CAUSALITY

Description: David Hume famously claimed that causality is the cement of the universe. Today, we can re-interpret his claim by saying that causality is the *cognitive* cement of the universe. It is indeed a central notion in the representation of action that governs the observed behaviour in many different species. It links eventualities and predicts the consequences of action. It is the origin of behaviours that allow animals, notably human animals, to act upon and thus shape their environments. Causality is also the cognitive basis for the acquisition and the use of categories and concepts in the child. As such, it is the basis for rationality.

The workshop will bring together psychologists, philosophers, primatologists, linguists to explore the difference between animal and human causality, the role of language in shaping our causal reasoning as well as the role of association and memory in causal understanding.
Starting from March 1st 2005, a new paper will be open to discussion every two weeks.

French Version
La causalité une notion centrale dans la représentation de l'action, notion qui gouverne le comportement observé chez de nombreuses espèces. La causalité lie les éventualités et prédit les conséquences des actions. C'est l'origine des comportements qui permettent aux animaux, et notamment aux êtres humains, d'agir sur leur environnement et de le modifier. A contrario, le comportement peut être l'indice d'un raisonnement causal chez les animaux. Qui plus est, la causalité est aussi la base cognitive de l'acquisition et de l'usage des catégories et des concepts chez l'enfant. En tant que telle, c'est la fondation de la rationnalité.

Le séminaire réunira psychologues, anthropologues, philosophes, linguistes, primatologues, pour essayer d'éclairer les questions suivantes: : y a-t-il une spécificité de la cognition causale chez l'espèce humaine et, si c'est le cas, le langage joue-t-il ou non un rôle dans cette spécificité ? Quel est le rôle de l'apprentissage et de la mémoire associative dans le raisonnement causal?

Moderators:
Gloria Origgi (CNRS, Institut Jean-Nicod), **Anne Reboul** (CNRS - Institut des Sciences Cognitives Lyon)

Guest Panel:
David Eagleman (University of Texas, Houston Medical School), **York Hagmayer** (University of Goettingen), **Daniel Povinelli** (University of Louisiana at Lafayette), Juan Rosas (University of Jaen, Spain), **Laurie Santos** (Yale University), **Michael Waldmann** (University of Goettingen).

This workshop is organized by the Institute for Cognitive Science in Lyon and the University of Geneve.

- *Similarities and differences between human and nonhuman causal cognition.*
- *Similarités et différences entre la causalité humaine et non humaine*
Anne Reboul (CNRS - Institut des Sciences Cognitives Lyon)

- *Consciousness, Intentionality and Causality*
Walter Freeman (Berkeley University)

- *Causal logic and the intentional stance*
John Watson (University of California, Berkeley)

- *Do young children possess dinstinct causalities for the three core domains of thought?*
Kayoko Inagaki (Chiba University)
Giyoo Hatano (University of the Air, Japan)

- *Physical causality in human infants*
Susan Hespos (Vaderbilt Univeristy)

- *Causality in Non-Humans*
Jennifer Vonk (University of Louisiana, Lafayette)

- *Left with the association, naked as if it were? Ideas from honeybee learning*
Martin Giurfa (Researcher, CNRS Centre de Recherche sur la Cognition Animale)

- *Inferring Causality and Making Predictions. Some Misconceptions in the Animal and Human Learning Literature*
Helena Matute (Universidad de Deusto, Spain) and **Miguel A. Vadillo** (Universidad de Deusto, Bilbao, Spain)

- *Thinking About Action. The Logic of Intervention in Reasoning and Decision Making*
Steven Sloman (Brown University)

- *Death as an Empirical Backdoor to the Representation of Mental Causality*
Jesse M. Bering (University of Arkansas)

- *Causal Inferences. Evolutionary Domains and Neural Systems*
Clark Barrett (UCLA Department of Anthropology) and **Pascal Boyer**

- *Associative Learning in Animals and Humans*
Leyre Castro (University of Iowa) and **Edward A. Wasserman** (University of Iowa)

-*Delusion as an abnormal causal reasoning process. A search for a common ground in schizophrenia and dementia in older people*
Sebastien Carnicella (University Institute of Technology, Strasbourg) and **Philippe Oberling** (Faculté de Médecine de Strasbourg)

- *Causality vs. Explanation. Objective Relations vs. Subjective Interests.*
Denis Hilton (Universite de Toulouse)

- *The Possible Influence of Perception of Causal Events on the Development of "if P then Q" Conditionals and Causal Reasoning*
Peter Ford F. Dominey (CNRS - Institut des Sciences Cognitives, Lyon)

- *Expressing Causality in Natural Language. A Pragmatic Perspective*
Jacques Moeschler (Unversité de Genève)

- Causal Maps and Bayes Nets. A cognitive and computational account of causal learning and therory formation
Alison Gopnik (University of California, Berkeley)

Similarities and differences between human and nonhuman causal cognition

Anne Reboul (Researcher CNRS - Institut des Sciences Cognitives Lyon)

(Date of publication: 28 February 2005)

Abstract: Hume noted that causality was inductively inferred by humans on the basis of three factors: spatio-temporal contiguity of cause and effect, contingency of the effect to the cause and precedence of the cause. He also noted that the inductive basis of causality was the association of the cause and the effect. The three factors which he saw as triggering the notion of a causal link are very near to those used in analyses of associative learning in animals. However, most researchers on animal cognition wouldn't want to claim that causal cognition in nonhuman animals is equivalent to its human counterpart. Given the similarities, what are the differences and where do they come from?

Introduction

That causality has a central role in cognition, whether human and nonhuman, is not controversial. What might be controversial is whether 'cause' means exactly the same thing in human and nonhuman cognition. Or, in other, more philosophical, words, would attribution of a common causal belief — e.g., "The fact that it rains will cause Mother not to take me for walk" — to my dog, Tolkien, and to my 11-year-old daughter, Abigaël, make sense? On a superficial view, we might say that both exemplify, mutatis mutandis, the same behavior, Abigaël is curled in an armchair reading a book, Tolkien has sneaked into another armchair and both are casting melancholy eyes at the rain beating on the windows. The question, at a deeper level, is whether there is more to Abigaël's causal belief than a mere association between rain and no walk and whether, if there is, it might be legitimate to attribute that additional feature to Tolkien's causal belief too. For instance, Abigaël may have a mentalist explanation to the effect that I believe that rain makes one wet and that I don't like to get wet which is why I choose to stay indoors when it rains. This explanation, presumably, is not something that it would make sense to attribute to Tolkien. Can we say exactly in what the difference between the causal belief as attributed to Abigaël and as attributed to Tolkien lies? It seems to me that the difference lies in the fact that Abigaël has an explanation for the association whereas Tolkien is left with the association, naked as it were. What is more, Tolkien is not, and would not be, interested by an explanation, while Abigaël would not, indeed should not, be satisfied with the naked association. Though it may be adventurous to see the difference between human and nonhuman causal cognition as lying in the existence in the first and absence in the second of explanation, this is the claim I want to make. I even want to go slightly farther and say that association is mostly between perceptible entities, while explanation, more often than not goes beyond the observable (as is the case with Abigaël's explanation for why I won't take her for a walk when it rains). As Hume (1975, 74) famously noted, "All events seem entirely loose and separate. One event follows another; but we never can observe any tie between them. They seem conjoined, but never connected". Hume deduced from that basic observation to the perceptibility of the association and the non-perceptibility of the causal link the inexistence of the second, but I will not be concerned with that metaphysical claim here.

Are humans associative animals?

I claimed above that causal cognition in humans is not or is not merely associative. This claim can be (and has been) cashed in different ways. To begin with, Premack (1995) distinguished between arbitrary causal knowledge (hereafter ACK), resulting from associative learning — dependent on contiguity and repetition —, and natural causal knowledge (hereafter NCK), strongly domain specific and a priori — not dependent on contiguity and repetition. Another way of putting it might be to say that ACK is based on induction, while NCK can be used as the basis for deduction. Typical NCK in humans is relative to folk psychology, folk physics and folk biology. It is difficult if not impossible to attribute it to nonhuman animals. To the extent that it is not based on associative learning, it does obviously justify my claim. NCK is however not what I want to discuss here. I'll concentrate on ACK. Regarding ACK, there are two possibilities:

- It is based solely on associative learning in both human and nonhuman animals;
- Though it is based on associative learning in both human and nonhuman animals, associative learning is not sufficient for ACK in human animals, though it is sufficient in nonhuman animals.

This raises a further question, which has to do with why association is not sufficient for ACK in humans. After all, if association is adaptive for nonhuman animals, why should it not be enough for humans? Another obvious question concerns what explanation exactly is.

My (tentative) answer will be that the response to both questions goes somehow through the fact that humans are the only linguistic species. The rest of this paper will be devoted to a short review of those experimental works which purport to show that association is not the whole story of human ACK and to some, admittedly speculative, hypotheses on the role of language in the difference between human and nonhuman ACK.

To begin with a pivotal point of contemporary philosophical literature on causality (now being investigated experimentally — see Roese 1994, Roese & Olson 1996, 2003, Pennington & Roese 2003), there is a strong link between counterfactuals and causal reasoning. A very common philosophical view is that saying "C caused E" is tantamount to saying two counterfactuals: "If C had occurred, then E would have occurred" and "If C had not occurred, then E would not have occurred". However, though it is plausible that counterfactual reasoning is uniquely human, it is not clear that the link between causal and counterfactual reasoning would support a more than associative view of human causal cognition. This is because counterfactual reasoning commonly bears on the associated facts rather than on explanation. This does not mean that counterfactuals have nothing to do with explanation: they often represent the idea of a necessary link, which may be the first point of departure between human and nonhuman animals. However, they do not, in and of themselves, constitute an explanation. So we come back to the first question: what is an explanation?

This may be the right place to introduce a distinction advocated by Waldmann (2000, 2001), between predictive and diagnostic learning: while predictive learning goes from cause to effect, diagnostic learning goes from effect to cause. It should be fairly clear that explanation, whatever other feature it might possess, is diagnostic in its direction. However, Waldmann goes farther than this simple distinction, saying that a sheer associationist model (e.g., Rescorla & Wagner 1972) cannot account for the whole of inductive causal — both predictive and diagnostic — learning, because it is basically indifferent to causal asymmetry, being concerned not with causes and effects, but with cues and outcomes, where cues and outcomes can be either causes or effects. Waldman goes on to show that depending on whether the task is a predictive or a diagnostic task, some well-known effects of associative learning (e.g., blocking and overshadowing) do not operate identically. In other words, causal directionality plays a role in causal learning, contrary to what associationist models would predict. In a series of papers, Waldmann and colleagues (Waldmann & Hagmayer 1998, Waldmann & Martignon 1999, Hagmayer and Waldmann 2004) have gone farther, defending a more abstract view of causal cognition, based on a Bayesian network model, arguing that human learners rely on abstract causal categories (e.g., multiple causes, multiple effects, causal chain) to learn causal relations.

This, then, is the first intimation of the fact that ACK does not only rely on association. However, the Bayesian model proposed by Waldmann and his colleagues still strongly relies on covariation of causes and effects. A further question is whether covariation really is the central factor in human causal learning. Dennis & Hahn (2001) examined order effect in causal relations judgment, through presenting subjects with the same covariation data but in different sequences. There was a strong primacy effect, suggesting that covariation is not the whole story behind causal learning and that a belief-updating process may be operative. An indirect support for this view could be found in Lovibond's (2003) experimental work (using a simple fear conditioning paradigm in humans) who has shown that associative models might profitably be reinterpreted as modeling inferential and propositional learning, given that fear conditioning in humans operates indifferently from physical stimuli or from linguistic instruction. Ahn and colleagues (1995) were also responsible for another body of work relevant to the problem, questioning the impact of covariation relative to that of mechanism information in causal attribution. Ahn and colleagues devised a series of tasks in which subjects, to give a causal explanation, could ask either covariational (who-, what-questions) or mechanism (how-questions). They found a strong preference for information on causal mechanisms rather than on covariation in all of their tasks. They went on to point out that explanation is more readily understood as mechanism-based (i.e., general laws) than as covariation-based, noting that mechanism-based explanations are truly explicative in that they are generative, allowing one to make predictions

about new situations in an abstract way. A view that indirectly agrees is that of Eagleman & Holcombe (2002), where the authors discuss Haggard et al. (2002) paper reporting on subjective judgments of the timing of events when the subject reports on the result of one of his/her own action versus an isolated similar event. There was a general reduction in the reports relative to the actual delay for the result of intentionally produced action. Eagleman & Holcombe explain this surprising result via the idea that "events known to be causally related are more likely to be close in time and space than unrelated events" (p. 325). This supports the view that you can deduce association (and, hence, temporal contiguity) from causal mechanism.

To sum up this section, then, both human and non-human animals do rely on association for ACK, but that's not the end of the story for humans, who also use abstract causal schemata and rely on general explanations, rather than staying with simple association. Finally, we are now in a position to give an (informal) definition of explanation: an explanation invokes a general mechanism that accounts for the correlation of a given effect with a given cause.

Is language the answer to why humans are not simply associative animals?

The first question to ask may be if indeed humans are the only animals not satisfied with association. Granted, my dog Tolkien does not go farther than association, but might not more cognitively complex animals, e.g., chimpanzees, actually go farther than association and use, as humans do, though maybe in a more restricted fashion, abstract causal models, as well as look for general explanations possibly based on invisible mechanisms? And if they don't, while we certainly do, what can explain the discrepancy between us and them?

The question of whether chimpanzees look for explanation was explored by Povinelli and Dunphy-Lelii (2001) in two ingenious experiments comparing preschool children and chimpanzees (9;4 to 10;3 year in the first experiment). The task in both cases was simply to stand several blocks on platforms covered with an irregular mat in which some holes with a regular surface had been made. In both experiments one sham block was proposed: in the first experiment, this block couldn't be made to stand because its ends had been beveled; in the second experiment, regular and sham blocks were visually identical, being L-shaped, though weights were placed in either the long or short size, making possible or impossible to stand the block on its long axis. The results were interesting: in the second experiment, where the difference between sham and regular blocks was not visible, 61% of the children investigated the sham block to try and find out why it couldn't be stood up in the wanted position, while chimpanzees didn't. This has led to the unobservability hypothesis (see Vonk & Povinelli in press), according to which "one of the important ways in which humans differ from other species is that our minds form and reason about concepts that refer to unobservable entities or processes" (p. 5). Vonk and Povinelli go on to "suspect that the underlying 'abstractive depth' that makes reasoning about unobservables possible co-evolved with natural language" (idem).

