reading time for each piece of individuating information was shortened to 3.5 seconds, which could make the task more difficult overall, and therefore increase the use of heuristics. (This time is close to the shortest reading times found in any condition of Pendry and Macrae's (1994) study, when not adjusted for the number of characters in the statements). Second, the category information (the fact that Hilda was a 75 year old woman) was presented earlier in the task, mixed in with the initial instructions, rather than directly before learning the individuating information. It is unclear when this information was presented in Pendry's original study, but presenting the category earlier may have allowed participants additional time to process it, therefore potentially increasing category use. Third, a free recall task was added, where participants listed as many of the individuating items as they could recall. As in Study 7, sample size (minimum sample size and stopping rules), exclusion criteria, hypotheses and analysis plans were preregistered.

Results

Exclusions

In studies 5 and 6, trials were excluded if four or more of the digits were recalled incorrectly in the number recall task, as participants may not have been sufficiently attentive to the number to increase cognitive load. This cutoff was established by Gilbert and Hixon (1991) and has been used by previous studies using this load manipulation (e.g. Macrae et al., 1993; Pendry & Macrae, 1994). This led to 91% of load trials being included in both study 5 and study 6. Results did not differ when including all trials.

Additional exclusion criteria were used in study 7 and 8 to ensure that participants were paying sufficient attention to the category information, to the distracting task in the load condition, and to the study in general. Specifically, participants were excluded if they did not list an age of 60 or over when asked to recall how old Hilda was, if they were in the load condition and did not get at least half of the questions correct that were based on the audio that they listened to, or if they did not answer at least 5/6 of the easy general knowledge questions correctly. Excluded participants were not analyzed, as per the preregistration.

Data Preparation

All trait ratings were centered at 0, so that possible responses ranged from -50 to 50.

Manipulation Checks

Manipulation checks in studies 7 and 8 supported the effectiveness of the load manipulation in these studies. In both cases, participants in the load condition reported attending less to the impression formation task (Study 7: b = -0.21, 95% CI [-0.33, -0.09], $p < .001$; Study 8: b = -0.30, 95% CI [-0.41, -0.19], $p < .001$), and reported that the impression formation task was more difficult (though this was only trending for study 8; Study 7: b = 0.18, 95% CI [0.04, 0.32], $p = 0.01$; Study 8: b = 0.10, 95% CI [-0.01, 0.22], $p = 0.08$).

Though the studies 3-6 did not include clear manipulation checks, there are several reasons to expect that cognitive load was successfully manipulated. For the eight-number recall task in studies 3 and 4, participants did in fact remember the number quite well, indicating they were engaged in this task (mean digits correct in included trials: study 3: 7.28; study 4: 7.73; similar to performance in previous studies; Gilbert & Hixon, 1991; Pendry & Macrae, 1994), while still frequently making errors (percent of included trials with at least one error: study 3: 29%; study 4: 25%), suggesting that people's working memory resources were overloaded (Kantowitz, 1974). Furthermore, a large number of previous studies have shown that this task depletes working memory resources, and can lead to relatively more category-based judgments using similar study designs, though perhaps driven by other factors than information type, as discussed in the introduction (Blessum et al., 1998; Gilbert & Hixon, 1991, 1991; Macrae et al., 1993; Pendry & Macrae, 1994; Wigboldus et al., 2004). Similarly, for the visual working memory task used in studies 5 and 6, participants showed decent recall while also making frequent errors (average number of correctly recalled dots out of 5: study 5: 3.52, study 6: 3.86; both significantly above chance, ps < .001), again suggesting that participants' working memory was engaged but overloaded. Furthermore, this performance level is similar to previous research where this manipulation was found to effectively interfere with processing (Trémolière et al., 2012).

Additional, though more indirect, checks of these manipulations can come from the free recall tasks included in studies 4, 5, 6 and 8, where participants were asked to recall as many of the items in the target descriptions as they could remember. If fewer items were recalled in the load condition than the no load condition, this would provide additional evidence that load interfered with cognitive processing. This was observed in study 8 (b = -0.23, p < .001), providing additional support for the effectiveness of this manipulation. However, there was no significant difference in studies 4-6 (Study 4: b = -0.16, p = 0.187; Study 5: b = 0.05, p = 0.156; Study 6: b = -0.07, p = 0.579; Study 8: b = -0.23, p < .001). Yet this does not mean that these other manipulations were ineffective, as previous research has found that cognitive load can affect relative category use without affecting overall recall or task performance (Gilbert & Hixon, 1991; Macrae et al., 1993; Sherman et al., 2004).

Analyses

In order to better understand the pattern of results across studies, all analyses were performed as meta-analyses. The first set of meta-analyses was used to answer the following questions: whether categories were used more than individuating information under cognitive load, and in normal processing conditions (as per claim 1), and whether the relative use of categories vs. individuating information increased under load, compared to no load conditions (as per claim 2).

Similar to previous studies (see study 3 for additional details), each study was analyzed by predicting the main target trait ratings from the trait ratings for the category and the relevant individuating information, interacted with their relative certainty. All predictors were also interacted with load condition (effect coded so 1 = load, -1 = no load). For study 7 and 8, if none of the individuating information was selected as relevant to a trait, the trait and certainty ratings of ‘someone who you know nothing about’ were used instead. As in previous studies, information use was examined at three levels of relative certainty. Rather than defining high and low relative certainty separately for each study, a consistent value of +/- 0.20 was used to make the results more comparable across studies. This value was chosen by computing 2 SD relative certainty for all six studies and using the smallest of these values, which ensured results were not extrapolating beyond the range of any study’s data. Using the regression models for each study, estimates and standard errors were computed for: 1) the difference between category and individuating information slopes within each condition (load and no load), and 2) for the change in this slope difference across conditions (i.e. the slope difference in load condition minus the slope difference in no load condition). Standard errors were computed mathematically for equal certainty cases, and using bootstrapping for high and low relative certainty cases.

To combine results across studies, a mixed-effects meta-analysis was performed. Separate meta-analyses were done for the slope difference in the load condition, the slope difference in the no load condition, and the change across conditions. Each meta-analysis included all six studies, and all three estimates for a given study based on different levels of relative certainty. The figures below display the individual results included in the meta-analyses, for comparison to the overall effects.

Are Categories used More than Individuating Information when their Relative Reliability is the Same?

Under Cognitive Load. The first question examined with these analyses was whether, in the cognitive load condition, categories would be overused – i.e. used more than individuating information of equal relative reliability (as per claim 1). The cognitive load condition should be most likely to limit effortful processing, and therefore most likely to lead to category overuse, if categories are in fact used as heuristics. However, even in this condition, meta-analysis results showed little evidence of category overuse: there was no significant difference in the use of category and individuating information (see Figure 4; Slope Diff. = -0.05, 95% CI [-0.12, 0.02], $p = 0.173$, Bias(23) = -1.09, 95% CI [-2.67, 0.48]; $Q = 43.45$, $p < .001$), and analyses accounting for prior beliefs produced similar results. Confidence intervals here indicate that if categories are used more than individuating information, the effect is likely negligibly small, with a judgment bias of less than 1 point at the upper confidence limit. Furthermore, as can be seen in Figure 4, in only one out of 18 tests were categories used significantly more than individuating information (in study 4, with low relative certainty).