This hypothesis seems to fall foul of a series of experiments made by Varley and her team (see Siegal et al. 2001, Varley & Siegal 2000, Varley et al. 2001) which show that agrammatic aphasics can nevertheless still solve reasoning, causal and theory of mind tasks (and ToM tasks, by definition, involve unobservable entities). Thus, operational language is not mandatory to succeed at such tasks. On the face of it, this seems to contradict Vonk & Povinelli's hypothesis about a link between the ability to conceptualize unobservables and language. However, Varley (1998) has herself observed that her patients had normal linguistic abilities until mit-adulthood, leading her to the conclusion that her "results have nothing to say about the role of language in the development of thinking. It may well be that language is necessary to configure central cognition for certain types of cognitive activity" (p. 145). More troubling might be the fact that pre-linguistic children are supposed to engage in sophisticated reasoning abilities as evidenced in NCK. Even this, however, should be nuanced: false belief test does not seem to be passed before language has set in (see Reboul 2004 for a review) and it is possible that babies' performances at folk physics habituation/dishabituation tests could be explained through more basic abilities than has been supposed, as proposed by Povinelli (2000). Let us in fact suppose that, as claimed by some researchers, NCK develops through time (which, by the way, does not contradict innateness factors). In this case, the apparent contradiction between the unobservability hypothesis and its link with language and prelinguistic or aphasic abstract thought disappears.

What is still mysterious, however, is how and in what way language is linked to the conceptualization for unobservables, that Vonk and Povinelli see as specific to humans by contrast to nonhuman animals. To try and clarify that link, let us go back to what is usually said about the evolution of language. The current opinion is that it evolved for communicative purposes. Apart from the fact that Chomsky thinks that it didn't evolve but just emerged, he has violently attacked the communicative account in a number of papers, of which I will only quote the most recent (Chomsky 2005). At the beginning of this paper, Chomsky quotes a number of eminent biologists (Jacob, Monod, Luria) to the effect that communication would not have produced a great selective pressure to produce language. As both a philosopher and a linguist, I entirely concur with this view, being firmly convinced of the cognitive import of language. However, it is interesting to ask what exactly is meant by language. A current and popular model of language evolution was given by Jackendoff (1994) who saw it as involving a series of steps or stages: From animal communication — protolanguage — to Chomskyan universal grammar (UG). Animal communication differs from proto language in being finite in number of items and in being unable of displacement (the ability to refer to absent or non-existent objects). Protolanguage has a non-finite lexicon and allows for two-words inference but has no function items (if, that, the, where, etc.) or morpho-syntax which distinguishes it from UG. According to most linguists (including Jackendoff), the big evolutionary step is from protolanguage to UG. Chomsky's view is interestingly different: UG — now reduced to very few operations — emerged as a function of complexity, being triggered by the necessity of linking isolated but numerous concepts in a generative (and potentially infinite) way, without any evolutive — in the major adaptive sense — process being involved (it should be noted that this Chomskyan hypothesis receives support through the mathematical models being developed by Nowak and colleagues: see Nowak 2001). If this is right, the major step was the augmentation of the number of available concepts, which, pace Anderson 2004, Maynard Smith & Szathmary 1999, may have been the decisive evolutionary step. In other words and supposing that the protolanguage hypothesis does make sense, the major step would have been going from the closed systems characteristic of nonhuman animal communication to the open systems characteristic of human cognition with their lexical and conceptual open-endedness. It has often been pointed out that displacement does not exist in nonhuman communication system and it can be argued that it is not clearly and uncontroversially present in the so-called talking apes (see Anderson 2004 for a discussion). This, I think, is what takes you from a closed to a truly open-ended system and this, one should emphasize, is what allows one to develop concepts for unobservables, of which it should be noted that they are strongly implicated in NCK, and, to close the loop, the explanations behind ACK frequently make use of NCK.

Conclusion

I've tried to show that causal cognition, though partly common in human and nonhuman animals through the associative basis of ACK, can nevertheless not be reduced to a simple associative process in humans, due to the fact that it involves a need for explanation which is not to be found in nonhuman animals. This major difference between human and nonhuman causal cognition has been explained by the unobservability hypothesis. I've tried in the last part of the paper to sketch an account of how and why the human ability to conceptualize unobservables is intimately linked with the human capacity for language.

Similarités et différences entre la causalité humaine et non humaine

Anne Reboul (r CNRS - Institut des Sciences Cognitives Lyon)
(Date de publication: 28 février 2005)

Résumé:
Hume remarquait que les humains déduisaient la causalité par induction sur la base de trois facteurs: la contiguïté spatio-temporelle de la cause et de l'effet, la contingence de l'effet à la cause et l'antériorité de la cause. Il remarquait aussi que la base inductive de la causalité était l'association de la cause et de l'effet. Les trois facteurs qu'il voyait comme déclenchant la notion d'un lien causal sont très proches de ceux utilisés dans les analyses de l'apprentissage associatif chez les animaux. Cependant, la plupart des chercheurs en cognition animale ne voudraient pas reconnaître que la cognition causale chez les animaux non humains est équivalente à sa contrepartie humaine. Etant donné les similarités, quelles sont les différences et d'où proviennent-elles?

(Traduction de l'original en anglais de Anne Reboul)

Introduction

Que la causalité ait un rôle central dans la cognition humaine ou non-humaine ne peut donner lieu à discussion. Ce qui peut être discuté en revanche c'est si « cause » signifie la même chose dans la cognition humaine et non-humaine. Ou, en termes plus philosophiques, l'attribution d'une croyance causale commune — e.g., « Le fait qu'il pleuve va causer le fait que Maman ne va pas m'emmener faire une promenade » — à mon chien, Tolkien, et à ma fille de 11 ans, Abigaël, fait-elle sens ? Une vision superficielle, basée sur un comportement, *mutatis mutandis*, similaire, ferait penser que oui : Abigaël lit un livre, Tolkien est couché sur un fauteuil et tous deux jettent des regards mélancoliques vers les vitres battues par la pluie. La question, à un niveau plus profond, est de savoir s'il y a plus à la croyance causale d'Abigaël qu'une simple association entre la pluie et une absence de promenade et si, si la réponse à cette première question est positive, on peut légitimement attribuer à la croyance causale de Tolkien ce trait additionnel. Par exemple, Abigaël pourrait avoir une explication mentaliste au terme de laquelle je crois que la pluie mouille et je n'aime pas être mouillée, raison pour laquelle je choisis de rester à l'intérieur quand il pleut. Cette explication, on peut le penser, n'est pas quelque chose que l'on puisse attribuer de façon sensée à Tolkien. Peut-on dire exactement ce qu'est la différence entre la croyance causale attribuée à Abigaël et celle attribuée à Tolkien ? Il me semble que la différence tient au fait qu'Abigaël a une explication pour l'association, alors que Tolkien n'a qu'une association, pour ainsi dire, nue. Qui plus est, Tolkien n'est pas et n'a pas à être intéressé par une explication, alors qu'Abigaël ne serait pas et de fait ne devrait pas être satisfaite par une association nue. Bien qu'il puisse être aventureux de concevoir la différence entre la cognition causale humaine et non-humaine comme l'existence dans la première et l'absence dans la seconde d'une explication, c'est la position que je défendrai ici. J'irai plus loin et je dirai que l'association se fait principalement entre des entités observables, alors que l'explication, le plus souvent, va au-delà de l'observable (comme c'est le cas dans l'explication que donne Abigaël du fait que je ne veuille pas l'emmener faire une promenade sous la pluie). Comme l'a noté Hume (1975, 74. Je traduis), « Tous les événements paraissent entièrement indépendants et séparés. Un événement suit l'autre, mais nous ne pouvons jamais observer aucun lien entre eux. Ils semblent conjoints, mais jamais coordonnés ». Hume déduisait de cette observation de base sur la perceptibilité de l'association et le caractère non perceptible du lien causal l'inexistence du second, mais je ne m'intéresserai pas ici à cette affirmation métaphysique.

Les humains sont-ils des animaux associatifs ?

J'ai affirmé ci-dessus que la cognition causale chez les humains n'est pas ou n'est pas seulement associative. Cette affirmation peut être (et a été) interprétée de façon différente. Tout d'abord, Premack (1995) distingue la *connaissance causale arbitraire* (ci-après CCA) résultant de l'apprentissage associatif — dépendant de la contiguïté et de la répétition — et la *connaissance causale naturelle* (ci-après CCN), fortement spécifique au domaine et *a priori* — indépendante de la contiguïté et de la répétition. Une autre façon d'aborder cette distinction est de dire que CCA est basée sur l'induction, alors que CCN peut servir de base à la déduction. La CCN chez les humains touche la psychologie naïve, la physique naïve et la

biologie naïve. Il est difficile, voire impossible de l'attribuer à des animaux non humains. Dans la mesure où elle n'est pas basée sur l'apprentissage associatif, elle justifie de façon évidente mon affirmation. Cependant, la CCN n'est pas mon objet ici. Je vais plutôt me concentrer sur la CCA. En ce qui la concerne, il y a deux possibilités :

- Elle est basée uniquement sur l'apprentissage associatif chez les animaux humains et non humains ;
- Bien qu'elle soit basée sur l'apprentissage associatif chez les animaux humains et non humains.

Ceci soulève une autre question qui a à voir avec la raison pour laquelle l'association n'est pas suffisante pour la CCA chez les humains. Après tout, si l'association est adaptative pour les animaux non humains, pourquoi ne suffit-elle pas aux humains ? Une autre question évidente concerne ce que c'est qu'une explication.

Selon moi, la réponse à ces deux questions passe par le fait que les humains sont la seule espèce linguistique. Le reste de ce papier sera dédié à une brève revue des travaux expérimentaux qui cherchent à montrer que l'association n'est pas le seul facteur de la CCA humaine et à quelques hypothèses, dont je reconnais le caractère spéculatif, sur le rôle du langage dans la différence entre la CCA humaine et non-humaine.

Pour commencer par un point central de la littérature philosophique contemporaine sur la causalité (maintenant examiné empiriquement, cf. Roese 1994, Roese & Olson 1996, 2003, Pennington & Roese 2003), il y a un lien fort entre les contrefactuelles et le raisonnement causal. Selon une vue philosophique courante, proférer « C a causé E » est équivalent à proférer les deux contrefactuelles « Si C s'était produit, alors E se serait produit » et « si C ne s'était pas produit, alors E ne se serait pas produit ». Cependant, bien qu'il soit plausible que le raisonnement contrefactuel soit spécifiquement humain, il n'est pas clair que le lien entre raisonnement contrefactuel et raisonnement causal justifie une vision de la cognition causale humaine qui y voit davantage que de l'association. En effet, le raisonnement contrefactuel porte généralement sur les faits associés plutôt que sur l'explication. Ceci ne signifie pas que les contrefactuelles n'ont rien à voir avec l'explication : elles représentent souvent l'idée du lien nécessaire qui pourrait être le premier point de différence entre les animaux humains et non-humains. Cependant, elles ne sont pas, en elles-mêmes, une explication. On en revient donc à la question primordiale : qu'est-ce qu'une explication ?

C'est probablement le moment d'introduire une distinction proposée par Waldmann (2000, 2001), entre apprentissage *prédictif* et apprentissage *diagnostique* : alors que le premier va de la cause à l'effet, le second va de l'effet à la cause. Il devrait être clair que l'explication, quelque autre caractéristique qu'elle possède, a la direction de l'apprentissage diagnostique. Cependant, Waldmann va au-delà de cette simple distinction, disant qu'un modèle purement associationiste (e.g., Rescorla & Wagner 1972) ne peut rendre compte de l'ensemble de l'apprentissage causal inductif — à la fois prédictif et diagnostique —, parce qu'il est fondamentalement indifférent à l'asymétrie causale, étant basé non sur des causes et des effets, mais sur des indices et des résultats, les indices et les résultats pouvant être indifféremment des causes ou des effets. Waldmann montre que selon qu'une tâche relève de l'apprentissage prédictif ou de l'apprentissage diagnostique, certains effets bien connus de l'apprentissage associatif (e.g., *blocking* et *overshadowing*) n'opèrent pas de façon identique. En d'autres termes, la direction causale joue un rôle dans l'apprentissage causal, contrairement à ce que prédiraient les modèles associationnistes. Dans une série de papiers, Waldman et ses collègues (Waldmann & Hagmayer 1998, Waldmann & Martignon 1999, Hagmayer & Waldman 2004) ont été plus loin, défendant un vision plus abstraite de cognition causale, basée sur un modèle de réseau bayésien, et montrant que les humains s'appuient sur des catégories causales abstraites (e.g., causes ou effets multiples, chaîne causale) pour apprendre les relations causales.

Ceci est la première indication du fait que la CCA ne s'appuie pas seulement sur l'association. Cependant le modèle bayésien proposé par Waldmann et par ses collègues s'appuie toujours fortement sur la covariation des causes et des effets. Une question ultérieure est de savoir si la covariation est vraiment le facteur central de l'apprentissage causal humain. Dennis & Ahn (2001) ont examiné l'effet d'ordre dans le jugement sur les relations causales en présentant aux sujets des données identiques du point de vue de la covariation, mais dans un ordre différent. Ils ont mis en évidence un fort effet de précédence, ce qui suggère que la covariation n'est pas le seul facteur de l'apprentissage causal et qu'un système de révision des croyances pourrait être un autre facteur. On pourrait trouver une justification indirecte de cette hypothèse dans le travail expérimental de Lovibond (2003) — utilisant un simple paradigme de conditionnement à la peur chez les humains — qui a montré que les modèles associatifs pourraient être réinterprétés de façon profitable comme modélisant un apprentissage propositionnel et inférentiel, étant

donné que le conditionnement à la peur chez les humains fonctionne indifféremment sur des stimuli physiques ou sur des instructions linguistiques. Ahn et ses collègues (1995) ont aussi réalisé d'autres expériences pertinentes sur l'impact de l'information sur la covariation relativement à celui de l'information sur le mécanisme dans l'attribution causale. Ils ont conçu une série de tâches dans lesquelles les sujets, pour donner ne explication causale, pouvait poser soit des questions sur la covariation (*qui, quoi*) ou sur le mécanisme (*comment*). Ils ont trouvé une préférence pour l'information sur les mécanismes causaux plutôt que sur la covariation dans toutes leurs tâches. Ils notent que l'explication est plus facilement comprise comme basée sur le mécanisme (i.e., sur des lois générales) que sur la covariation. Les explications basées sur le mécanisme ont en effet l'avantage d'être réellement explicatives parce qu'elles sont génératives, permettant d'une façon abstraite de faire des prédictions sur des situations nouvelles. Eagleman & Holcombe (2002) s'accordent indirectement avec cette vision des choses dans leur discussion de l'article d' Haggard *et al.* (2002) qui rapporte des jugements temporels subjectifs sur le déclenchement du résultat d'une action selon que le sujet considère le résultat d'une de ses propres actions ou non. Les délais sont jugés plus courts dans le premier cas. L'explication proposée par Eagleman & Holcombe est que « des événements connus comme en relation causale ont plus de chances d'être proches dans le temps et dans l'espace que des événements sans lien entre eux » (p. 235), ce qui justifierait le fait que l'on peut déduire l'association, et donc la contiguïté temporelle, du mécanisme causal.
En bref, les animaux humains et non-humains s'appuient les uns et les autres sur l'association pour la CCA, mais les humains ne s'en arrêtent pas là et utilisent aussi des schémas causaux abstraits et des explications générales. Enfin, on peut donner une définition informelle de l'explication : une explication invoque un mécanisme général qui permet de corréler un effet donné à une cause donnée (les éléments associés).

Le langage est-il la raison pour laquelle les humains ne sont pas des animaux simplement associatifs ?

La première question à poser pourrait être de savoir si les humains sont bien les seuls animaux qui ne se satisfont pas de l'association. Certes, mon chien Tolkien ne dépasse pas l'association, mais des animaux cognitivement plus complexes, e.g., des chimpanzés, ne pourraient-ils pas aller plus loin et utiliser, comme le font les humains, bien que probablement de façon plus limitée, des modèles causaux abstraits et chercher des explications générales, peut-être basées sur des mécanismes invisibles ? Et, s'ils ne le font pas, alors que nous le faisons clairement, comment expliquer cette divergence entre eux et nous ?
La question de savoir si les chimpanzés cherchent des explications a été examinée par Povinelli et Dunphy-Lelii (2001) dans deux expériences ingénieuses comparant des enfants d'âge maternelle et des chimpanzés (de 9,4 à 10,3 ans dans la première expérience). La tâche dans les deux cas étaient de poser des blocs debout sur une plateforme couverte d'un revêtement irrégulier, mais où des trous livrant une surface régulière avaient été pratiqués. Dans les deux expériences, il y avait un bloc trafiqué : dans la première, ce bloc ne pouvait tenir debout parce que ses extrémités avaient été arrondies ; dans la seconde, les blocs normaux et le bloc trafiqué étaient visuellement identiques et en forme de L, mais des poids placés soit dans la grande longueur, soit dans la courte, permettaient ou rendaient impossible de faire tenir le bloc debout sur sa grande longueur. Les résultats sont intéressants : dans la seconde expérience, où la différence entre les blocs normaux et le bloc trafiqué n'était pas visible, 61% des enfants ont examiné le bloc trafiqué pour essayer de comprendre pourquoi on ne pouvait pas le mettre dans la position voulue. Par contraste, aucun chimpanzé ne l'a fait. Ceci conduit à l'*hypothèse de l'inobservabilité* (cf. Vonk & Povinelli, sous presse, 5. Je traduis), selon laquelle « une des différences importantes entre les humains et les autres espèces est que nos esprits forment des concepts qui réfèrent à des entités ou à des processus inobservables et raisonnent sur cette base ». Vonk et Povinelli poursuivent en proposant « que la « profondeur abstractive » sous-jacente qui rend possible le raisonnement sur les inobservables a probablement coévolué avec le langage naturel » (idem).
Cette hypothèse semble contredite par une série d'expériences faites par Varley et son équipe (cf. Siegal *et al.* 2001, Varley & Siegal 2000, Varley *et al.* 2001) qui montrent que des aphasiques agrammatiques peuvent réussir des tâches de raisonnement, de causalité et de théorie de l'esprit (et les tâches de ToM impliquent par définition des inobservables). Ainsi, un langage opérationnel n'est pas nécessaire pour réussir de telles tâches. Ceci semble donc contredire l'hypothèse de Vonk et Povinelli sur un lien entre la capacité de conceptualiser des inobservables et le langage. Cependant, Varley (1998, 45. Je traduis) a elle-même observé que ses patients ont eu des capacités linguistiques normales jusqu'au milieu de l'âge

adulte et en conclut que ses « résultats n'ont rien à dire du rôle du langage dans le développement de la pensée. Il se peut que le langage soit nécessaire pour configurer la cognition centrale pour certains types d'activités cognitives ». Il pourrait être plus troublant que les enfants pré-linguistiques soient supposés capables de raisonnement sophistiqué dans la CCN. Même ceci, cependant, devrait être nuancé : le test de la fausse croyance n'est pas passé avant que le langage n'ait commencé à se développer (cf. Reboul 2004 pour une discussion) et il est possible, comme le propose Povinelli (2000), que les performances des bébés dans les tests de physique naïve puissent s'expliquer par capacités plus basiques que celles qui ont été généralement invoquées. Supposons de fait que, comme l'affirment certains chercheurs, la CCN se développe au cours du temps (ce qui ne contredit pas l'hypothèse de facteurs innés). Dans ce cas, la contradiction apparente entre l'hypothèse de l'inobservabilité et son lien avec le langage d'une part et la pensée abstraite pré-linguistique ou aphasique disparaît.

Ce qui reste mystérieux, c'est la façon dont le langage serait lié à cette conceptualisation des inobservables, dont Vonk et Povinelli pensent qu'elle est spécifique à l'espèce humaine. To essayer de clarifier ce lien, revenons-en à ce qui est généralement dit de l'évolution du langage. L'opinion majoritaire est qu'il a évolué pour la communication. Indépendamment du fait que Chomsky ne pense pas qu'il ait évolué — il aurait simplement émergé —, il a violemment attaqué cette vision communicative dans un grand nombre de papiers, dont je ne mentionnerai que le dernier (Chomsky 2005). Au début de cet article, Chomsky cite un certain nombre de biologistes éminents (Jacob, Monod, Luria) qui doutent que la communication ait pu exercer une pression sélective suffisante pour produire le langage. Comme philosophe et comme linguiste, j'adhère à cette vision des choses, étant convaincue que le langage a un impact cognitif important. Cependant, il est intéressant, dans cette perspective, de se demander ce que l'on entend généralement par « le langage ». Un modèle populaire de l'évolution du langage est celui de Jackendoff (1994) qui y voit une suite d'étapes : communication animale — protolangage — grammaire universelle chomskyenne (UG). La communication animale diffère du protolangage en ce qu'elle comporte un nombre d'items finis et qu'elle est incapable de déplacement (la capacité à référer à des objets absents ou inexistants). Le protolangage a un lexique ouvert et permet des énoncés de deux mots, mais il ne comporte ni items fonctionnels (*si, que, le, où,* etc.) ni morphosyntaxe, ce qui le distingue de UG. La vision de Chomsky diffère de celle-ci de façon intéressante : UG — maintenant réduite à un petit nombre d'opérations — a émergé comme une fonction de la complexité et cette émergence a été déclenchée par la nécessité de lier des concepts isolés, mais nombreux de façon générative (et potentiellement infinie) sans qu'aucun processus évolutif — dans le sens adaptatif fort — ait été impliqué. On notera que cette hypothèse chomskyenne est soutenue par les modèles mathématiques développés par Nowak et ses collègues (cf. Nowak 2001). Si la vision chomskyenne est correcte, l'étape majeure a été l'augmentation du nombre de concepts, qui, *pace* Anderson (2004) et Maynard Smith & Szathmary 1999, pourrait bien avoir été l'étape évolutive décisive. En d'autres termes et en supposant que l'hypothèse du protolangage fasse sens, l'étape majeure aurait été le passage des systèmes clos caractéristiques de la communication animale aux systèmes ouverts caractéristiques de la cognition humaine avec leur absence de limites lexicales ou conceptuelles. On a souvent observé que le déplacement n'existe pas dans les systèmes non-humains de communication et on peut même argumenter pour le fait qu'il n'est pas clairement présent chez les soit-disant primates parlants (cf. Anderson 2004 pour une discussion). C'est précisément, selon moi, ce qui permet de passer d'un système clos à un système authentiquement ouvert et c'est, on le notera, ce qui permet de développer des concepts d'inobservables, dont on remarquera qu'ils sont fortement impliqués dans la CCN. Enfin, pour boucler la boucle, on notera que les explications sous-jacentes à la CCA utilisent fréquemment la CCN.

Conclusion

J'ai essayé de montrer que la cognition causale, bien que en partie commune aux animaux humains et non-humains par la base associative de la CCA, ne peut néanmoins pas être réduite à un simple processus associatif chez les humains, parce qu'elle implique un besoin d'explication qui ne se retrouve pas chez les animaux non-humains. Cette différence majeure entre la cognition causale humaine et non-humaine a été expliquée par l'hypothèse de l'inobservabilité. J'ai essayé dans la dernière partie de l'article d'esquisser une analyse de la façon et des raisons pour lesquelles la capacité humaine à conceptualiser les inobservables est intimement liée à la capacité humaine pour le langage.

References

Ahn, W-K., Kalish, C.W., Medin, D.L. & Gelman, S.A. (1995), "The role of covariation versus mechanism information in causal attribution", in Cognition 54, 299-352.

Anderson, S.R. (2004), Doctor Dolittle's delusion: animals and the uniqueness of human language, New Haven/London, Yale University Press.

Chomsky, N. (2005), "Three factors in language design", in Linguistic Inquiry 36/1, 1-22.

Dennis, M.J. & Ahn, W-K. (2001), "Primacy in causal strength judgments: the effect of initial evidence for generative versus inhibitory relationships", in Memory & Cognition 29/1, 152-164.

Eagleman, D.M. & Holcombe, A.O. (2002), "Causality and the perception of time", in TICS 6/8, 323-325.

Haggard, P., Clark, S. & Kalogeras, J. (2002), "Voluntary action and conscious awareness", in Nature Neuroscience 5/4, 382-385.

Hagmayer, Y. & Waldmann, M.R. (2004), "Seeing the unobservable — inferring the probability and impact of hidden causes", in Proceedings of the 26th animal conference of the Cognitive Science Society, Mahwah, NJ, Erlbaum.

Hume, D. (1975), Enquiries concerning human understanding and concerning the principles of morals, Oxford, Oxford University.

Jackendoff, R. (1994), Patterns in the mind: language and human nature, New York, Basic Books.

Lovibond, P.F. (2003), "Causal beliefs and conditioned responses: retrospective revaluation induced by experience and by instruction", in Journal of experimental psychology: Learning, memory & Cognition 29/1, 97-106.

Maynard Smith, J. & Szathmary, E. (1999), The origins of life: from the birth of life to the origins of language, Oxford, Oxford University Press.

Nowak, M.A. (2001), "Evolution of universal grammar", in Science 291, 114-118.

Pennington, G.L. & Roese, N.J. (2003), "Regulatory focus and temporal distance", in Journal of experimental social psychology 39, 563-576.

Povinelli, D. (2000), Folk Physics for apes, Oxford, Oxford University Press.

Povinelli, D.J. & Dunphy-Lelii, S. (2001), "Do chimpanzees seek explanations? Preliminary comparative investigations", in Canadian journal of experimental psychology 52/2, 93-101.

Premack, D. (1995), "Cause/induced motion: intention/spontaneous motion", in Changeux, J.P. & Chavaillon, J. (eds), The Origins of the human brain, Oxford, Clarendon.

Reboul, A. (2004), "Evolution of language from theory of mind or coevolution of language from theory of mind?", Webconference Issues in the coevolution of language and theory of mind, available at URL: http://www.interdisciplines.org/coevolution/papers/1.

Rescorla, R.A. & Wagner, A.R. (1972), "A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement", in Black, A.H. & Prokasy, W.F. (eds), Classical conditioning II. Current research and theory, New York, Appleton-Century-Crofts.

Roese, N.J. & Olson, J.M. (1996), "Counterfactuals, causal attributions, and the hindsight bias: a conceptual integration", in Journal of Experimental Social Psychology 32, 197-227.

Roese, N.J. & Olson, J.M. (2003), "Counterfactual thinking", in Nadel, L., Chalmers, D., Culicover, P., French, B. & Goldstone, R. (eds): Encyclopedia of cognitive science, New York, Macmillan.

Roese, N.J. (1994), "The functional basis of counterfactual thinking", in Journal of personality and social psychology 66/5, 805-818.

Siegal, M., Varley, R.A. & Want, S. (2001), "Mind over grammar: reasoning in aphasia and development", in TICS 5, 296-301.

Varley, R.A. & Siegal, M. (2000), "Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient", in Current biology 10/12, 723-726.

Varley, R.A. (1998), "Aphasic language, aphasic thought: propositional thought in an apropositional aphasic", in Carruthers, P. & Boucher, J. (eds), Language and thought: interdisciplinary themes, Cambridge, Cambridge University Press.

Varley, R.A., Siegal, M. & Want, S. (2001), "Severe impairment in grammar does not preclude theory of mind", in Neurocase 7, 489-493.

Vonk, J. & Povinelli, D. (in press), "Similarity and difference in the conceptual systems of primates: the unobservability hypothesis", in Zentall, T. & Wasserman, E. (eds), Comparative cognition, available at URL: http://www.cognitiveevolutiongroup.org/site100-01/1001369/docs/preliminary_similarity.pdf.

Waldmann, M.R. (2000), "Competition among causes but not effects in predictive and diagnostic learning", in Journal of experimental psychology: Learning, memory and cognition 26/1, 53-76.

Waldmann, M.R. (2001), “Predictive versus diagnostic causal learning: evidence from an overshadowing paradigm”, in Psychonomic Bulletin & Review 8, 600-608.
Waldmann, M.R. & Hagmayer, Y. (1998), “How categories shape causality”, in Hahn, M. & Stenoss, S.C. (eds), Proceedings of the 21rst annual conference of the Cognitive Science Society, Mahwah, NJ, Erlbaum.
Waldmann, M.R. & Martignon, L. (1999), “A Bayesian network model of causal learning”, in Gernsbasher, M.A. & Derry, S.J. (eds), Proceeding of the 20th annual conference of the Cognitive Science Society, Mahwah, NJ, Erlbaum.

Discussion

Naming, Predicting and Diagnosing, Computing, and Transporting
Robert Stonjek
Feb 28, 2005 21:32 UT

A notable difference between humans and animals is our use of language. Baby humans and chimps are remarkably similar until the child starts to ‘babble’, showing that the language apparatus is warming up ready for deployment.

This ‘warming up’ stage can also be observed in a child’s understanding of causality. Children will make up causal links, often amusingly. These linkages are the equivalent of babble on a whole new level.

Just as a child does not like to let a new object go un-named, even if they have to make up a name for themselves, they do not allow potential causal links to go unmade.

It is clear that chimps don’t make it to the first layer of babble – the second, then, must be entirely inaccessible to them.

Are humans associative animals?

Before any useful debate on the possibility of human-like causal linkages in non-human animals can be seriously contemplated, we must establish just what such linkages achieve in humans. From there we can ask if the same ends are met in non-human animals and then the question becomes “How?”, by the same or some other mechanism as that known to exist in humans?

To better understand the utility of rudimentary causal associations we must look to tribal myths and legends. Let’s take the example of the Australian Aboriginal story of Kapali ~ ‘old possum man’.

This story tells of how the possum acquired its behavioral nature through the story of a cantankerous old man that was eventually transformed or became a possum.

The story simultaneously tells about cantankerous people, that the old can become cantankerous (a warning to both the old who may be becoming cantankerous and to the young who may not know that old people can become like this) about the treatment of such people – how they should be approached, how a cantankerous person may behave, but also tells of the possum’s nature, how to approach possums and what to expect etc.

Such stories are both Predictive and Diagnostic, but have another important feature – the causal chain can be applied to more than one phenomena (is Transportable) and underlying causal connections can be

merged, generalized or independently manipulated to predict outcomes never before observed or to explain phenomena not previously explicable (Computable) .

The tribal story is the early version of this and can guide us to the roots of such casual predisposition in humans. A more modern form of the same thing is arithmetic eg if I have two objects and I acquire three objects I have five objects ~ 2+3=5. The causal form here is entirely abstract and thus may be the epitome of early causal thinking. It can be merged, applied to multiple scenarios simultaneously and can be independently manipulated. Arithmetic is Predictive, Diagnostic, Computable (can be manipulated) and Transportable.