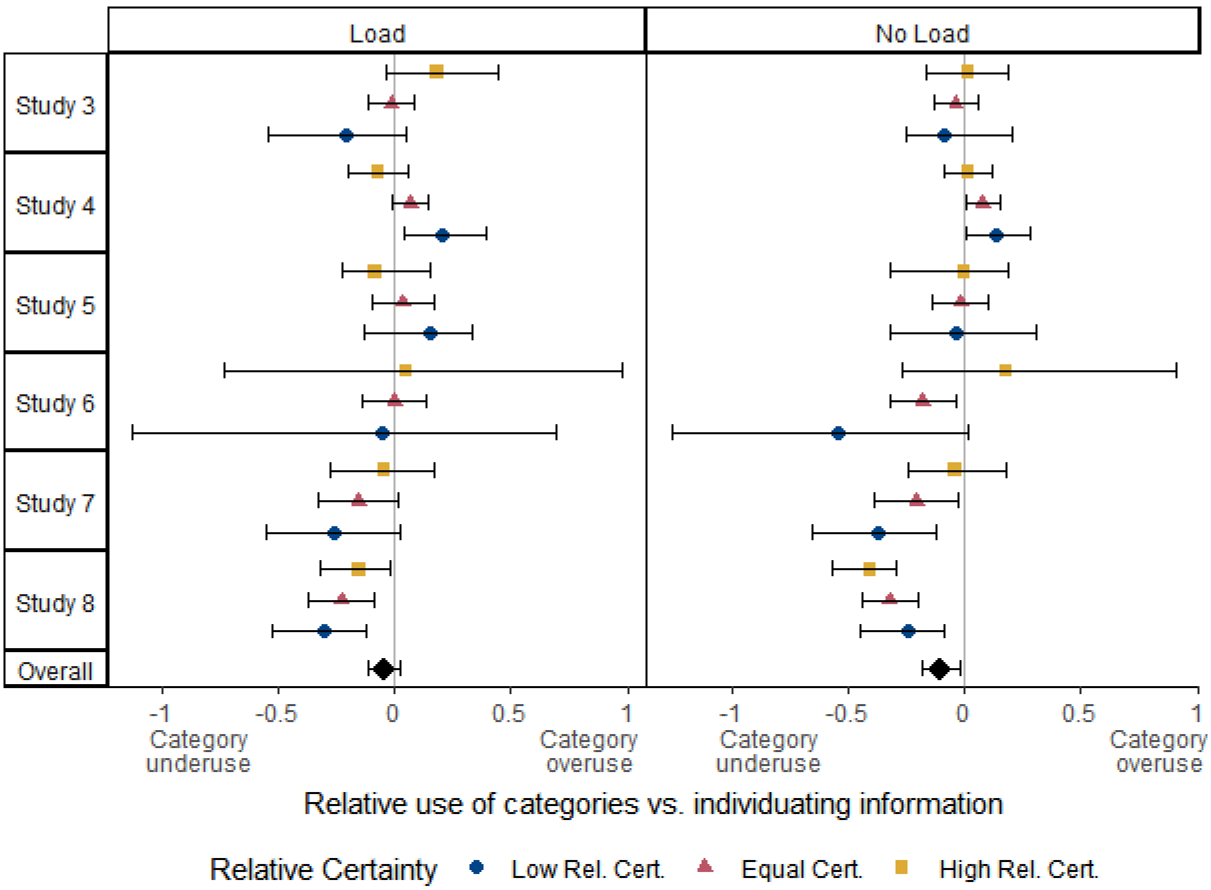
Instead of overusing categories, the overall effect indicated a trend towards overusing individuating information. This effect was strongest in studies 7 and 8, and was significant for all three tests in study 8. This is perhaps surprising, as studies 7 and 8 included several factors expected to increase the use of categories compared to individuating information (e.g. presenting categories first, more concisely, and never under load). However, these factors appear to have had the opposite effect. This could have occurred, for example, if the longer, more detailed format of the relevant individuating information made it more salient and therefore easier, rather than more difficult, to process.

Overall, these results indicate that, contrary to claim 1, categories were not overused compared to individuating information, even when effortful processing was limited by cognitive load. This provides further evidence against the idea that people overuse categories as a heuristic to simplify judgments. Instead, categories were used similarly to individuating information, or at times underused compared to individuating information, though the latter effect was likely driven by differences in how the two types of information were presented.

In Normal Processing Conditions. The next question examined was whether categories overused compared to individuating information in normal processing conditions – that is, in the no load condition. Meta-analysis results again showed that this did not happen, and instead individuating info was on average overused compared to categories (see Figure 4; Slope Diff. = -0.11, 95% CI [-0.19, -0.02], $p = 0.012$, Bias(23) = -2.42, 95% CI [-4.31, -0.53]; $Q = 85.78$, $p < .001$). Analyses accounting for prior beliefs produced similar results, though the effect did not reach significance. As shown in Figure 4, this appeared to be driven mainly by studies 8-10, where this effect was significant in six out of nine individual tests for these studies. This overuse of individuating information is therefore likely again due to differences in how these two types of information were presented, as all three of these studies presented the category first, never under load, and, in study 7 and 8, more concisely. On the other hand, there appeared to be little difference in how categories and individuating information were used in studies 3-5, where the information was presented in similar formats. Therefore, like studies 1 and 2, these results provide further evidence that categories are not overused under normal processing conditions, contrary to claim 1. These results also provide support for theories which propose that both types of information are used similarly, since studies 3-5 presented both types of information similarly, and found that neither underuse nor overuse of individuating information was observed.

Figure 4

Categories were not used more than Individuating Information of the Same Relative Certainty, either under Cognitive Load or in Normal Processing Conditions



Note. In the load condition, on average, there was no significant difference between the use of categories and individuating information of equivalent relative certainty. In the no load condition, on average, individuating information was used more than categories of equivalent relative certainty. Values above (/below) 0 on the x-axis indicate that categories were used more (/less) than individuating information of equivalent relative certainty; values of 0 indicate that they were used equally. Error bars represent 95% confidence intervals. Relative certainty is plotted at equal certainty (0 relative certainty), and high and low relative certainty (+/- 0.20 relative certainty).

Does the Relative use of Categories compared to Individuating Information Increase under Cognitive Load? The next set of analyses examined whether people would make relatively more category-based judgments under cognitive load, compared to the no load conditions, which should occur if cognitive load increases people's reliance on categories as heuristics (as per claim 2). Meta-analysis results did not support this claim. Overall, there was no significant effect of cognitive load on relative information use (see Figure 5; Change in Slope Diff. = 0.04, 95% CI [-0.01, 0.10], p = 0.13, Change in bias(23) = 1.97, 95% CI [-0.57, 4.52]; Q = 15.14, p = 0.59). Furthermore, only 1/18 tests were significant: only study 8 showed increased relative category use, and this was found only at high relative certainty. Analyses accounting for prior beliefs overall showed a trending but small effect towards greater relative category use, again driven primarily by study 8 (the only study with a significant effect in this version of the analyses). These results therefore do not provide strong support for the claim that people make relatively more category-based judgments under cognitive load. Furthermore, even though study 8 may provide partial support for this claim, these results may have been driven by differences in how categories and individuating information were presented, rather than inherent differences in how these types of information are represented and processed.