There are a number of advantages to this form of thinking, the main one being the ability to share causal links with others. We may know, personally, that if it rains we won't go for a walk, but this knowledge passed on that form is either accepted or rejected outright. If we include the causal chain then others can be convinced and, further, can add to the causal chain from their own experience. Two or more people can then discuss the causal chain, manipulate it, apply it to other phenomena and so on (Compute it, Transport it, test its capacity for Prediction and Diagnosis).

Combined with the root of language, 'Naming', we have the basis of all human conscious though processes ie Naming, Predicting and Diagnosing, Computing, and Transporting. Importantly, in this form, more than one thinker can contribute to a thought process giving human intellectual capacity enormous leverage.

Kind Regards, Robert Karl Stonjek

A reply to Robert Karl Stonjek
Anne Reboul
Mar 2, 2005 16:46 UT

Thanks to Robert Karl Stonjek for his nice comment. I take it that he is in broad agreement with my paper, which, of course, makes it much more difficult to reply.

I'll just come back to my example of both my daughter and my dog predicting, from the fact that it rains, that I will not take either of them for a walk. In that very simple example, I take it that the predictive power of their respective causal knowledge (of which I claimed that in my dog it was based exclusively on association, while in my daughter, it went farther than association) is very much the same. I would say that it is on their diagnostic power that they differ. Strictly speaking, I don't think that my dog is all that interested in diagnosing the causes of effects, if only because he presumably does not distinguish between cause and effect. On the other hand, my daughter uses in her (diagnostic) explanation of why I won't take her for a walk when it rains not only the observable facts of its raining and the absence of a walk but the further (unobservable) fact of my belief (rain will make me wet) and desires (not getting wet) as well as how I combine them to make my desire come true. This, of course, is still very much bound in the situation, not because, as Robert says, the knowledge that I don't go for walks in the rain can be rejected by other people, but because other people may have other beliefs or desires relative to rain (one of my brothers goes for walks when it rains because he likes walking in the rain). However, my daughter can apply her general knowledge about beliefs, desires and the connection between them (which is precisely what enabled her to diagnose the reason why I don't take her for a walk) to other people. My dog can't.

A final comment: I don't think that the main cognitive import of language is sharing causal knowledge. It is of course extremely important and that's the way science progress. But, before communicating causal knowledge, one has to be able to get it and this is where I think language does make a difference.

A question regarding Predictive vs. Diagnostic
Walter Freeman
Mar 7, 2005 20:09 UT

This distinction is new to me and raises my question whether language is essential for the latter and not for the former. As my example, some years ago I took my dog for a ride on my motorcycle by having her jump onto the gas tank between my legs and arms. At the first turn (left) she fell over onto my right arm. At the next turn (right) she leaned into the turn. This was one-trial learning that I would maintain is Diagnostic cognition, if I understand the word correctly. Days later after a number of rides, when I let her out to pee in the morning, she didn't come back. I found her sitting on my bike in the garage, pre-empting the space that otherwise one of my children would soon occupy. I offer this as an example of Predictive cognition. Further, because dogs can't compute, I suggest that this complements my view, derived from studies of
brain dynamics, that neither form is computational, nor need either be linguistic.

Perceiving cause-effect relations in apes

Josep Call
Mar 2, 2005 12:04 UT

Anne Reboul argued that nonhuman animals detect associations between stimuli whereas humans are also capable of explaining cause-effect relations. Although I do not dispute the second part of Reboul's argument, there is some data on apes (Call, 2004, in press) that, in my opinion, undermines the first part of the argument. Below, I summarize the data.

We contrasted two types of problem with similar superficial cues and reward contingencies but that differed in their causal structure. We presented two opaque cups (1 baited, 1 empty) to the apes. In the causal problem, we shook both cups, which resulted in the baited cup making a noise. In the arbitrary problem, we taped onto both cups but we only produced a noisy tapping onto the baited cup. Thus, in both problems a sound indicated the presence of food, but they differed in the causal structure (i.e., the food caused the noise in the causal but not in the arbitrary problem). Results showed that apes selected the baited cup above chance in the causal but not the arbitrary problem, even though we ran the arbitrary problem after the apes had succeeded in the causal problem. Additional tests indicated that substituting the tapping noise in the arbitrary problem with a tape recording of the shaken food produced identical results.

In another test, we probed further the apes understanding of the relation between movement and noise. In particular, we investigated whether subjects were able to make inferences regarding the location of food. Again, we presented subjects with two cups. This time, however, we only shook the empty cup and lifted the baited one so that no sound was produced from any of the cups. If subjects knew that shaken food produces a noise, they should select the lifted cup because if a shaken cup makes no noise, then the food must be in the other cup. Results indicated that subjects selected the baited cup significantly more often than in the control condition in which we lifted both cups. Additional experiments showed that this result could not be explained as a learned avoidance response to the silent shaken cup since they did not avoid the shaken silent cup when it was paired with a rotated silent cup. Furthermore, a minority of subjects was able to appreciate that certain movements but not others can produce a noise. In this test, we presented a silent shaken cup (i.e., lateral movement) paired with a silent stirred cup (i.e., circular movement). Here, both cups had a movement, but this movement would only produce a sound if the food was inside the shaken, not the stirred cup.

In summary, our results show that subjects perform better in problems that have a causal rather than an arbitrary structure. Furthermore, apes can make inferences regarding the location of food based on the differential movements of cups in the absence of any sound cue. One interpretation is that apes do not simply associate a sound with the presence of food, but they understand that the food is responsible for the sound.

References Call, J. (2004). J. Comp. Psych., 118, 232-241. Call, J. (in press). In S. Hurley & M. Nudds (eds.). Rational animals. Oxford University Press.

Reply to Josep Call
Anne Reboul
Mar 2, 2005 17:14 UT

I've had the ocasion to read Josep's (2004) paper since posting mine on monday and I can only say that I'm very sorry that I hadn't read it before I wrote my paper. It's a fascinating paper on all counts.

I've no doubt, having read it carefully that it shows that nonhuman animals, at least apes, are able of natural causal knowledge (NCK). This, of course, strictly contradicts the first part of my paper: humans are not the only animals in whom causal knowledge goes beyond association and I take it that Josep has uncontroversially demonstrated that. However, it may also be that Josep's experiments partly contradict the second part of my paper (though he doesn't claim that, of course). Indeed, what is not entirely clear to me is whether this non-associative and non-arbitrary causal knowledge really entails looking for an explanation. Explanation, as I defined it in my paper is diagnostic in direction: it goes from the effect to the cause. However, in Josep's experiment, the apes saw the cups being baited, eventhough they couldn't see which specific cup was being baited. Supposing that they had the causal knowledge, this would lead them to predict that shaking a baited cup would make a noise, while this would not be the case for a non-baited cup. Alternatively, of course, they could not have had any previous relevant causal knowledge and could have diagnosed the presence/absence of food from the noise/absence of noise. Only the second case would count as explanation. I'm not quite sure that Josep's experiments can discriminate between these two (predictive and diagnostic) accounts. I take it however that even if successfully completing the task entails looking for an explanation, the specific explanation involved does not rely on unobservables. To sum up, Josep's experiments certainly contradict the first part of my paper and I'm quite ready to accept that some nonhuman animals have access to natural causal knowledge. I'm not sure that the experiments also partly contradict the second part of my paper. I'm reasonably sure that the natural causal knowledge involved does not rely on unobservables. What I'm not sure about though is whether anything like looking for an explanation (as opposed to using causal knowledge predictively) was involved. If it was, then the experiments partly contradict the second part of my paper. The only thing I can say in conclusion is read Josep's paper!

The Difference is Better Software
Eric Baum
Mar 2, 2005 15:52 UT

When my dog discovered in a neighbor's yard the idea of digging under fences, he instantly applied this in my yard and began to form plans such as: whining to be let out, so that he could beeline to the back of the yard (not visible from where he whined) and probe for a weak point, so that he could escape, presumably in search of some unseen goal.

When Heinrich's ravens discovered, after hours, that they could ratchet up a rope attached to their perch to pull up suspended meat, they understood that if startled they should drop the meat, or they would be jerked back at the end of the rope; and I'll bet they would have applied this reasoning to a different desirable object suspended from a different kind of rope in a different cage.

Since Turing we know that the brain processes involved in thinking are equivalent to computations, and it is helpful to understand them in these terms and ask what the computations look like, and how they arise. Presumably, a concept like digging under a fence, or theory of mind, is representable by computer code. Most likely this code is modular, with modules calling other modules in complex ways. A new concept represents a new module added to the code, or at least some alteration of it. Complexity theory and experience tells us that finding meaningful and useful code is a hard problem, requiring extensive computation and search. It is inconceivable that human mental abilities sprung from thin air, rather they must have been built on a program already present in animals.

In What is Thought? (MIT Press, 2004) I proposed a theory of how such a program evolved and acquired meaning through a generalized version of the formal Occam's razor much studied in the computational

learning theory literature over the last 20 years. Much of the code at the level of animal intelligence is in this view essentially programmed into the genome (more precisely, the genome encodes algorithms that interact with sensory data to reliably build executable brain structures encoding meaningful computational modules). The Occam hypothesis holds that such modules arise and acquire meaning in a sufficiently compact program that solves sufficient number of problems presented by the world. Such program can only be so compact and so powerful through code reuse, being composed of modules that exploit underlying structure in the world and recombine in multiple ways to solve new problems presented by the world (old ideas apply to a new fence).

It is then natural to understand words as labels for computational modules. Metaphor then indicates code reuse. Linguistic expressions then allow humans to communicate programs, (more precisely to guide the listener to construct a program).

One can now explain the difference between human and animal reasoning solely through language as a communicative medium. Recall, discovery of meaningful modules is a hard computational problem, involving extensive search. Animals can more or less only engage in discovery of new programs through a single lifetime. Humankind, through our ability to guide listeners to construct programs, has discovered over generations more powerful programming superstructure built on top of the concepts coded in the genome.

This theory is consistent with all data of which I'm aware. For example, data indicate that human theory of mind could be so constructed on top of computational modules present in plovers and apes, a more powerful program perfected over generations and communicated to children through bedtime stories and books.

Eric Baum http://www.whatisthought.com

Reply to Eric Baum: the question is where the better software came from
Anne Reboul
Mar 2, 2005 17:33 UT

I'm not quite sure what Eric's objections are, but I'll do my best to try and answer them nonetheless. I think that I can assure Eric that I agree that thinking is a matter of computation. Where I think we disagree is that he sees (much as Robert Karl Stonjek) the cognitive import of language as being communicational, whereas I see it as being not only communicational. Strictly speaking, however, I think that there could be a meeting ground. I am not claiming in this paper that one has to actually talk to be able of causal reasoning (if anything the short part of the paper which discusses Varley's work should have made this clear). My claim is that there is something in common between the abilities for language (at least some of them) and the abilities for causal reasoning in human beings. This common feature may be the one that is exemplified in human language by displacement and in causal reasoning by reliance on unobservables. By the way, this is where Eric's analogy between digging under a fence and theory of mind breaks down: first of all, though there may be A concept of digging under a fence, I very much doubt that ToM can be seen as involving A concept; what is more, digging under a fence is clearly a concept for an observable, while theory of mind clearly involves concepts for unobservables (e.g., belief). Finally, this does not mean that I think that the use of language to communicate causal or noncausal knowledge is trivial. Clearly, it has played a major role in a major cognitive endeavor, i.e., scientific research, since antiquity. But that does not mean that the cognitive import of language is limited to communicating knowledge: it could also be useful to build knowledge. There is certainly nothing contradictory about defending both claims at the same time. Similarly, saying that language evolved for cognitive rather than communicative reasons in no way contradicts the fact that it is used to communicate.

reply to Anne Reboul's reply
Eric Baum
Mar 3, 2005 16:34 UT

The evidence that animals can't think about things not present is weak IMO and contradicted by much other evidence. My dog seems to think about going under the fence while in the house, where the fence is not visible. He also seems to dream of chasing squirrels, when they are not present. Dolphins and apes taught the rudiments of language have no problem referring to objects not present.

The evidence that people require words for abstract thought, as opposed to for acquiring methods for abstract thought, is also weak. Introspection denies it (there are many quotes of mathemeticians on the subject) and some people have been observed who lost verbal ability temporarily or permanently due to lesions or epilepsy yet can still reason.

Moreover, I don't understand how words themselves could enable thought. To have meaning, the words must summon the modules implementing the computations. But then it is the computational modules that do the actual work. The discovery of speech might have involved discovery of a new method of interface among modules, but after examining this question at some length in What is Thought? I concluded that the data seems to be explainable without postulating such, so the principle of simplicity mitigates against it.

All the empirical data of which I'm aware where animals can not perform mental tasks that humans can are readily understandable by humans having acquired some additional programming, and the hypothesis of speech enabling discovery of this additional programming over time seems reasonable in every case.

My point above was also that the causal concepts animals employ are quite sophisticated, when you think of what's involved in the program. A concept like "digging under a fence" must call many other modules, it must be integrated into the overall program in a powerful way if it is to enable the thought that the dog can access various regions so that he can decide to go into the backyard after acquiring it when he wouldn't before acquiring it, as observed, and if it is applied in new situations in an appropriate way, as observed. Speaking about it as simply "learned by association" seems to understate the complexity of somehow building the code for a new useful computational module, from a single experience, in just the right way that it generalizes correctly. A concept like pulling up meat on a rope may be observable in the sense that you can see if it works, but the ravens who figured it out understood the rope would jerk them back if they flew holding the meat without observation that this would happen. A mental state is observable from its consequences, which isn't so different from the concept of digging under a fence. All of these things are explainable from the Occam picture in a unified way.

A final reiteration
Anne Reboul
Mar 7, 2005 9:23 UT

I get the rather weird impression that Eric Baum is not commenting on either the paper or the answer I wrote. The point about unobservables is not and cannot be answered through such remarks as "My dog seems to think about going under the fence while in the house…". There is a major distinction between thinking about something of which you have had a perception or which you have done and conceiving of something which you a) haven't perceived, b) haven't done. What is more, animals have desires and desires are about things which are not present at the given time (or else, they would be beliefs, not desires). However, desires are related to known things. Unobservables in the sense I borrowed from Povinelli and Vonk are not abstractions from perceived objects to build a category. They are such things as force in naive physics, essence in folk biology and belief in ToM. And though some mental states may be perceived (feelings for instance), beliefs

cannot: you can only infer a belief from a behavior if you already have the concept of a belief. Contrary to what Eric blithely claims, you cannot observe a mental state from its consequence, unless you already have a concept of the mental state in question. Additionally, though it may be that animals are able of diagnostic learning (of which this would be an instance), it is far from being shown that this is the case. Assuming it simply begs the question. What is more, I think that nothing I said goes against the idea that digging under a fence may be linked to other things in the dog's mind: for instance, chasing rabits of which there are more on the other side of the fence or meeting with the neighbor's nice looking bitch. The idea that concepts are holistic in the sense that no concept can exist in isolation is around since Quine and I've certainly not said anything to contradict it. But what has that to do with unobervables? All the passage on people being able to think without words has both been discussed in my paper and my first reply. Reiterating it without answering my arguments is not productive. As said before, I did not say that words were necessary for thought. Nevertheless, it might interest Eric Baum to learn that children with specific language impairment (SLI) very frequently see their nonverbal IQ regress if their difficulties are not solved. This is a final answer to Eric's argument.

the linguistic properties of explanations
Jacques Moeschler
Mar 3, 2005 11:01 UT

One major point of Anne Reboul's paper, as I understand it, is the difference between a mere association in nonhuman causal cognition and the presence of explanation in human causal cognition. I would contribute to the discussion by anticipating a little bit my paper for this workshop: explanation, from a linguistic point of view (that is, its linguistic surface property), is a backwrad inference, not a forward one. In other word, when an utterance explains an other one, the formal order is 1. the utterance to be explained, 2. the utterance explaining. A very trivial example is: "John fell. Bill pushed him". An argument for this analysis is when you want to make the explanation relation explicit with a connective ("because"), the consequence-cause order is preserved ("John fell because Bill pushed him"). Moreover, causal connectives in different languages behave similarly. An last argument is when you inverse the consequence-cause order, you get another interpretation ("Bill phushed John. He fell" or "Bill pushed John, and then he fell") and the explanation interpretation is out. Now the question is the following: if causes preceed their consequences (effects) in the world (causality implies a temporal asymetry), why do languages impose the consequence-cause order in causal and explanarion discourses. One possible answer would say that language allowes more than a strict causality: when utterances are about causes and effects, speakers mean to explain facts, events, states, etc. But this answer is incomplete because there is no linguistic a priori reasons why the presentation of causal relation should follow the consequence-cause order, and because all languages have linguistic strategies (lexical, syntactic) to represent causal relations between events: some lexical items (verbs) have a causal meaning (sink, open, break, kill, stop, etc.) and there are constructions, called causatives, allowing an additional causal meaning ("Mary made the children eat", in French "Marie a fait manger les enfants"). These facts are not new and not surprising from a strict linguistic point of view. But I think they are very interesting from a cognitive one. Indeed, causal reasonings seem to be more efficient from the causes to the effects than from the effects to the causes (Ahn & Nosek 1998), but causal discourses and explanation discourses goes from effects to causes, implying a backward inference.

Reference: Ahn W. & Nosek B. (1998), « Heuristics used in reasoning with multiple causes and effects », in Proceedings of the 20th Annual Conference of the Cognitive Science Society, Mahwah (NJ), Erlbaum, 24-29.

A reply to Jacques Moeschler
Anne Reboul
Mar 7, 2005 9:38 UT

Jacques Moeschler's comment is very welcome in as much as it highlights the fact that the diagnostic direction of explanation is directly to be found in human languages. This does NOT mean that you couldn't have explanation without language, but it is nevertheless striking that diagnostic

learning has not really been evidenced in animals (it may be that Josep Call is very near to having shown it, see his comment, but I don't think that he is quite there yet, though it would be absolutely fascinating if he did succeed in showing it). I just want to use one of Jacques' remark on Ahn & Nosek's paper to go slightly deeper into the matter. What Ahn & Nosek have shown is that causal reasoning is easier for humans in predictive rather than in diagnostic tasks. One of the points frequently made by Daniel Povinelli is that, for instance, in ToM, the specificity of human cognition is not thar it radically departs from the behavior predictions clearly made by, for instance, apes. It is rather that humans add to that predictive ability other abilities, involving unobservables, abilities which I think are clearly diagnostic. At the same time, as rightly pointed out by Walter Freeman, we most of the time live our lives without stoping for deliberation. Indeed, again one of Povinelli's main points is that one can go very far with just predictive systems without unobservables. I'd say, regarding Ahn & Nosek's results that the evidence that diagnostic reasoning is more costly than predictive reasoning agrees with Povinelli's views. We, as most other animal species, are well attuned to associative learning and, as such, predictive reasoning is easy for us. Diagnostic reasoning by contrast is more deliberative, more costly and, possibly, more likely to be conducted explicitly (through language for instance).

Intentionality in Causal Cognition
Walter Freeman
Mar 4, 2005 17:58 UT

Reboul bases her discussion of causal cognition in differences between humans having language and animals with at best ‘protolanguage’. She focuses on the inadequacy of association (a prime mechanism for adaptation by animals) to explain learning in humans; the distinction between ACK and NCK (induction and deduction); and human reasoning from ‘unobservables’ (mental constructs requiring abstraction and generalization).

In her terms deriving from linguistics and philosophy, the capacities of nonhuman animals for cognition seem to be excessively limited. In my experience as an experimental neurobiologist, humanists often underrate the capacities for cognition of brainy animals such as mammals, octopuses lobsters, and prelingual children. These animals cannot explain causality, yet they behave as if they grasp causal situations intuitively, including their own causal roles.

Reboul asks: “After all, if association is adaptive for nonhuman animals, why should it not be enough for humans?” If as appears she concludes that association suffices for nonhuman animals, I disagree. My observations on nonhuman brains show that they abstract and generalize cinematographically many times each second, doing so at the first cortical synapses (Freeman, 2005). Sensory information selects pre-existing neural activity patterns and modifies them, whereupon it is discarded, not ‘processed’. There is no other way brains could survive continuous massive influx and not drown in information. Cortical output following a parsed frame of sensory input corresponds to the class to which cortex assigns accepted input. Assignment is abstraction owing to deletion of background and generalization by signaling class membership rather than features of input. Association is performed on classes, not events.

These perceptual operations were well known to Lashley (1942): “Generalization is one of the primitive basic functions of organized nervous tissue, … almost universal in the activities of the nervous system (p 302),” and to Piaget (1984), who showed that infants evolve through universal patterns in their prelingual sensorimotor phase, when they learn visuomotor coordination through assimilation and then cognitive skills for navigation, homing, searching, categorizing, and memorizing. As examples of declarative vs. procedural memory, where can a squirrel find nuts it hid months before, and a child find a matching pair of socks it believes its mother put away? These low-order cognitive skills are precursors of high-order processes that require language for symbol manipulation. Can we truly describe the latter without first understanding the former?

This evidence blurs the distinctions drawn by Reboul. yet is fully compatible with philosophical traditions of Thomas Aquinas and Nominalists like David Hume. My conviction is that the bedrock of cognition is not language but intentionality prior to consciousness (Núñez and Freeman, 1999). Most intentional behaviors

in human and nonhuman animals flow purposively without interruptions for deliberation. Seen from the engineering context of artificial intelligence in autonomous devices, these are astonishing feats of cognition, extremely difficult to simulate, unlike the relative ease of programming digital computers to play chess.

I propose that our arena be broadened to include the processes of intentionality, abstraction, and generalization as essential for causal cognition in both human and nonhuman animals.

References Lashley KS (1942) The problem of cerebral organization in vision. In: Cattell J (ed.) Biological Symposia VII: 301-322. Núñez R, Freeman WJ (1999) Reclaiming Cognition. Thorverton UK: Imprint Academic. Piaget J (1984) Adaptation and Intelligence. Chicago IL: Univ. Chicago Press. Freeman WJ (2005) Origin, significance, and role of background EEG. Part 3: Neural frame classification. Clinical Neurophysiology, in press.

Reply to Walter Freeman

Anne Reboul
Mar 7, 2005 9:02 UT

I'm grateful to Walter for giving me the oportunity to clarify a few things. First of all, I don't think that association operates on events: it is triggered by the perception of specific events, but it clearly produces abstraction on these events, and I certainly would agree that the end product of associative learning is the construction of relations between classes of events. This is clearly necessary if any prediction is to be possible and no one who has had a look on the literature of associative learning would deny that prediction is exactly what occurs (indeed it is the test that associative learning has taken place). Thus, what I am not claiming is that there are no abstract representations in animals (there clearly are); neither am I claiming that there is no associative learning in humans (there is). What I am claiming however is that causal cognition in humans goes farther and implicates other processes. This, indeed, is made clear by Waldmann's work on blocking in predictive vs. diagnostic associative learning. Blocking is one of the best-described aspect of conditioning. If you condition an animal (a rat, for instance) to associate a given conditional stimulus (CS; e.g. a sound) with a given uncondition stimulus (US; e.g., an electric shock), the animal, when it perceives the CS will freeze (i.e. manifest a fear behavior showing that it anticipates the US). Then, you add another CS, for instance an odor and the animal now perceives the first CS (the sound) and the second CS (the odor) simultaneously and both are followed by the same US. The animal, when tested with the second CS, will not freeze. The same thing happens with humans who do manifest blocking in predictive tasks. However they do not manifest it in diagnostic tasks. Thus, I'm not saying that associative learning does not involve abstraction and generalization. I'm merely saying that in humans causal cognition is not only predictive, but also diagnostic. It may also be diagnostic in animals, but presumably that's not something which might be evidenced through the classical tasks of conditioning, which are designed to evidence prediction. Now, turning to language and unobservables: the unobservables that, following Povinelli and Vonk, I had in mind, have strictly nothing to do with the kind of abstraction and generalization described by Walter (and of which I would certainly not say that they don't occur in animals). It has to do not with building categories from entities that can be perceived, such as CS and US, but with entities that cannot be perceived, such as belief in ToM, force in naive physics and essences in naive biology. Additionally, it should be said that these entities have a strong role to play in diagnostic explanation, but not necessarily in prediction and I certainly (in agreement with Povinelli 2000) think that you can go very far in terms of both behavioral and cognitive sophistication without them. Thus, I certainly agree with Walter when he says that "most intentional behaviors in human and nonhuman animals flow purposively without interruptions for deliberation". This however has probably most to do with prediction and not with diagnosis. Thus the bedrock of cognition is certainly not language: this would be tantamount to refuse cognition to nonhuman animals and prelinguistic children, certainly not a view that I would endorse. This does not mean however that language does not play an important role in human cognition. It does not mean either that human cognition is necessarily better: we tend to overdo causality and see causal relations on extremely flimsy basis. A final point:

I don't think that animals have protolanguage: they have extremy diverse and sophisticated systems of communication, the study of which is among the most fascinating fields in animal studies. Some animals (parrots and apes) have been taught some language amounting to protolanguage.

Is language the prerequisite for NCK?
Giyoo Hatano & Kayoko Inagaki
Mar 8, 2005 11:12 UT

Anne Reboul's argument starts with a sharp contrast between the two living entities, a 11-year-old girl and a dog. Undoubtedly, only the former possesses abilities to use fully syntactic languages, conceptualize and label unobservables, and offer causal explanations. This contrast is suggestive as well as entertaining, but we are a bit afraid that it is misleading when we discuss roles of language and sensorimotor experiences in causal cognition as well as causal differences between human and nonhuman animals.

The well educated girl is likely to offer an explanation like "she wants to stay indoors because she hates getting wet in the rain." However, explanations do not always take this explicit form. Young children often fail to offer explanations but they can make coherent, reasonable and differentiated predications based on a proper causal device in the domain, which probably reveals a basic explanatory schema (Inagaki & Hatano, 2002). As Keil & Wilson (2000, p. 3) put it, "the ability to express explanations explicitly is likely to be an excessively stringent criterion" and also paying attention only to explicit causal explanation excludes the possibility of attributing causal understanding to preverbal infants and non-human animals.

The unobservability hypothesis certainly echoes Tomasello (1999), his assumption of the uniquely human capability to understand the mediating forces responsible for regularities of external events, and probably reflects a shared belief among dominant primate cognition investigators. It seems to us, however, that this capability is often overemphasized by Western researchers who love to draw a sharp dividing line between humans and great apes. We would like to point out, first, humans are not very good at abstracting and explaining unobservable entities or processes. Human species needed thousands of years to find germs as micro-organisms mediating the transmission of diseases. Without accumulated scientific knowledge and education, very few individuals could understand the unobservable process of contagion. Secondly, nonhuman animals, especially those engaging in a complex social life, may have some ability to infer unobservable mental states of conspecifics, if not hidden processes in natural events. For example, not a few examples of deception, which is impossible without inferring others' mental states, have been reported for chimpanzees in the context of obtaining food (e.g., Hirata & Matsuzawa, 2001).

We agree with Reboul as to the importance of language for causal cognition. Language provides us with devices to represent unobservables, and allows us to readily generalize from one situation to others. However, it will not work effectively without sensorimotor foundations. We propose that human infants/toddlers come to grasp "protocausality" when they are serving a causal agent, in other words, their varied actions produce the correspondingly differentiated changes in spatially connected and temporarily following external entities/events. According to Piaget (1954), this occurs at Stage 5 of the development of sensori-motor intelligence (12-18 months of age). As Mandler (2004, p. 101) conjectures, a notion of causal force can be derived from perceptual analysis of the transfer of motion between two objects combined with "bodily experiences of pushing against resistance and being pushed."

A reply to G. Hatano & K. Inagaki
Anne Reboul
Mar 9, 2005 8:26 UT

In their comment, Gyoo and Kayoko make a few points and, strange as it may seem, I'm on the whole in agreement with these points. I'll just go back briefly to them in the order they are presented in the comment. Beginning with the comparison between my daughter and my dog, of course, Gyoo and Kayoko are right: it was just meant to make a contrast in a light way. However, it wasn't meant to say that the existence of verbally explicit explanations is necessary to attribute NCK. Indeed, I suspect that the names "natural" causal knowledge and "arbitrary" causal knowledge are

misleading in as much as they may lead us to think that associative learning can only detect bogus causality (as that common between CS and US in conditioning, where the CS does not "cause" the US). Indeed associative learning by itself will not discriminate between bogus and authentic causality, but that does mean that it does not play any role in causal learning and reasoning. Regarding specifically NCK, I think that children manifest NCK at pretty early age, and certainly before they can give verbally explicit explanation of what their causal reasoning is based on. What this may mean is that their causal reasoning at that stage is mainly of the predictive rather than the diagnostic kind. Neither am I claiming that one could not reason without the ability to speak: I take it that the reverse has been shown by Varley and her team. I'm only saying that some of the abilities which make language possible may be intimately linked with some of the abilities evidenced in NCK, among which the ability to conceive of unobservables in Povinelli's sense. These may well be left intact in patients with aphasia. Note by the way that presumably a fair part of NCK does not need unobservables in that sense, though some part of it does. I entirely agree that humans are not all that great at explaining unobservable processes such as contagion, which as Gyoo and Kayoko say is shown by the relatively slow process of scientific discovery. At the same time, though scientific discovery aims at accounting for authentically causal processes, nevertheless it is far from such domain specific causal knowledge as is evidenced by folk psychology, folk physics and folk biology (which I think was what Premack meant by NCK). And in these fields, we do seem to be both fairly competent and different from other animals. Note that this does not in any way discount the notion of protocausality, resting on agentive experience of sensorimotor action. Why should it? Finally, regarding the ability of great apes (chimpanzees) to deceive, I'm not discounting it lightly: I know that there is strong anecdotal evidence (and I don't use "anecdotal" as a disparaging term) for it. There has even been some experimental confirmation for it (Tomasello et al. 2003), though this has been disputed (Povinelli & Vonk 2003). My position is that right now the evidence is still not clear enough to say whether the examples of deception which have been reported in the literature are entirely convincing or not. If they are, I'm quite ready to change my position, but the debate is still going on. In fact, this is an area where I would bow to the dominant position, not because it is necessarily right, but because I don't have the necessary expertise to decide.