To further test whether load might lead to relatively more category-based judgments, a second meta-analysis was done to more closely replicate analyses used in previous research (including Pendry, 1998). This analysis examined more directly whether cognitive load caused target judgments to be more in line with the category rather than individuating information. For this analysis, judgments were only included if participants viewed the category and individuating information as being judgment-inconsistent (i.e. the category was judged above average and the individuating information below, or vice versa). The main target trait ratings were recoded so that relatively more category-based judgments would always lead to higher values, by reverse coding cases where the category was rated below average. For each study, the recoded target trait ratings were predicted from load condition (effect coded as in the previous analyses). This allowed for testing whether judgments made under load, compared to no load conditions, were relatively more in line with the category rather than the individuating information. Effect estimates and standard errors from each study were used in a mixed-effect meta-analysis.

Results of this additional test again did not show any effect of load on people's relative information use. On average, load had no significant effect (b = 0.72, 95% CI [-0.18, 1.62], p = 0.118; Q = 1.31, p = 0.934), and the results suggested that, if a real effect exists, it is likely negligibly small (the upper confidence limit suggested at most a 1.62 point change in judgments on the 100 point scale used here). Furthermore, no single study provided significant results. A follow up meta-analysis examined only study 9 and 10 in combination, the two studies that attempted to replicate Pendry (1998). Here, the average effect of load was again not significant (b = 1.39, 95% CI [-0.48, 3.25], p = 0.145; Q = 0.18, p = 0.675), with any real effects again likely being very small (the upper confidence limit suggested at most a 3.25 point change in judgments).[13] We interpret this as a failure to replicate the main results of Pendry's study, where cognitive load was found to lead to relatively more category-based judgments.

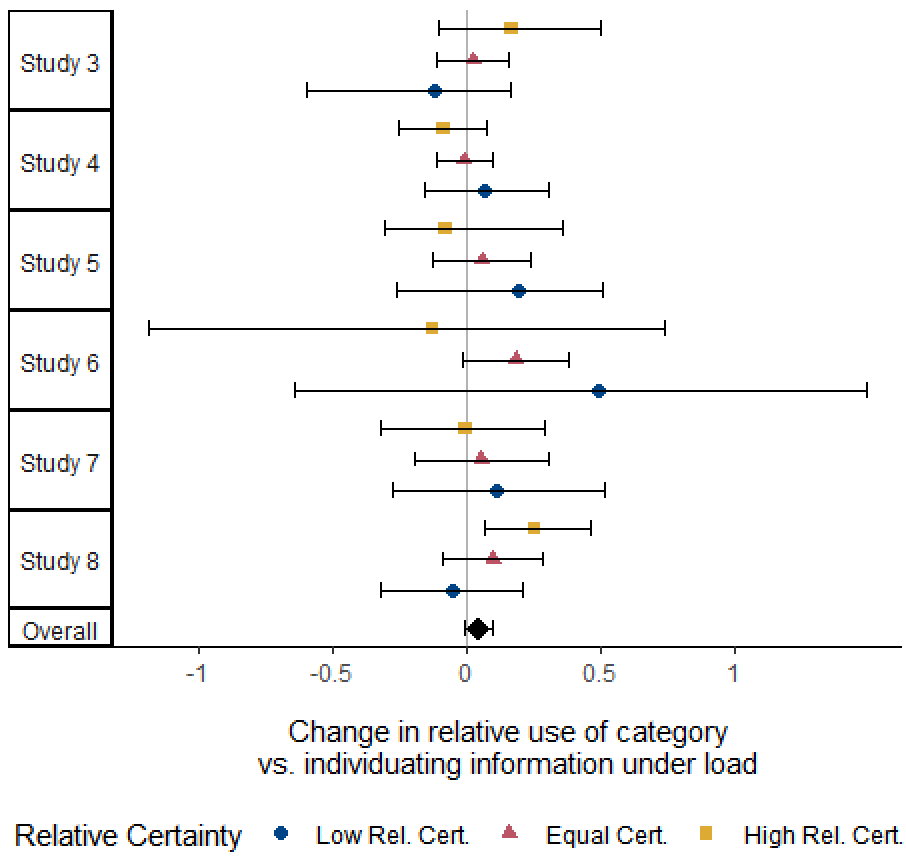
Overall, these results do *not* support the claim that people will make relatively more category-based judgments under load, due to a greater reliance on categories as heuristics (as per claim 2), with only study 8 providing weak and inconsistent support for this claim. Instead, cognitive load generally led to negligible changes in relative information use, even in studies that

[13] This effect was even smaller when the analysis was made more similar to Pendry's (1998): i.e. reverse coding target trait ratings based on if they were considered atypical of the category in pre-testing, and including all category-relevant trait ratings regardless of judgment-consistency (for study 7 and 8 combined: b = 0.27, 95% CI [-0.94, 1.47], p = 0.665; Q = 0.07, p = 0.791).

included several other factors that could increase the relative reliance on categories, and even when closely replicating the methods of previous research.

Figure 5

Cognitive load did not Significantly Change the Relative use of Categories compared to Individuating Information



Note. Values above (/below) 0 on the x-axis indicate that the use categories compared to individuating increased (/decreased) under cognitive load; values of 0 indicate no change. Error bars represent 95% confidence intervals. Relative certainty is plotted at equal certainty (0 relative certainty), and high and low relative certainty (+/- 0.20 relative certainty).

Discussion

Studies 3-8 looked for evidence that categories would be used as heuristics in conditions of varying cognitive load. Overall, there was little support for this idea, despite the fact that cognitive load should reduce people's ability to engage in effortful processing, and thus increase people's reliance on heuristics. Specifically, these studies found that categories were not overused compared to individuating information (contrary to claim 1), either in general or under cognitive load, providing evidence against theories of categories as heuristics. Instead, categories were used similarly to individuating information, or, in some cases, underused relative to individuating information (likely due to differences in presentation format). These results are in line with theories proposing that categories and individuating information are represented and processed similarly, and thus are not used differently unless there are other extraneous differences between the two types of information (e.g. differences in how they are presented). This further suggests that the overuse of individuating information observed in study 1 and 2 is not necessarily a generalizable effect, and thus is unlikely to reflect inherent or reliable differences in how these two types of information are processed or represented.

The current studies also did not generally find that cognitive load led to relatively more category-based judgments, contrary to what would be expected if people increased their use of categories as heuristics under load (i.e. contrary to claim 2). This general lack of effect was observed in studies that removed confounds from previous research (studies 3-5), as well as studies that tried to more closely replicate previous research, by including some of these confounds (studies 6-8) or by attempting to fully replicate all aspects of an existing study design (studies 7-8). Only one study, study 8 provided some evidence of relatively more category-based judgments under load; however, these effects were generally small and not always reliable across analyses, and may have been driven by differences in presentation format. Manipulation checks suggested that load manipulations were effective in at least study 7 and 8; however, even in these cases, load effects were at best unreliable. These null or unreliable results provide evidence against the idea that categories are easier to process and use, and thus tend to be relied on as a heuristic to simplify inferences. Moreover, the failure to replicate previous research in studies 6-8 (especially the direct replications in studies 7 and 8) also suggest the need to re-evaluate other evidence about how cognitive load will affect these judgments.