REFERENCES Povinelli, D. & Vonk, J. (2003): Chimpanzee minds: suspiciously human, in TICS 7/4, 157-160. Tomasello, M., Call, J. & Hare, B. (2003): Chimpanzees understand psychological states — the question is which ones and to what extent, in TICS 7/4, 153-156.

A 1rst reply to W. Freeman query on predictive vs diagnostic reasoning

Anne Reboul

Mar 9, 2005 9:18 UT

This is a reply to W. Freeman's query about predictive vs. diagnostic reasoning (to be found under R.K. Stonjek's comment at http://www.interdisciplines.org/causality/papers/1/1/2#_1). To begin with, I'm not sure whether language is necessary for diagnostic causal learning or reasoning. My claim is (more restrictedly) that some common abilities underly both language and the formation of concepts for unobservables in Povinelli's sense). As explanation fairly often uses such concepts for unobservables, this suggest that diagnostic as opposed to predictive reasoning and learning might be more plausible in humans that in nonhumans. However, this is very far from proving that diagnostic reasoning is impossible for non-linguistic creatures, given that not all causal explanation has to invoke unobservables. To return to diagnostic vs. predictive, I'm not sure that I would say that Walter's dog adjusting her position in order not to slide (even after only one trial) is an example of diagnostic learning as such. The reason is that to show uncontroversially that diagnostic reasoning has taken place, it would be necessary to test animals from the effect alone. I think as a matter of fact that this can be done and that a slight modification in Call's experiment (see Call's comment) might do the trick: in Call's original experiment, great apes were shown the baiting of cups (though they couldn't see which cup of the two had been baited), and then were made to choose one of the two cups. In the relevant condition, either the cup with food in it was shaken, producing a noise, or the empty cup was, not producing a noise. Some of the apes were able to identify the baited cup in both conditions, i.e., they would choose respectively the shaken and noisy cup or the cup which had not been shaken and they would do that, just as Walter's dog, during the first trial. This is extremely impressive as it supposes a very sophisticated reasoning in the second (silent) condition, including the use of modus tollens (if p, then q; if not q, then not p; not q, hence not p) and exclusive

inference (either a or b; not a; then b). By the way, Call was able through a series of experiments to discount any simple associative process, i.e., the apes discriminated betwen authentic causality (the cup is noisy because baited) and simple association (the experimenter makes a noise by a cup). I don't think that it quite shows diagnostic reasoning however, because the apes had access to both the causes (baiting and shaking) and the effect (noise or absence of noise). I think however that diagnostic reasoning could be shown provided that the experiment was run with only the effect and one cause (i.e., the apes would not be shown the baiting) and with maybe a slight change such as for instance replacing solid food by liquid (e.g., fruit juice). If the apes choose the right cup in an experiment of that sort, I would be quite ready to admit that they have shown themselves able of diagnostic reasoning. And I should add that, given Call's very impressive results, I would be very astonished if they didn't succeed. That would most certainly show that nonhuman animals are able of diagnostic reasoning, though it should be noted that this specific case does not involve unobservables. (see A 2nd reply … for a conclusion)

A 2nd reply to W. Freeman query on predictive vs. diagnostic
Anne Reboul
Mar 9, 2005 9:19 UT

I would just like, as a conclusion, to discuss one other instance of diagnostic reasoning, where it would be expected but does not occur among animals. In their very detailed survey of vervet behaviour, Cheney & Seyfarth (1990) discuss the use that vervets make of different kinds of signals to "infer" the presence of predators (I'm not sure in what terms Cheney & Seyfarth would actually describe the way vervets use these signals: a purely associative account is presumably possible). Vervets are quite good at associating various auditory signals, from the alarm calls of nonconspecifics (the superb starling) to the mooing of cows (Masai cattle) and even the ringing of the bells hanging from their necks with specific predators (birds of prey and humans). However, there are a number of signs of the approach or presence of predators which vervets do not seem to be able to understand. These are visual signs (the cloud of dust which Masai cattle makes when they're approaching, the carcasses that cheetahs leave in trees, the track of pythons) and this is how Cheney and Seyfarth characterize them, expressing their puzzlement at this unability of vervets. I would like to point out that all of these are effects from which one can deduce the cause (the presence of the predators). One could explain this inability of vervets through the fact that these are not communicative signals (while all the auditory signals mentionned before either are or could be — the bells — taken as communicative) and though they're rather and properly cautious about it, I suspect that this is what Cheney & Seyfarth had in mind. Now, I think that an alternative explanation is possible: deducing from any of these visual signs the presence of this or that predator would be a clear case of diagnostic reasoning (which, by the way, does not involve unobservables). My take on the matter is that vervets are not capable of that kind of reasoning though they are quite able of predictive reasoning. This of course does not show that other species of nonhuman animals are not able of diagnostic reasoning (Call presumably will be able to show that apes are), but I think it would be interesting to know if among primates, excluding apes, other species of monkeys are able of such diagnostic reasoning. And I would be very interested to know whether the ability to detect predators through visual signs of a comparable nature has been reported in great apes in the wild. This would be a nice ethological confirmation of their abilities at diagnostic reasoning, pending the result of Call's investigation.

REFERENCES Cheney, D.L. & Seyfarth, R.M. (1990), How monkeys see the world. Inside the mind of another species, Chicago/london, University of Chicago Press.

Prospective, predictive, retrospective, diagnosis
John Watson
Mar 16, 2005 18:26 UT

I believe Anne Reboul’s attention to prospective/predictive and retrospective/diagnosis in causal analysis (in her text and responses to Walter Freeman) is of particular importance in the pursuit of understanding species differences in causal detection/conception. Her notion that humans (or higher order primates) may be especially attentive to retrospective/diagnostic relations between events is provocative, at least to me. About 40 years ago, I thought I sensed a form of retrospective analysis in Piaget’s (1952) description of the human infant’s enactment of secondary circular reactions (around 3 months). This contrasted from

what I viewed as the prospective conception of operant learning as evidenced in how response-reinforcement contingencies were depreciated in “schedules of reinforcement” (Ferster & Skinner, 1957). Combining these views, I developed a proto-causal model of “contingency perception” that eventually incorporated a formal distinction between prospective and retrospective conditional probability. I found experimental evidence that 4-6 month-olds are sensitive to variation in both, and the relation between, these conditional probabilities in instrumental learning contexts (Watson, 1979). A conceptual point worth noting is that for any causal relation (XàY) the two conditionals (i.e. P(Y/Xt), the probability of Y given time span t following X, and P(X/tY), the probability of X given time span t preceding Y) are free to vary from one another with the only exception being that if one is greater than zero then the other must be greater than zero. This mean that while an animal may have perfect capacity to cause Y with behavior X, it is possible that the event Y may occur frequently for other reasons and, in that case, the retrospective probability, (P(X/tY), will be very low. The animal learning research on “learned helplessness” and “freeloading” would seem to imply that many lower order animals (dogs, fish) are sensitive to this manipulation of the retrospective conditional (Seligman, 1975). From this I would tentatively conclude that “retrospection” in and of itself is not a critical distinction between higher primates and other animals. But the data are not sufficient to draw a strong conclusion.

Logical retrospection, on the other hand, would seem a stronger candidate for the kind of species demarcation Reboul is suggesting (again, whether or not language dependent). The negation of disjunction (A or B, not A, therefore B) used by Call (as noted in preceding discussion) and by me and my colleagues (Watson, Gergeley, Csanyi, Topal, Gacsi, & Sarkozi, 2001) is a retrospective frame (the event “not A” only carries a useful behavioral cue by reference back to the event establishing “A or B”). The cue presumably derives from the animal’s sense of implication (i.e. the capacity to feel the logical entailment), but great care is necessary to avoid false positives from associative experience. In sum, I think Reboul’s analysis highlights the potential value of focusing on species differences in retrospection and a sense of logical implication and possibly the unique combination of these cognitive acts.

Watson, J. S. (1979) Perception of contingency as a determinant of social responsiveness. In E. B. Thoman (Ed.) Origins of the infant’s social Responsiveness. Hillsdale, N.J.: Erlbaum.

Watson, J. S., Gergely, G., Csanyi, V., Topal, J., Gacsi, M., and Sarkozi, Z. (2001) Distinguishing logic from association in the solution of an invisible displacement task by children (Homo sapiens) and Dogs (Canis familiaris): Using negation of disjunction. Journal of Comparative Psychology, 115, 219-226.

Consciousness, Intentionality and Causality

Walter Freeman (Professor of Neuroscience, Berkeley University)
(Date of publication: 14 March 2005)

Abstract: On the basis of brain dynamics I conceive and describe causality as our phenomenological experience of the intention-action-perception-assimilation cycle that underlies all human knowledge, and infer that causality is not a property or law of the material world but a quale and percept derived in the brain.

Introduction

To develop a vocabulary with which to interrelate neuroscience, philosophy and psychology, I summarize three classes of theories relating mind and brain. These complementary classes are materialist/empiricist, idealist/intellectualist, and existentialist/enactionist. I describe examples for each class of its origins, successes and major problems. I emphasize the concepts of "meaning" and “intentionality”, because the main problem that each class of theories faces is best understood through its conception and treatment of meaning. I develop the proposition that ‘Animals’ create and grasp meaning through intentional actions in social contexts to the extent that they can establish the contexts.

Materialist

In materialist/empiricist views minds are flows of matter and energy that carry information. Among ancient Greeks humors were the basic elements of which the world was made. Blood, phlegm, yellow bile, and black bile, were conceived as combinations of air, earth, fire, and water. These components gave the material basis for the function of the body; health and disease depended on their balance, and perceived imbalances prescribed treatment. For example, excess water and insufficient air caused congestion; treatment was bleeding to restore balance, demonstrating cause and effect. Modern chemists modified this alphabet of elements by replacing humors with atoms and combinations into neurochemicals and genes that are conceived to control behaviors of brains and bodies. Diseases are caused by chemical and genetic imbalances that indicate treatments.

These theories have striking successes. Examples are Pavlovian psychology and behaviorism, in which all behavior consists of hierarchies of reflexes. Accordingly pain is treated by cutting pathways or blocking synapses with analgesics to close the “pain gate”. Chemicals serve to control behavior, mood, and emotional state. More problematic are proposals that undesirable behaviors such as violence, obsession, or hoarding might be treated by replacing specific genes.

The intractable problem for this approach is the nature of qualia, the subjective experiences that accompany neural activity. Chronic pain is very difficult to treat in the confines of materialist theories, and gene therapy for psychiatric disorders is only a dream. John Searle raised this problem with his question of how the meaningless firing of neurons can cause qualia, such as pain.

Philosophers often view meaning in this context as the awareness of feelings and emotions that are add-ons to stimuli and responses. Neurobiologists ask how the limbic system can contribute feeling to the neural activity that is generated in the cognitive hemispheres of the cerebral cortex, and how emotions can be attached to the motions controlled by the basal ganglia and cerebellum. The lack of a central place for emotion poses a great obstacle for a materialist foundation for psychiatry.

Idealist

In the idealist or intellectualist views minds are collections of representations and symbols. These views also originated in ancient Greece. In the Platonic world of ideas there was a fixed set of ideal patterns that were imperfectly realized in the humors of the material world. Use of intellect was necessary to comprehend these ideal forms by processing raw sense data. Plato characterized his approach to perception as passive in using the metaphor of the cave, in which light from outside cast shadows, the forms of which we now describe as information. Descartes mathematized this view in explaining brain function through the primacy of thought. Kant elaborated it into innate categories of mind, and concluded

that all that we can know about the real world is filtered through our senses, which distort our apprehension of the world and prevent us from knowing it as it actually is.

This conception is widespread today, as for example in Noam Chomsky's linguistics inferring the existence of deep structure that is wired into brains. The neurobiology is not his concern; he postulates simply that there is innate connectivity common to all languages, the details of which are realized diversely in different societies and cultures from instruction by parents. In my field the theory takes the form of neural networks. McCulloch, a neurobiologist, and Pitts, a logician, conceived neurons operating as binary switches doing Boolean algebra. The action potential is regarded no longer as an energy wave but as a binary digit of information — on-off, yes-no, one-zero — that carries information for processing according to rules deduced by neurobiologists.

This model was an inspiration for John Von Neumann to invent programmable digital computers. Information flowing from sources to sinks fills an ecological niche by replacing energy and matter as the vector for the flow of ideas. Crucial is the divorce of meaning from information, explicitly in Shannon-Weaver information theory. Engineers are unconcerned about contents of telephone messages, only about their bandwidth and the number of bits required to transmit them. The degree to which uncertainty is reduced gives meaning to the information.

This approach has the “hard problem” of David Chalmers, in which the attempt to realize artificial intelligence as an alternative to brains has raised the question: What is the nature of consciousness? Cognitivists debate the “mystery of consciousness” as Searle calls it: how it is experienced, how it is caused by brains, if at all, and how it causes changes in behavior, not only of the firing in networks of neurons, but of the movements of the body. In its aspectual dualist form, this approach led Whitehead to “pan-experientialism”, by which there are no particles, only events; each particle hasa material aspect and an experiential aspect. Consciousness precedes intent and is conceived as the agent of intention.

The unsolved problem for the cognitive scientists is that in minds, being collections of representations, tokens, or symbols, meaning is defined as the relationship among symbols. This is circular, because when one wants the meaning of a word, one looks for it in the dictionary, which gives more words. Reference to the real world is not defined within or by a computer or a cognitivist brain. Stevan Harnad calls the problem of how to attach meaning to representations the “symbol grounding” problem.

Computational neurobiologists typically suppose that representations are embodied in the firings of neurons, each action potential symbolizing a feature, object or person like a grandmother; that these symbols are operated upon in the cortex according to neurobiological rules; and that the output is passed through the amygdaloid nucleus, where a meaning or an emotion is attached before the message is sent into the motor system. The symbol grounding problem is unsolved in neurobiology and artificial intelligence and is perhaps, along with the “hard problem”, unsolvable.

Existential

I have called the third class of views “existential” in the terminology of Maurice Merleau-Ponty. In his view mind is the structure of behavior. The mind creates behavior and behavior creates mind. This approach has multiple origins, first among them being Aristotle, who championed a view of perception as active in contrast to the Platonic view of passive intake of forms (information). Aristotelian perception required action — probing, cutting, burning, etc. Chief among his successors was St. Thomas Aquinas, who proposed that each person is a unified being using the body to probe the world. Such action expressed 'intent', from the Latin word 'intendere', which meant 'stretching forth' and shaping the self from the consequences of the action by assimilation (“adequatio”) followed by accommodation. He made no use of representation or consciousness; those are modern ideas. It is important to recognize that the Thomist interpretation of “intention” differs markedly from that of Anglo-American analytic philosophers, who follow Brentano in re-defining intention as the “aboutness” of explicit representations – a belief or thought is “about” something in the world.

The idea of the importance of action for understanding was developed in great detail by Merleau-Ponty, relying on medical literature from World War I, and by Jean Piaget and inter alia the school of pragmatism.