General Discussion

The current research re-examined how people integrate social categories (e.g. demographics, occupations and roles) and individuating information (e.g. personality traits, behaviors, and other attributes) to make inferences about others. Previous theories have argued that categories tend to be easier to use than individuating information, so that, when both types of information are available, people may overuse categories as a heuristic to simplify their judgments (Bodenhausen et al., 1999; Fiske & Neuberg, 1990). However, other theories suggest that these two types of information will be used similarly, unless there are other extraneous differences between them (e.g. differences in informativeness or presentation format; Köpetz & Kruglanski, 2008; Kunda & Thagard, 1996). The current research therefore re-examined two key claims about how these types of information will be used, that follow from theories of categories as heuristics. Claim 1 is that categories will be overused compared to individuating information in a way that provides a less accurate strategy for inferences, either in general or in conditions that reduce effortful processing. Claim 2 is that people will rely on this heuristic more, and therefore make relatively more category-based judgments, in conditions that reduce effortful processing (i.e. greater informational fit or greater cognitive load).

Despite the large amount of research on this topic, the current studies provide some of the first clear and systematic tests of these two claims. Several important theoretical and methodological improvements enabled this. First, the current work clarified what counts as category overuse by defining it in contrast to valid inference, and developed a specific Bayesian model to identify category overuse in inferences about continuous traits. Second, the current work identified and removed several extraneous factors that may have led to previous results (e.g. experimental confounds and statistical artifacts), thus providing some of the first tests of whether categories would be used differently than individuating information, due to inherent differences in how these two types of information are processed or represented. Third, the current work systematically tested these claims across a range of content and conditions, including conditions of varying informational fit and cognitive load, thus helping ensure the generalizability of the results.

Using this approach, a series of eight studies found little to no support for these two key predictions of theories of categories as heuristics. Specifically, there was no evidence of category overuse in any case examined (going against claim 1). Instead, categories were either underused compared to individuating information, or else both types of information were used similarly. Furthermore, when testing if informational fit and cognitive load would lead to relatively more category-based judgments (claim 2), support for this claim was weak at best. In terms of informational fit, there was tentative support for this claim when varying category-consistency, though these effects were small and unreliable, while judgment-consistency did not have any clear effect. In terms of cognitive load, load also generally had little effect on relative information use. Study 8 provided the one possible exception to this, where judgments may have been slightly more category-based under load, though effects here were again small and unreliable, and may have been due to differences in presentation format.

Overall, these results provide evidence against the idea that people tend to systematically overuse categories compared to individuating information as a heuristic to simplify their inferences. This suggests that theories of categories as heuristics may need to be re-evaluated, including the assumption that these two types of information tend to be represented and processed differently. Instead, these results fit with theories which suggest that both types of information are represented, processed, and used similarly, except in cases where other factors that can affect either type of information (e.g. presentation format) happen to affect one type more than the other. More broadly, this work suggests that to understand people's impressions and inferences, as well as the ways these can go wrong (e.g. by overusing categories), we may need to move beyond these broad content distinctions and focus more directly on underlying psychological mechanisms, to bring greater clarity to claims about heuristics, biases, and other errors in people's inferences, and to explore the variety of other reasons people might make these errors.

Theoretical Implications

Dissolving the Distinction between Categories and Individuating Information

One key implication of this work is that categories do not appear to be, on average, any easier to use than individuating information. This suggests that these two types of information may not tend to be represented or processed differently either. This contrasts with the assumptions of theories of categories as heuristics (Bodenhausen et al., 1999; Fiske & Neuberg, 1990). However, it is consistent with competing theories, which assume that these two types of information are represented, processed, and used similarly (Köpetz & Kruglanski, 2008; Kunda

& Thagard, 1996), and that assume that any factors that affect information use (e.g. differences in salience) can equally affect either type of information.

In support of this view, the current work highlighted how many of the apparent differences in how these two types of information are used can be explained by other factors (see also Beckett & Park, 1995; Köpetz & Kruglanski, 2008; Skorich & Mavor, 2013). For example, as discussed, previous findings of category overuse/underuse (e.g. Locksley et al., 1982), as well as most previous effects of informational fit and cognitive load (e.g. Lambert, 1995; Pendry, 1998), may have been driven by other factors like differences in how information was presented or statistical artifacts. Similar explanations might also account for other findings used to support theories of categories as heuristics. For example, findings that increased accuracy motivation leads to relatively less category-based judgments might be driven by similar factors to cognitive load effects, such as always presenting the category first (Fiske & Neuberg, 1989). The current studies therefore tested more directly whether these two types of information were used differently, by controlling for these other factors. In most cases, no differences were observed, and the few differences found were inconsistent either within or across studies.

One explanation for this lack of consistent differences is that, in contrast to theories of categories as heuristics, the content captured by the category-individuating information distinction may not map reliably onto differences in underlying processes or representations. For example, these theories propose that categories tend to have richer schemas than individuating information that allow them to better organize other information (Bodenhausen et al., 1999; Fiske & Neuberg, 1990). While this idea was supported by on some initial work comparing the schemas for ‘stereotypes’ – occupations (e.g. politician), roles (e.g. housewife), and other social types (e.g. bully) – to schemas for personality traits (e.g. extraverted, calm) (Andersen et al., 1990; Andersen & Klatzky, 1987), it is unclear how generalizable these differences in schema-richness are. This is because this initial work used only a limited set of content: it used a relatively small set of items, did not include things like demographic categories or behaviors, and specifically selected stereotypes that were expected to have richer schemas. These observed differences in schema-richness may therefore not reflect generalizable differences between categories and individuating information as a whole. Instead, schema-richness may simply reflect a dimension along which both categories and individuating information can vary widely.

Rather than trying to map these broad content distinctions onto different psychological mechanisms (processes or representations), a more productive research approach for future research may be to focus directly on these psychological mechanisms. This approach has been taken in work on entitativity and essentialism, which directly measures factors related to schema-richness and examines their impact on information use (Yzerbyt et al., 2004). A similar approach could be used to examine other ways that categories and individuating information are thought to differ. For example, judgments based on categories may often be seen as less socially or morally acceptable than those based on individuating information (Banaji & Bhaskar, 2000; Yzerbyt et al., 1998) (e.g. it is ok to use past behavior in hiring decisions, but not gender). Yet, as with differences in schema-richness, this apparent pattern may be much less robust than it seems at face value. Judgments based categories are sometimes considered quite acceptable (e.g. perhaps assuming that a mathematician is good at math), judgments based on individuating information may sometimes be considered unacceptable (e.g. perhaps assuming that a women who wears revealing clothes is seeking sexual attention), and people can differ widely in how acceptable they view the same judgment (Cao et al., 2019; Carter et al., 2006; Montemurro & Gillen, 2013; Uhlmann et al., 2010). Therefore, rather than trying to make claims at the level of these broad information types, it may be more useful to look at the underlying psychological mechanisms – e.g. directly examining the perceived acceptability of using a given piece information for a judgment, as well as the causes and consequences of these perceptions. Shifting the focus from broad content distinctions to the underlying mechanisms in this way may provide deeper insights into how people make social inferences, and the various factors that can affect this process.

Clarifying Epistemic Standards regarding Social Category use

A second key contribution of this work is bringing clarity to the epistemic standards used to evaluate social inferences. The current work does this by clearly defining valid inferences for a specific set of cases: when combining categories and individuating information to make inferences about continuous traits (see online supplement for further assumptions about the cases being considered). In contrast, existing research often makes evaluative or normative claims about category-based inferences, without specifying the standards being used for these evaluations (see Arkes & Tetlock, 2004; Jussim et al., 2009). For example, referring to category use as a 'heuristic' or 'bias' (e.g. Bodenhausen, 1990; Fiske & Neuberg, 1990; Greenwald & Krieger, 2006) implies that some sort of error is being made, but this is sometimes done without specifying exactly what kind of error is involved, or what would count as correct. Indeed, in some cases, 'biases' may not reflect epistemic errors at all, but instead reflect moral errors, such as violating egalitarian norms (Arkes & Tetlock, 2004; Banaji & Bhaskar, 2000). This unclarity can make it difficult to empirically establish whether a given type of error is occurring. In contrast, Bayesian models can help bring the clarity required to test for these errors. The current work provides a prime example of this, where clearly defining valid inferences allowed for testing whether people overuse categories in the ways predicted by theories of categories as heuristics.

The current model, and this Bayesian approach more broadly, can also bring greater clarity to more explicit discussions of how social categories should be used to make accurate inferences (Arkes & Tetlock, 2004; Jussim et al., 2009). In particular, the mathematical formalism of the Bayesian approach allows for greater precision, nuance, and justification than is typically found in purely intuitive, verbal discussions of this question. For example, compare the current perspective to a related argument for how categories can sometimes help make accurate inferences about individuals (Jussim et al., 2009). According to this argument, categories can sometimes help make inferences more accurate on average, at least when people have fairly accurate beliefs about the category's average trait, and when they lack fully informative individuating information. While these claims are largely in line with the current Bayesian model, the current model highlights additional factors that affect the accuracy of the resulting inferences. For example, using a category might still reduce the accuracy of people's inferences in these cases, if people misestimate the category's reliability (e.g. due to thinking the group is overly homogeneous), misrepresent its relationships to other characteristics (e.g. due to being unaware of causal relationships involved), or misuse the information due to biases in their inference processes (such as those discussed in the current work). As this example highlights, by formalizing the beliefs, assumptions, and inference patterns involved in making valid inferences, Bayesian modelling can bring further nuance to these types of normative discussions. Clarifying how categories should be used to make accurate inferences can in turn help understand the various ways people may fail to do this.

Future work can leverage the power of this Bayesian approach to further examine how people make inferences about others. Using the current model, future work can examine other reasons people might overuse categories compared to individuating information. The current model can also be extended, for example, to model various causal relationships, or to model how beliefs about groups are themselves inferred. In each case, these models can be used to clarify what counts as a valid inference, and to test if people deviate from this. This in turn can help develop new theories to understand how people make these social inferences, and when and why errors occur in this process. This approach has been enormously productive in related areas, including causal reasoning, categorization, and theory of mind (Baker et al., 2009; Gopnik & Wellman, 2012; Sanborn et al., 2010). However, applications of Bayesian models to stereotyping and impression formation are just beginning to be explored (e.g. Cao et al., 2017; Lau et al., 2018; Locksley et al., 1982; McCauley & Stitt, 1978; Vélez & Gweon, 2020), making this a promising avenue for future research.

Alternative Explanations for Observed Differences between Categories and Individuating Information

Though the current work largely did not find reliable differences in how categories and individuating information were used, two cases where potential differences were found deserve further discussion. One case was in studies 1 and 2, where individuating information was overused compared to categories, despite being presented in a similar format. However, these results were inconsistent across studies, as no overuse was found in studies 3-5. The mixed results here, as well as in previous research (Beckett & Park, 1995; Locksley et al., 1980, 1982; Rasinski et al., 1985), suggest that the observed differences do not stem from reliable differences in how categories and individuating information are processed or represented. Instead, when they occur, they may have other explanations. One possibility is that, for certain questions and content, people bring in additional assumptions beyond those in the current Bayesian model (see online supplement). For example, one of the questions from study 1 was, “Someone is a bodybuilder and lifts weights very often. How physically fit is this person?" People might bring in additional beliefs about the causal relationships in this scenario: e.g. that being a bodybuilder (the category) causes one to lift weights often (the individuating information), which in turn makes one physically fit. With this causal structure, the category provides no additional information once the individuating information is known, meaning that apparent overuse of the individuating information here could actually reflect a valid inference (Rottman & Hastie, 2014). Future work modelling these types of causal relationships can examine whether they can account for the results of the current studies, and whether there might be other scenarios in which these types of causal structures will lead to the opposite effect: the apparent overuse of categories.

Another case where the use of these two types of information differed was in their response to category-consistency in study 1. Though the results were somewhat unreliable and should be treated with caution, they tentatively fit with predictions of theories of categories as heuristics, as category-consistency appeared to increase the use of categories, while decreasing the use of individuating information. This increased use of categories is largely in line with previous findings (Craig & Bodenhausen, 2018; Fiske et al., 1987; Monroe et al., 2018; Peters & Rothbart, 2000), though the decreased use of individuating information differed from previous work, where either no change or increased use was observed (Fiske et al., 1987; Monroe et al., 2018). If the current results turn out to replicate, future research will need to examine what causes them, since the overall results of current work cast doubt on existing explanations. For example, consider the explanation given in the continuum model (Fiske & Neuberg, 1990): while categories tend to be used initially as organizing frameworks, it is primarily when informational fit is poor that people are motivated to incorporate additional individuating information. However, the lack of other evidence that categories are used initially as organizing frameworks (e.g. the lack of cognitive load effects) and the lack of other informational fit effects (i.e. based on judgment-consistency) in the current work suggest that this explanation might not hold. Other explanations that assume asymmetries between these two types of information, without explaining, testing, or justifying them, run into similar issues. For example, another proposed explanation for category-consistency effects is that when targets are atypical of the category (e.g. a Chinese person who is also bad at math), the category becomes less salient, and so has less impact on judgments (Craig & Bodenhausen, 2018). However, this logic works equally well in reverse: that when targets are atypical of the individuating information (e.g. someone bad at math who is also Chinese), the individuating information could become less salient, and so have less impact on judgments. Therefore, open questions remain about whether and why category-consistency might affect these two types of information differently. A more detailed examination of people's underlying beliefs, as well as Bayesian models that account for them, could help understand any differences more fully.

Alternative Explanations for Cases where Categories and Individuating Information did Not Differ

Aside from the two cases just discussed, the remaining results suggested that categories and individuating information were used similarly. Studies 3-5 did not find differences in how much these two types of information were used when their relative reliability was the same. Furthermore, judgment-consistency and cognitive load had little clear effect on information use in studies 2-8. This held even when including several additional factors that may have led to cognitive load effects in this previous work (studies 6-8), and even when trying to closely replicate the methods of previous research (studies 7-8). Though there are some reasons to be cautious about over-interpreting these null results, especially where they directly conflict with previous work, these results still provide evidence against theories of categories as heuristics, and the idea that these two types of information tend to be used differently.