A focus for Merleau-Ponty was the dissociation that brain injury revealed between the abilities to grasp and to point, by which he developed his concept of “the tendency toward a maximum grip”. This was also a concern for Piaget in his somato-motor phase preceding acquisition of higher order cognitive skills. These different abilities appear at different stages of development. Children are born with the grasp reflex; their ability to manipulate objects appears in their first few months; the ability to point follows several months later. The remarkable feature of pointing is that this is a social action; the child points with the intent of communicating a target of interest primarily to its mother. The infant looks to its mother to see if she is looking where it is pointing. Grasping merely orients the body to an object in the outside world intentionally without necessary relation to another intentional being. These operations are deep in brain organization. In the existential views meanings exist in social relations that are established by intentional action.

A related but differing view was developed by Gestalt psychologists, principally Wolfgang Köhler and Kurt Koffka, dealing not with brains but with fields of force that mediated between objects and egos. One of the critical differences between the existential view and the Gestaltist view is the passive role of the perceiver, the ego, for Gestaltists. For example, Gibson (1979) conceived of information contained in an object in the form of an affordance, which was what an observer could do with it. “The object offers what it does because it is what it is (p. 139)." Meaning was carried by the information from the object as it flowed through the senses into the nervous system, where it was extracted by resonances that caused appropriate choices. Despite its emphases on action and reciprocity, Gestaltism is idealist and ultimately passive. Affordance does not constitute expectancy.

The existential views have been marginalized largely by their central problem: what causes intentional action? According to both the materialist and intellectualist views, actions are determined by stimuli. A laboratory rat is trained to sit inert until a stimulus comes, and then to do something. Its behavior is stimulus-dependent. A computer terminal waits for keystrokes before computing. Wild animals and people aren't like that. They continually engage the environment, reaching into it purposively. How do those endogenous actions arise? How do neurons cause intentional actions that don't require stimuli to initiate them? For the materialist, intent implies teleology, which is inadmissible; for the cognitivist, intent implies aboutness, which is opaque; for the enactionist, intent is a mystery. For this neurobiologist intent is the doorway into neurodynamics. Brains are self-organizing systems that are future-oriented. The biological role of brains is to construct expectations of future states and schemata of actions by which those future states may be brought about and evaluated. What causes the expectations?

Causality

We have a choice among meanings for the verb to "cause": (i) to make, move and modulate (an agency in linear causality); (ii) to explain, rationalize and blame (cognition in circular causality without agency but with top-down-bottom-up interaction); or (iii) to flow in parallel as a meaningful experience or by-product (Humean invariant relation). The most vexing problem is “agency”, leading to a variety of synonyms for causes: dispositions (Aquinas), regularities (Hume), tendencies (J S Mill), propensities (Popper), capacities (Cartwright), explanations (Donaldson), risk factors (statisticians), etc., in order to distinguish them from mere correlations and coincidences.

It is easy to conceive of how a reflex response is caused by a conditioned stimulus, but it is not easy to explain the causes of exploratory behavior. Some psychologists say that a rat has a curiosity drive or a motivation for novelty that provides the cause, but these are only words. “Intent” is what the animal did or will do; “motive” is the reason we infer from what we observe and experience by introspection. Brains have masses of interconnected neurons that generate the evolving patterns of behavior we observe as sequences of ordered, intelligible states. We perceive the behavior as goal-directed. This animal is looking for something. That animal is running away from something. So we call the actions intentional, what neurologists call “voluntary” by invoking consciousness. Materialists assert that brains cause behavior without explaining intent; idealists claim that consciousness causes behavior without explaining consciousness; Merleau-Ponty opposed both views for what he called their “linear causality”.

Linear causality is universal determinism in causal chains such as, A causes B causes C. A stimulus excites sensory receptors; they excite the sensory cortex; that cortex transmits to the motor cortex; that

activates motor neurons; muscles contract. That is a sequence of linear events; at the core of cognitivist, computationalist, and materialist is linear causality. Instead, Merleau-Ponty introduced circular causality. Every action and every sensation is both a cause and an effect in what he called the "intentional arc". It could not be reduced to linear causation but rather was inherently circular. Consciousness was tertiary, a judgmental act late in this action-perception cycle, and certainly not the ego-driver of action.

This is the deep problem for existential theories: What is the origin of the structure in behavior, which is, according to Merleau-Ponty, mind? In these approaches meaning is created by the interactions of intentional beings with each other through the world. It exists in the engagements, not merely or only in the brains of the participants. It is ontologically in the world and is epistemologically grasped through introspection by inferences from observing goal-directed actions of others' and one's own. Recent advances in brain imaging have made it possible to observe the brain activity patterns of humans and other nonhuman animals who are engaged in intentional behaviors by which meaning is created.

Neurodynamics

Concomitant advances in the engineering and physical sciences in nonlinear dynamics have major import for solving the mind-brain problem, as it is understood through existential theories. My starting point is the Thomist postulate of the self as an entity that distinguishes itself from all aspects of the world outside itself. The self according to Tomaso has unity and integrity of action that precludes mutual interpenetration with the environment. A modern analogy would be the immunological self that disallows exchanges among bodies of foreign constructions such as organ transplants and requires that all foodstuffs be reduced to simple molecules before absorption in the gut. Comparably, all sensory input is disintegrated into the action potentials of individual sensory receptors, leaving the spatiotemporal structures of stimuli behind in the world.

Brains self-organize their internal states to create assimilations of those spatiotemporal structures. More importantly, survival requires predictions of future states in the search for fuel, shelter, and companionship. Images of future states emerge from memories of past states and prior actions that succeeded or failed to achieve those pleasurable and avoiding those painful. An important consequence now verified by brain imaging is that, during perception, brains create expectations about the world in the form of landscapes of chaotic attractors, each corresponding to a class of stimuli. Each attractor is surrounded by a basin of attraction. Brains command actions to test their expectations by importing information through the senses in the form of stimuli. They use a stimulus to select a basin of attraction and update it in the process of assimilation. Convergence to the attractor deletes extraneous information in the stimulus by the process of abstraction. The neural activity pattern configured by the attractor gives the class to which the stimulus belongs in the process of generalization. The residual information serves to modify the attractor in the process of accommodation. Thus brains do not "process information" by storage, retrieval, and pattern matching. They use it for selectivity and then discard it.

The unity of action is not just momentary but spans the whole life of the self by creating knowledge, storing it by synaptic modification, and continually up-dating it with each new sensory impact. As an example, an individual H.M. lost this capacity through a surgical operation intended for relief of epilepsy: removal bilaterally of large parts of his temporal lobes, including the hippocampus. H.M. retains most experience from before the operation but cannot add anything new. He lives eternally in the present. He has lost the ability to construct his life story. Another example of the property of wholeness is revealed in the field of surgery, where by long tradition healing is described as by first or second intention. First intention is formation of a clean scar; second intention is by infection with scar formation. Healing expresses the biology of intention in re-establishing the integrity of the body after injury. The concept can also apply to the healing of psychic wounds of the self.

It is critical for my view and my interpretation of Merleau-Ponty that intention does not require consciousness and most often proceeds without it. This poses a difficulty for phenomenologists, whose 'raw materials' have the form of qualia, and whose 'tools' are provided by awareness through the languages of literature, sport, art and music. My reliance on neurodynamics enables me to conceive how self-organizing brain operations can explain processes described by Merleau-Ponty of humans learning new skills through language and logic, with reversion to smooth, thought-free execution after sufficient

practice.

I conceive the emergence and execution of an intentional act as follows. A self-organized global pattern of synchronized neural activity emerges, linking by interactions masses of neurons in various parts of the forebrain that generate an expectation of a future state. This dynamical structure evolves into schemata of motor commands in specific modules that are delivered to the motor systems, with preafference copies sent to the sensory systems. There they guide the construction of expectations of the changes in sensory inputs that the forthcoming actions will entail. Feedback through the cerebellum confirms that the motor commands will soon be executed, and the proprioceptive sensors in joints and muscles feed back confirmation that the intended actions are taking place. This feedback on preparing and executing an action comes later to awareness in the form of a "cause".

The sensory consequences of the action enter chains of neurons and ascend to cortical modules, whence still later they are assembled in the limbic system and disseminated through the forebrain. This closes the intentional arc by setting the stage for the next frame in a cinematographic process (Freeman, 2005). The up-date comes to awareness in the form of an "effect". The cause-effect relation of these frames is linear, and the sequence is invariant. We act; we perceive. That is the essence of cause-effect, which is the foundation of all knowledge and social organization. All we can know is what we can predict, and what the results of our tests are. Consciousness is a late judgmental process, not an agency.

Conclusion

David Hume with legendary unassailability challenged the notion that causality is anything more than a feeling of certainty based on constancy of contiguity and temporal order. Elsewhere I have marshaled my own evidence that an awareness of cause is the quale that accompanies some cognitive constructs (Freeman, 1999). My aim here is to offer the explanation that people search for the causes of things because that is the way their brains work. As Hilary Putnam (1990) wrote: "Perhaps the notion of causality is so primitive that the very notion of observation presupposes it? "(p. 75) Why do some philosophers insist on asking how consciousness causes an object like a brain or a collection of neurons to move the body? This is like asking, what causes the sun to move across the sky? Animists conceived a chariot with horses. Newton created a larger framework in which the agency was the pull of gravity. Einstein created a still larger explanatory framework, in which a calculus of relations between matter and observer replaced agency.

In my view we can create a new framework for causal cognition by combining nonlinear neurodynamics with Thomist intentionality, within which causality can be redefined. I believe that the sense of causation is an ethical necessity for assignment of responsibility and for defending the belief in one's own power to make choices. We educate children and train animals by nourishing their feelings of guilt, in order that they perceive their actions as causes, so that we and they can control them. But the assignments of causation to objects like neurons, brains, and billiard balls, or to concepts like consciousness and related unobservables, appear to me as category errors. I propose this conclusion as a challenge in understanding causal cognition and its larger role in epistemology.

Acknowledgment

This essay is condensed from my notes for a course on "Intentionality in Philosophy, Neurobiology and Cognitive Science" given in the 1990s by myself and Hubert Dreyfus, Professor of Philosophy, University of California at Berkeley.

References

Dreyfus HL (2002) Intelligence without representation — Merleau-Ponty's critique of mental representation. Phenomenology and the Cognitive Sciences 1: 367-383. Here is revealing commentary on the views of Merleau-Ponty.
Freeman WJ (1999) How Brains Make Up Their Minds. London UK: Weidenfeld & Nicolson. Here is a concise introduction to nonlinear brain dynamics.
Freeman WJ (1999) Consciousness, intentionality, and causality. Journal of Consciousness Studies 6:

143-172. Here is a complete list of my sources and references.
Freeman WJ (2005) Origin, structure, and role of background EEG activity. Part 3. Neural frame classification. Clinical Neurophysiology, in press.
Gibson JJ (1979) The Ecological Approach to Visual Perception. Boston: Houghton Mifflin. Information-based description of interactions with environments.
Putnam H (1990) Realism With a Human Face. Cambridge MA: Harvard U.P. The father of functionalism has had changes of heart, similar to those of Jerry Fodor, yet both are still groping in the wilderness.

Discussion

The Utilitarian Genesis of the Causation Issue
Robert Stonjek
Mar 14, 2005 21:17 UT

Regardless of what the causal chain is or even if there really is one, the utility of a belief in causation and the utility of the associated quale can be found in the evolutionary development of human beings, particularly as their brain became large enough and/or sophisticated enough to contemplate such things as causation.

Taking a vastly reduced subjective world view (perhaps artificially so, to illustrate this point) an early human may believe that everything in nature is caused by some God or nature spirit, even the actions of others. Thus there are only two causal agents in this simplified world view – the individual and God.

As the outside causal agent can not be controlled, only those events directly cause by the self require attention and the taking of responsibility for those events. When the quale of personal causation accompanies an observation, the individual's mind is stimulated to take responsibility for the outcome ie if the outcome is not as intended then the individual must 'think' of how to change it.

It is this 'thinking' which is stimulated by the knowledge of personal responsibility. Identifying causation, then, from the earliest times was essential in applying frugal intellectual resources only to those tasks that could be influenced by a particular individual.

This notion persisted into modern times where some religionists still feel that we should not try to influence "God's work" or go against "God's will", more so among primitive people.

Thus the seeking of causation and accompanying quale had its utilitarian genesis in the need to parsimoniously assign intellectual resources only to those events that can be reliably influenced by such contemplation and subsequent action. Kind Regards, Robert Karl Stonjek.

> **Prime Mover and Ultimate Knowledge**
> Walter Freeman
> Mar 20, 2005 3:43 UT
>
> I agree with Robert that the belief in causation and the accompanying qualia of the experiences of responsibility and guilt have undoubtedly been essential for the evolution of human societies over the past 2 to 3 million years. No one would write a book or run for public office or undertake the arduous educational preparation for a profession without the belief that through personal action one can make a difference. However, the tradition of attribution to the world of a First Cause or Prime Mover, which is alive and well in the Big Bang, is an extrapolation from personal experience to universal law. Were it not for the numerous instances in my experimental career when I have been led significantly astray by causal inference, I would let sleeping dogs lie. My attempts to distinguish between the utility of causation in personal and social action versus the useless spinning of wheels in the tiresome debates about free will and determinism are inspired by my vision of a higher order of discursive framework, in which unanswered and possibly unanswerable questions can be

bypassed in search of relational calculi by which we can all assimilate more closely to the ultimately unknowable “Ding-an-Sich”. This effort is not only compatible with Eric’s postulate, it requires it, as the only truly competent method for devising such an epistemological framework.

Computation, Consciousness, Intentionality and Causality
Eric Baum
Mar 15, 2005 14:54 UT

Turing gave compelling arguments that any possible brain process, and thus presumably any possible thought, corresponds precisely to a particular computation, thus computation is a language capable of describing thought formally. Thus there is no need to talk separately about flows of energy or about neurodynamics or even about behavior. One can talk in a more precise fashion about any of these things by talking about appropriate computations.

At first blush, this unification seems empty. Since everything is a computation, merely saying these things are computations seems to tell us little. But if we can find natural laws of computation, we can then understand thought as well and in the same way as we understand physics and chemistry. One relevant such natural law that has been conjectured is P!=NP; and my book What is Thought? conjectures (based on results in computational learning theory) another, a generalized Occam's razor: a sufficiently compact program that behaves well enough interacting with its environment will generalize to behave well in new situations and to solve new problems not previously seen.

In particular the puzzles you pose: eg how meaning arises from syntax ("the symbol grounding problem") and "the mystery of consciousness", can then be answered in computational terms. Meaning arises through Occam's razor because evolution coded into the compact genome a program that is constrained precisely so that its computations have meaning, i.e. that exploit the underlying structure of the world in an appropriate way. And, similarly, one can see that evolution would naturally build a program having all the aspects of consciousness and qualia. That is, the execution of the program can be seen to include computations corresponding in a natural way to qualia, sense of awareness, of self, and so on.

Meaning then is computational exploitation of the underlying structure of the world. This structure includes causality. If the world weren't causal, it would presumably be much less predictable. Our brains predict by executing code mostly discovered by evolution that exploits the underlying causal structure of the world. Our causal reasoning, and our understanding of causality, is execution of this code.