In general, one reason to be cautious about interpreting null results is that they may simply be caused by inadequately powered studies. However, this seems unlikely in the current case, as several aspects of the current research helped ensure adequate power. Specifically, the current studies included large sample sizes or meta-analyses across studies, multiple responses per person, and, in many cases, within-participant manipulations of informational fit and cognitive load. Furthermore, studies 7 and 8, which attempted to directly replicate previous work (Pendry, 1998), were each estimated to have over 96% power to find effects of the same size as in previous work. While there is no guarantee that the current studies had enough power to detect true effects, the reported confidence intervals suggest that, in many cases, any effects would likely have to be quite small, thus may be of little practical significance.

Turning specifically to the judgment-consistency and cognitive load results, another reason one might hesitate to interpret these null results is that, in previous research, both of these conditions appear to have led to relatively more category-based judgments. Yet, as discussed, these previous findings could have been driven by various statistical artifacts or experimental confounds. In contrast, studies 2-5 provided the first tests of these effects without these limitations, and found null results. Therefore, the current balance of evidence suggests that judgment-consistency and cognitive load do not alter the relative use of categories and individuating information, at least not due to inherent differences in how these two types of information are represented and processed, as suggested by theories of categories as heuristics.

A more significant conflict with previous research exists in studies 6-8, which failed to replicate previous cognitive load effects (e.g. Pendry, 1998) even when including additional confounds that may have led to these effects in previous work (i.e. presenting categories first, more concisely, and never under load), and even when trying to fully replicate all aspects of a previous study design. It is unclear what to conclude about load effects in these specific conditions. The much larger sample sizes in the current studies compared to most of this previous work suggests that previous findings may have been false positives, or overestimations of negligibly small effects – which could be quite prevalent if other null results in this area remained unpublished (the file drawer problem). Alternatively, the current work may not have fully recreated the conditions required to find previous effects. For example, perhaps participants' motivation was too low, so that they used categories as heuristics equally in both the load and no load conditions. Yet this seems unlikely, given that there was no evidence of category overuse in either condition, and manipulation checks indicated that load was successfully manipulated in at least studies 7 and 8. Even if some other aspect of the current methods accounts for these different results, this lack of robustness across contexts suggests that, if cognitive load does increase relative category use at times (even if this is due to differences in presentation format), the conditions in which this occurs are likely even narrower than previously thought.

Might Social Categories Still be Overused at Times?

Though the current studies found no evidence that categories were systematically overused compared to individuating information (i.e. in ways that deviate from valid inferences given people's beliefs), this does not mean that categories will never be overused in this way. Instead, it suggests that we may need to go beyond theories of categories as heuristics to understand when and why this occurs. It may be that category overuse occurs only for certain types of categories, for certain people and/or in certain conditions other than those examined here. For example, people who are especially prejudiced against a certain group might overly attend to and thus overuse information about that group. Similarly, categories might be overused when the category information happens to be presented in an easier to process format – e.g. upon first seeing a new person, their age, race, and gender may be rapidly and easily processed (Fiske & Neuberg, 1990), but subtler visual cues (e.g. whether their clothes are freshly ironed) might be more difficult to notice. Indeed, any factor known to increase the salience or use of a category might plausibly lead its overuse. An important goal for future research will therefore be to test which of these factors in fact lead to category overuse, and which might instead reduce category underuse, or reflect valid inferences in all cases. For example, cognitive load may sometimes increase category use by reducing people's ability to suppress categories that they believe are informative, yet morally problematic to use (e.g. using race or gender in hiring decisions; Sczesny & Kühnen, 2004). However, this does not necessarily mean the category is overused under load. Instead, cognitive load might decrease category under-use, so that inferences are in fact more in line with people's beliefs about the category, despite being less in line with people's moral values. As another example, perceiving a group as more homogeneous or as more of a coherent entity may also increase use of that category (Hamilton et al., 2009). Yet if people see these types of categories as more informative, this increased category use may reflect valid inferences in all cases. By providing tools to clearly identify category overuse, the current work allows us to begin answering these types of questions.

Even when people are reasoning validly from their beliefs, future work can also examine if people overuse categories in a different sense: through having incorrect beliefs, such as believing an exaggerated stereotype about a group. This type of overuse can also reduce the accuracy of people's resulting inferences, and lead to problematic social consequences (e.g.

increasing discrimination). To understand category overuse in this broader sense, it is important to understand when and why these kinds of inaccurate beliefs will occur. As highlighted by the current Bayesian model, this applies not only to beliefs about the group's average traits, as most commonly studied (Jussim et al., 2009), but also to the various other beliefs that go into people's inferences, such as beliefs about a group's homogeneity or reliability (Judd et al., 1991; Judd & Park, 1993; Vrantsidis & Cunningham, in press) and beliefs about how different types of information are causally or otherwise related (Carvalho et al., 2021).

Conclusions

The current research re-examined fundamental questions about how categories and individuating information are combined to form impressions and make inferences about others. In particular, it focused on the idea that categories tend to be overused compared to individuating information as a heuristic to simplify judgments (Bodenhausen, 1990; Fiske & Neuberg, 1990) (in contrast to alternative theories proposing these two types of information are used similarly; Köpetz & Kruglanski, 2008; Kunda & Thagard, 1996) and re-examined two key claims that follow from these ideas. The current work made several important contributions to this area. It clarified what counts as category overuse, through using Bayesian modelling to define valid inferences. It identified and overcame important methodological challenges that prevented most previous work from fully testing these claims. It then provided some of the first clear and systematic tests of these claims, and found almost no support for them, suggesting that categories are not overused due to their role as heuristics, and instead may often be used similarly to individuating information. More broadly, by challenging existing theories and assumptions about the differences between categories and individuating information, and by demonstrating the powerful conceptual and methodological tools provided by a Bayesian approach, this work opens up the potential to better understand when and why categories might be overused. This can be extended to understand the myriad of ways that social inferences might go wrong, and thus help understand and address the negative social consequences (e.g. increased discrimination) that this can lead to.

References

Andersen, S. M., & Klatzky, R. L. (1987). Traits and Social Stereotypes: Levels of Categorization in Person Perception. *Journal of Personality and Social Psychology*, *53*(2), 235–246. https://doi.org/10.1037/0022-3514.53.2.235

Andersen, S. M., Klatzky, R. L., & Murray, J. (1990). Traits and social stereotypes: Efficiency differences in social information processing. *Journal of Personality and Social Psychology*, *59*(2), 192–201. https://doi.org/10.1037/0022-3514.59.2.192

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, *2*(1), 1–9. https://doi.org/10.1037/h0021966

Anderson, N. H. (2014). *A functional theory of cognition*. Psychology Press.

Andrews, K. (2017). More stereotypes, please! The limits of 'theory of mind' and the need for further studies on the complexity of real world social interactions [Peer commentary on "Précis of Social Perception and Social Reality: Why accuracy dominates bias and self-fulfilling prophecy," by L. Jussim]. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X15002290

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: Psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*, *19*(4), 452–458. https://doi.org/10.1016/j.conb.2009.06.008

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?" *Psychological Inquiry*, *15*(4), 257–278. https://doi.org/10.1207/s15327965pli1504_01

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Banaji, M., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In Schacter, D. L. & Scarry, E. (Eds.), *Memory, brain, and belief* (pp. 139–175). Harvard University Press.

Beckett, N. E., & Park, B. (1995). Use of Category Versus Individuating Information: Making Base Rates Salient. *Personality and Social Psychology Bulletin*, *21*(1), 21–31. https://doi.org/10.1177/0146167295211004

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 12–23. https://doi.org/10.1037/0096-1523.14.1.12

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The Automaticity of Race and Afrocentric Facial Features in Social Judgments. *Journal of Personality and Social Psychology*, *87*(6), 763–778. https://doi.org/10.1037/0022-3514.87.6.763

Bless, H., Schwarz, N., Bodenhausen, G. V., & Thiel, L. (2001). Personalized versus Generalized Benefits of Stereotype Disconfirmation: Trade-offs in the Evaluation of Atypical Exemplars and Their Social Groups. *Journal of Experimental Social Psychology*, *37*(5), 386–397. https://doi.org/10.1006/jesp.2000.1459

Bless, H., Schwarz, N., & Wieland, R. (1996). Mood and the impact of category membership and individuating information. *European Journal of Social Psychology*, *26*(6), 935–959. https://doi.org/10.1002/(SICI)1099-0992(199611)26:6<935::AID-EJSP798>3.0.CO;2-N

Blessum, K. A., Lord, C. G., & Sia, T. L. (1998). Cognitive Load and Positive Mood Reduce Typicality Effects in Attitude-Behavior Consistency. *Personality & Social Psychology Bulletin*, *24*(5), 496–504. https://doi.org/10.1177/0146167298245005

Bodenhausen, G. V. (1990). Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychological Science*, *1*(5), 319–322. https://doi.org/10.1111/j.1467-9280.1990.tb00226.x

Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In Chaiken, S. & Trope, Y. (Eds.), *Dual-process theories in social psychology* (pp. 271–290). Guilford Press. https://escholarship.org/uc/item/3rm7v8wk

Bond, C. F., & Brockett, D. R. (1987). A social context-personality index theory of memory for acquaintances. *Journal of Personality and Social Psychology*, *52*(6), 1110–1121. https://doi.org/10.1037/0022-3514.52.6.1110

Bond, C. F., & Sedikides, C. (1988). The recapitulation hypothesis in person retrieval. *Journal of Experimental Social Psychology*, *24*(3), 195–221. https://doi.org/10.1016/0022-1031(88)90036-4

Canty, A., & Ripley, B. D. (2019). *boot: Bootstrap R (S-Plus) Functions.* (1.3-22) [R package].

Cao, J., Kleiman-Weiner, M., & Banaji, M. R. (2017). Statistically inaccurate and morally unfair judgements via base rate intrusion. *Nature Human Behaviour, 1*(10), 738–742.

Cao, J., Kleiman-Weiner, M., & Banaji, M. R. (2019). People Make the Same Bayesian Judgment They Criticize in Others. *Psychological Science*, *30*(1), 20–31. https://doi.org/10.1177/0956797618805750

Carter, J. D., Hall, J. A., Carney, D. R., & Rosip, J. C. (2006). Individual differences in the acceptance of stereotyping. *Journal of Research in Personality*, *40*(6), 1103–1118.

Carvalho, K., Peretz-Lange, R., & Muentener, P. (2021). Causal Explanations for Weight Influence Children's Social Preferences: Biological-Essentialist Explanations Reduce, and Behavioral Explanations Promote, Preferences for Thin Friends. *Child Development*, *92*(2), 682–690.

Chasteen, A. L. (2000). The Role of Age and Age-Related Attitudes in Perceptions of Elderly Individuals. *Basic and Applied Social Psychology*, *22*(3), 147–156. https://doi.org/10.1207/S15324834BASP2203_3

Chun, W. Y., & Kruglanski, A. W. (2006). The role of task demands and processing resources in the use of base-rate and individuating information. *Journal of Personality and Social Psychology*, *91*(2), 205–217. https://doi.org/10.1037/0022-3514.91.2.205

Craig, M. A., & Bodenhausen, G. V. (2018). Category (Non)Fit Modulates Extrapolative Stereotyping of Multiply Categorizable Social Targets. *Social Cognition*, *36*(5), 559–588. https://doi.org/10.1521/soco.2018.36.5.559

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press.

Eva, B., & Hartmann, S. (2018). Bayesian argumentation and the value of logical validity. *Psychological Review*, *125*(5), 806–821. https://doi.org/10.1037/rev0000114

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In *Dual-process theories in social psychology* (pp. 231–254). The Guilford Press.

Fiske, S. T., & Neuberg, S. L. (1989). Category-based and individuating processes as a function of information and motivation: Evidence from our laboratory. In *Stereotyping and prejudice* (pp. 83–103). Springer.

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Academic Press. https://doi.org/10.1016/S0065-2601(08)60317-2

Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, *23*(5), 399–427. https://doi.org/10.1016/0022-1031(87)90038-2

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology*, *33*(5), 573–589. https://doi.org/10.1002/ejsp.166

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517. https://doi.org/10.1037/0022-3514.60.4.509

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085.

Greenwald, A. G., & Krieger, L. H. (2006). Implicit Bias: Scientific Foundations. *California Law Review*, *94*(4), 945–967. https://doi.org/10.2307/20439056

Hahn, U. (2014). The Bayesian boom: Good thing or bad? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00765

Hamilton, D. L., Sherman, S. J., Crump, S. A., & Spencer-Rodgers, J. (2009). The role of entitativity in stereotyping. In Todd D. Nelson (Ed.), *Handbook of prejudice, stereotyping and discrimination* (pp. 179–198). Psychology Press.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, *100*(1), 109.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology*, *61*(3), 366–379. https://doi.org/10.1037/0022-3514.61.3.366

Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In Nelson, Todd D (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–227). Taylor & Francis Group.

Kantowitz, B. H. (1974). Double stimulation. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 83–132). Erlbaum.

Köpetz, C., & Kruglanski, A. W. (2008). Effects of Accessibility and Subjective Relevance on the Use of Piecemeal and Category Information in Impression Formation. *Personality and Social Psychology Bulletin*, *34*(5), 692–705. https://doi.org/10.1177/0146167207313730

Krueger, J., & Rothbart, M. (1988). Use of categorical and individuating information in making inferences about personality. *Journal of Personality and Social Psychology*, *55*(2), 187–195. https://doi.org/10.1037/0022-3514.55.2.187

Kunda, Z., & Sherman-Williams, B. (1993). Stereotypes and the construal of individuating information. *Personality and Social Psychology Bulletin*, *19*(1), 90–99.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308. https://doi.org/10.1037/0033-295X.103.2.284

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/jss.v082.i13

Lambert, A. J. (1995). Stereotypes and social judgment: The consequences of group variability. *Journal of Personality and Social Psychology*, *68*(3), 388–403. https://doi.org/10.1037/0022-3514.68.3.388

Lambert, A. J., Chasteen, A. L., Payne, B. K., & Shaffer, L. (2004). Typicality and group variability as dual moderators of category-based inferences. *Journal of Experimental Social Psychology*, *40*(6), 708–722.

Lau, T., Pouncy, H. T., Gershman, S. J., & Cikara, M. (2018). Discovering social groups via latent structure learning. *Journal of Experimental Psychology: General*, *147*(12), 1881.

Leitgeb, H., & Pettigrew, R. (2010). An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. *Philosophy of Science*, *77*(2), 236–272. https://doi.org/10.1086/651318

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.3.5) [R package]. https://CRAN.R-project.org/package=emmeans

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *39*(5), 821–831. https://doi.org/10.1037/0022-3514.39.5.821

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, *18*(1), 23–42. https://doi.org/10.1016/0022-1031(82)90079-8

Lord, C. G., Lepper, M. R., & Mackie, D. (1984). Attitude prototypes as determinants of attitude–behavior consistency. *Journal of Personality and Social Psychology*, *46*(6), 1254–1266. https://doi.org/10.1037/0022-3514.46.6.1254

Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, *23*(1), 77–87. https://doi.org/10.1002/ejsp.2420230107

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, *66*(1), 37–47. https://doi.org/10.1037/0022-3514.66.1.37

Madon, S., Guyll, M., Hilbert, S. J., Kyriakatos, E., & Vogel, D. L. (2006). Stereotyping the Stereotypic: When Individuals Match Social Stereotypes. *Journal of Applied Social Psychology*, *36*(1), 178–205. https://doi.org/10.1111/j.0021-9029.2006.00057.x

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, *36*(9), 929.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621.

Monroe, B. M., Koenig, B. L., Wan, K. S., Laine, T., Gupta, S., & Ortony, A. (2018). Re-examining dominance of categories in impression formation: A test of dual-process models. *Journal of Personality and Social Psychology*, *115*(1), 1–30. https://doi.org/10.1037/pspa0000119

Montemurro, B., & Gillen, M. M. (2013). How clothes make the woman immoral: Impressions given off by sexualized clothing. *Clothing and Textiles Research Journal*, *31*(3), 167–181.

Neys, W. D. (2006). Dual Processing in Reasoning: Two Systems but One Reasoner. *Psychological Science*, *17*(5), 428–433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

Oruç, İ., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468. https://doi.org/10.1016/S0042-6989(03)00435-8

Pendry, L. F. (1998). When the mind is otherwise engaged: Resource depletion and social stereotyping. *European Journal of Social Psychology*, *28*(2), 293–299. https://doi.org/10.1002/(SICI)1099-0992(199803/04)28:2<293::AID-EJSP869>3.0.CO;2-7

Pendry, L. F., & Macrae, C. N. (1994). Stereotypes and Mental Life: The Case of the Motivated but Thwarted Tactician. *Journal of Experimental Social Psychology*, *30*(4), 303–325. https://doi.org/10.1006/jesp.1994.1015

Peters, E., & Rothbart, M. (2000). Typicality Can Create, Eliminate, and Reverse the Dilution Effect. *Personality and Social Psychology Bulletin*, *26*(2), 177–187. https://doi.org/10.1177/0146167200264005

R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Rasinski, K. A., Crocker, J., & Hastie, R. (1985). Another Look at Sex Stereotypes and Social Judgments: An Analysis of the Social Perceiver's Use of Subjective Probabilities. *Journal of Personality and Social Psychology*, *49*(2), 317–326. https://doi.org/10.1037/0022-3514.49.2.317

Rohde, M., Dam, L. C. J. van, & Ernst, M. O. (2016). Statistically Optimal Multisensory Cue Integration: A Practical Tutorial. *Multisensory Research*, *29*(4–5), 279–317. https://doi.org/10.1163/22134808-00002510

Rosenkrantz, R. D. (1992). The Justification of Induction. *Philosophy of Science*, *59*(4), 527–539. https://doi.org/10.1086/289693

Rottman, B. M., & Hastie, R. (2014). Reasoning about Causal Relationships: Inferences on Causal Networks. *Psychological Bulletin*, *140*(1), 109–139. https://doi.org/10.1037/a0031903

Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology*, *75*, 54–70. https://doi.org/10.1016/j.jesp.2017.11.009

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144.

Sczesny, S., & Kühnen, U. (2004). Meta-Cognition about Biological Sex and Gender-Stereotypic Physical Appearance: Consequences for the Assessment of Leadership Competence. *Personality and Social Psychology Bulletin*, *30*(1), 13–21. https://doi.org/10.1177/0146167203258831

Sherman, J. W., Conrey, F. R., & Groom, C. J. (2004). Encoding flexibility revisited: Evidence for enhanced encoding of stereotype-inconsistent information under cognitive load. *Social Cognition*, *22*(2), 214–232.

Sherman, J. W., Macrae, C. N., & Bodenhausen, G. V. (2000). Attention and Stereotyping: Cognitive Constraints on the Construction of Meaningful Social Impressions. *European Review of Social Psychology*, *11*(1), 145–175. https://doi.org/10.1080/14792772043000022

Skorich, D. P., & Mavor, K. I. (2013). Cognitive load privileges memory-based over data-driven processing, not group-level over person-level processing. *British Journal of Social Psychology*, *52*(3), 469–488. https://doi.org/10.1111/j.2044-8309.2012.02099.x

Trémolière, B., Neys, W. D., & Bonnefon, J.-F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, *124*(3), 379–384. https://doi.org/10.1016/j.cognition.2012.05.011

Uhlmann, E. L., Brescoll, V. L., & Machery, E. (2010). The Motives Underlying Stereotype-Based Discrimination Against Members of Stigmatized Groups. *Social Justice Research*, *23*(1), 1–16. https://doi.org/10.1007/s11211-010-0110-7

Vélez, N., & Gweon, H. (2020). Preschoolers use minimal statistical information about social groups to infer the preferences and group membership of individuals. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *The Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* (pp. 227–233). Cognitive Science Society.

Vidotto, G., & Vincentini, M. (2007). A general method for parameter estimation of averaging models. , 12, 1-2, 211-221. *Teorie & Modelli*, *12*, 211–221.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vrantsidis, T. H., & Cunningham, W. A. (in press). The Effect of First-Hand and Second-Hand Knowledge on Perceived Group Homogeneity and Certainty about Stereotype-Based Inferences. *Social Cognition*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* https://ggplot2.tidyverse.org

Wigboldus, D. H. J., Sherman, J. W., Franzese, H. L., & Knippenberg, A. van. (2004). Capacity and Comprehension: Spontaneous Stereotyping Under Cognitive Load. *Social Cognition*, *22*(3), 292–309. https://doi.org/10.1521/soco.22.3.292.35967

Wilson, A. R. (2016). *How African American Male Typicality Affects In-Group Stereotyping and Belonging: A Cross-Sectional Analysis* [Doctoral dissertation, UC Santa Cruz]. eScholarship. https://escholarship.org/uc/item/1q84n6zx

Yzerbyt, V., Dardenne, B., & Leyens, J.-P. (1998). Social judgeability concerns in impression formation. In *Metacognition: Cognitive and social dimensions* (pp. 126–156). Sage Publications, Inc. https://doi.org/10.4135/9781446279212.n8

Yzerbyt, V., Judd, C. M., & Corneille, O. (Eds.). (2004). *The psychology of group perception: Perceived variability, entitativity, and essentialism.* Psychology Press.

Zalinski, J., & Anderson, N. H. (1991). Parameter estimation for averaging theory. In N. H. Anderson (Ed.), *Contributions to information integration theory. Vol. 1: Cognition* (pp. 353–394). Erlbaum.