Walter, we have common ground here. Eg you also emphasize behavior, and your discussion of strange attractors is a way of talking about certain computations. I think, however, it is possible to speak more precisely and say more by speaking in computational terms. Eric Baum http://www.whatisthought.com

Computational Simulations are not the Ding-an-Sich
Walter Freeman
Mar 20, 2005 3:31 UT

Let me begin my response to Eric’s comments by citing two recent books that present irreconcilable views: Steven Wolfram’s “The New Science” and Jerry Fodor’s “That’s Not How Brains Work”. Wolfram’s views strongly agree with Eric’s regarding the universality of computation as the basis for the dynamics of the world and brain and for the construction of meaning. Fodor expresses dismay at the degree to which his work on sensory functionalism has been generalized by others to model cognitive brain function. I might paraphrase Fodor’s position by analogizing brain function to “Darkest Africa” as conceived by 19th century Europeans. Computational understanding is comparable to a fringe of civilized settlements on the coasts of the continent, while the interior remains unknown, untapped, untamed, and noncomputational. Perhaps I misinterpret Fodor’s writing, because I find his views are so congenial with my own.

I am in complete agreement with Eric that models of brain function are not only best when they are computational; they can only be computational. However, the isomorphism of the models with the systems that they describe does not support the inference that brains work the same way. A simple but conclusive example is the function of the eye. The eye refracts light. We model that function by computing Fourier transforms. Success in modeling does not imply that the eye computes Fourier transforms. There are no numbers in eyes nor, for that matter, in brains.

The best thing about Eric's "computational exploitation" is that his models are acausal; time can run whichever way he wants it to. But its Achilles Heel is its inability to compute with real numbers, which must be truncated to rational numbers. No matter how many bits in word size, sooner or later the simulation of chaotic dynamics crashes owing to numerical instabilities from attractor crowding.

causality as a quale
Anne Reboul
Mar 17, 2005 13:03 UT

Walter Freeman points out that Hume reduced causality to "a feeling of certainty based on constancy of contiguity and temporal order", which he reinterprets in more modern terms as awareness of causality being "the quale that accompanies some cognitive construct". I suspect that Walter is, so to say, an atheist regarding causality, while I tend to be an agnostic. His account of the acquisition by young children and animals of causal cognition through education and training aiming at forstering responsability and moral sense is quite credible. There are however two questions I would like to ask: the first one concerns baby before they can move by themselves; the second concerns animals. There has been a wealth of experimental papers — exploiting mostly the habituation-dishabituation paradigm — by psychologists since the 80s, investigating naive physics in babies, which have shown that babies, before they can move by themselves, have expectations about the way material objects move and, more generally, interact. These expectations can be, indeed have been, described in causal terms. Though there is a developmental aspect — some "impossible" events will be accepted as normal at a younger age but perceived as astonishing later on — some aspects of naive physics seem operative at a very early age (around 3 months) before any agentive explanation through personal experience is tenable. What is more, there does not seem to be anything like personal responsability involved. So the question is: should one consider these experiments as evidencing something other than causal knowledge, albeit primitive, of the physical world, and if so, what exactly would be involved? Additionally, if it is causal cognition, how does it fit in Walter's account of the acquisition of causality through training or learning? From what I understood, Walter's claim is that causal cognition, initially developped in agency to foster personal responsability and ethics (i.e. in the intentional domain), tends to be extended to non intentional, physical domains. However, if causality is evidenced in infants' performance in the naive physics experiments, the chronology seems the reverse. Regarding animals, I think that Walter is right that we train them for guilt (that's pretty obvious with dogs and even with domestic cats), though whether we only interpret their behavior as showing guilt or whether they have a similar quale to our own is debatable. But, if awareness of causality is a quale, would that show that dogs have a quale fot causality? I'm not sure it is, given that surely our guilt qualia are not causality qualia: for one thing, we can be not only guilty but proud when we cause something positive; for another, awareness of causality, even when it is misguided, is not limited to events in which we are agents. So does Walter mean that, in domestic animals, there is a quale for causality, which, as it is in humans, is independent from and indeed different from guilt or pride, and that this quale extends to the nonintentional domain? If this is what Walter means, what kind of empirical evidence could show it to be the case?

Causality is a Concept, not a Quale
Walter Freeman
Mar 20, 2005 3:29 UT

Anne's penetrating questions show me that my writing was not as clear as it should have been. First, no, I am not atheistic regarding causality nor agnostic like herself but nominalist like Hume, and I regard those attributing causation to inanimate objects as deists and animists. She asks whether the experiments in naïve physics show that infants learn physical causation before

personal responsibility, because she thinks I believe that infants learn causality by training. My view is that causality is learned not by training but by trial-and-error manipulation of limbs in the first few months of life, increasingly under visual guidance. By three months of early experience with somatomotor control infants could uncritically use that experience to interpret manipulanda provided by experimenters. Regarding the development of a sense of personal responsibility, this could only come after the infant had begun to distinguish itself from not-self, a complex process the timing of which is not explicitly stated in her allusive description of the experiments. I am not sufficiently experienced in the field of child development to handle this topic.

Anne prefaces further questions regarding quale with the conditional statement, "… if awareness of causality is a quale…", which appears to attribute that belief to me. On the contrary, I don't see causality as quale, but as cognitive attribution of agency. The quale is the 'aha!' feeling of certainty that accompanies a conclusion that one has found a necessary 'causal connection' and not merely a fortuitous correlation. Finally, I share Thomas Nagel's view on the inaccessibility for humans of qualia in animals. While I'm sure some nonhuman social species also experience qualia of guilt, shame, joy and pride, I cannot know, and in any case would definitely classify causality as a cognitive abstraction that is inaccessible to nonhuman animals and prelingual children.

Quale of causal logic?
John Watson
Mar 25, 2005 20:28 UT

Walter has presented a complex philosophical position that raises many interesting questions. I find some of his proposed answers to these questions, if I understand them correctly, quite compatible with my own biases (e.g. the possibility of intentional/purposive action without consciousness, the neural basis of intentionality, and as I will argue in my paper, the idea that humans at least may be endowed with a determinist stance that when combined with causal logic provides a basis for appreciating hidden causes such as intentions in others).

I must admit that I misread Walter in the same manner as did Anne regarding the notion that there is a special quale to causal perception/detection. I find, however, that Walter's clarification to the effect that the quale "is the 'aha!' feeling of CERTAINTY that accompanies a CONCLUSION that one has found a NECESSARY 'causal connection'…"(emphasis added) raises some need for additional clarification, at least for me. While I initially (erroneously) thought Walter's conception was like Micheott's (1963) notion of perceived 'ampliation' in causal collisions (even though the discussion of roots in infant motor action sounded more like Piaget), I now believe Walter is binding his conception of causal conception/perception to the individual's capacity to sense logical implication. Lewis Carroll's (1895) infamous Tortoise who failed to feel logical implication, and could not be logically forced to do so, would thus, if I've got it right this time, experience life as if in a non causal world. But would he not see (or feel) Michottean 'ampliation'? And, more importantly for the coherence of Walter's position (if I understand it), how does one reconcile the position that perception of causality is anchored in young infant's "trial and error manipulation of limbs" and yet hold that causality is "cognitive abstraction that is inaccessible to nonhuman animals and prelingual children?"

Carroll, L. (1895) What the Tortoise said to Achilles. Mind, 4, 278-280.

Michotte, A. (1963). The perception of causality. London: Methuen.

Distinctions of perceiving vs. conceiving causality
Walter Freeman
Apr 6, 2005 15:30 UT

Thank you, John, for these insightful comments and questions. My intent is to separate use of the concept of causality into two domains: on one side the physical, neural, and social sciences, on the other side the ethical and legal domains of human conduct. None of us has difficulty in assigning responsibility in such statements as "Hitler caused the downfall of Germany", etc. We have great

difficulty in deciding how consciousness causes neurons to fire, etc. Consider a simple distinction. When I served as an intern in Pathology, I was taught not to write that a patient died “of” a condition but “with” it. We had no scientific or medical proof of cause of death. Assignment of cause was the legal responsibility of the coroner or his delegates. Causality comes into play when humans call upon themselves to take action to change the world for better or worse. My experience is that when I attribute causation to nonhuman (or better said, nonintentional) entities, in science I open myself to the high probability of making mistakes.

I also want to separate the logical and epistemological understanding of causality by philosophers from the intuitive grasp of causal relations experienced by all of us before we could talk about them. A child intuitively grasps the control of its limbs through practice in trial and error and perceives the relations between its intents and the consequences of its actions. It does not perceive (in your words) causality. That is what we adults perceive, though I would prefer to say that we conceive causality.

Thank you for this opportunity for further clarification.

Causal logic and the intentional stance

John Watson (Professor of Psychology, University of California, Berkeley)

(Date of publication: 11 April 2005)

Abstract: Dennett introduced the concept of the 'intentional stance' as a strategy used by humans in their coping with the behavior of each other and with the behavior of complex animals. This stance attributes mental states as causal factors for the behavior being explained or predicted. In this paper, I argue that the intentional stance may well be more than a pragmatic choice. It can be viewed as a logically necessary choice given certain observed behavior and the causal logic of a primitive 'determinist stance.' Implications and potential assessment of this logical trigger are discussed.

Introduction

In a study of the evolution of purposive intentionality in artificial life, I reviewed the historical criteria for attributing purposiveness and intention to behaving systems (Watson, 2005). I found three primary criteria for the assessment of purposiveness: equifinality (displayed equivalent outcomes by varying behavior across instances of non-equivalent situations as illustrated in Fig. 1), rationality, and perseverance. Two additional criteria appeared relevant to a claim that the purposive system was also intentional. If the system met a criterion of "equi-origin" (displayed variable behavior across instances of equivalent situations as illustrated in Fig. 2), the system was said to meet what I termed "weak intentionality". If behavior supported a claim that it was directed by a representation of the outcome (that it met the philosophical notion of "aboutness"), the system was said to meet "strong intentionality". Although the analysis would surely not satisfy everyone, I tried to accommodate important criteria that have arisen in psychology, philosophy, and jurisprudence while not straying too far from the man on the street. In this paper, I propose that equi-origin may logically trigger the intentional stance.

Taking a Stance

Recent theorizing about cognitive-perceptual development has been influenced by Dennett's use of the concept of stance. His proposal and analysis of the intentional stance, the design stance, and the physical stance (1971, 1987) has had seminal influence on many developmental psychologists in their consideration of when children detect intentionality in others (Gergely, Nadasdy, Csibra, & Biro, 1995; Gergely & Csibra, 1997; Keleman, 1999, Leslie, 1995; Premack & Premack, 1995). The more recent proposal of the infant's teleological stance (Csibra & Gergely, 1998; Csibra, Gergely, Biro, Koos & Brockbank,1999) is a distinct concept but is clearly derivative in form. In each case, the stance is proposed as a cognitive-perceptual filter or bias that affects how the infant interprets events. Theorists have proposed specific cues that govern when the infant will assume the stance in question. For example, many have proposed that the cue of agency (self propelled motion) will invoke the intentional stance (Baron-Cohen, 1994; Leslie, 1995; Premack & Premack, 1995; and see Heider & Simmel, 1944 and Heider, 1958 for related discussions). Equifinality of action is proposed as a provocative stimulus for the teleological stance (Csibra, Gergely, Biro, Koos & Brockbank, 1999). The work of Lewis (1990) raises the possibility that emotional behavior in relation to success and failure of instrumental behavior may provide cues to intentional states.

Dennett's original argument for why we take the intentional stance is that it is pragmatic to do so. It is an evolved strategy for coping with the behavior of others. Evolution also provided selective sensitivity to certain cues, as noted above, regarding when it would be helpful to take the stance. From this view, cues do not logically force us to claim the existence of internal states, they simply alert us to the potential benefit of doing so.

The objective of this paper is to propose that the intentional stance can be viewed as logically forced by a more primitive stance: the determinist stance. From this view, the attribution of intentionality can be a logical necessity under certain conditions. What then is this more primitive determinist stance?

The Determinist Stance

The determinist stance 1 is a view of the world as lawful. Events are caused by other events. An event may (or may not) cause other events to happen. When a salient event occurs, it provokes the quest to find its cause. Formulating the lawfulness of the world allows one to predict future events and to understand/explain past events. An important assumption about the determinist stance is that it frames lawfulness as complete and universal. That is, a valid or good law is not sometimes true and sometimes false. Neither is it somewhere true but untrue in other places. From the determinist stance, a law is valid only if it is universally applicable. In the case of complete determinism: All things being equal, if a sufficient cause occurs then the effect should occur. Even from the perspective of probabilistic determinism: All things being equal, if a sufficient cause occurs then the effect should occur with specified probability over repeated instances (e.g. not .8 sometimes and .2 at other times).

A notable feature of good laws from the determinist stance is that they are not symmetrical in the relation of cause and effect. With a good law, if you know the causal context is complete, then you know the exact effect that will occur (or, in probabilistic determinism, the probability of that effect). But the reverse is not entailed. Alternative causes may exist for a specific kind of effect.2 In such cases, knowing that a specific effect has occurred leaves open the question of which of the alternative causes was the determinant.

Fixing a Failure of Universality

When a determinist formulates a law of behavior, the objective is to identify the situational event (S) that causes the behavior (B). This lawful relation is often rendered graphically as:

S ->B, or by the equation B = f S. If this law is meant to apply to people, then it needs to be applicable to all people all of the time. There are two ways this assumption of universality can fail, however. It might turn out that it seems valid for some people but not others. Or it might seem valid sometimes but not at other times for any given person. The determinist stance can not tolerate either failure because there is no wiggle room in this stance for free will or any other indeterminacy. Historically, determinists have found a conceptual cure for both of these potential failures of the universality principal. The cure comes in the form of dispositional property attribution (Armstrong, 1968,1969, 1993; Carnap, 1938; Prior, 1985; Ryle, 1949; Watson, 1995).

Regaining Universality with Dispositional Properties

Dispositions of Kind. When a behavioral law appears to vary across people, psychologists introduce dispositional constructs of individual difference or personality. These are dispositional properties of kind. An example is given in fig. 3.

Disposition of State. When a behavioral law appears to vary across time for particular people, psychologists introduce dispositional constructs of state (as in appetitive and emotional states) or stage (as in stages of learning and development). An example is given in Fig. 4.

A logical trigger for mental state attribution

A determinist tries to find a situational contrast to explain an individual's change of behavior from one time to another, but failing that, is logically forced to introduce a dispositional property of state. The logical pressure derives from a judgment that the situations are equivalent, and thus, by exclusion (or successive negation of disjunction), a non equivalence in individual is implied. The principle of universality requires that this claim of non equivalence be framed so as to be, in principle, a claim about any individual. The dispositional property claim meets that requirement. All people are comparable in reference to the dispositional property variation just as all situations are comparable in reference to their stimulus variation. Thus, when determinists are lead to believe that in equivalent circumstances the actor has or could have acted differently, they are forced to introduce a dispositional factor in order to avoid the failure of determinant lawfulness.

To believe that an individual could have acted differently in the same circumstances is a counterfactual assumption (in jurisprudence, support for this assumption is central to assessment of intent). If such

counterfactual claims were derived solely from imagined alternatives, then there would be little justification for introducing dispositional properties as additional factors in empirical laws that relate situations to behavior. Empirical laws are not obliged to explain imagined events. But the case is very different when the alternative behaviors have been observed in equivalent situations. These alternatives, if they are to be accounted for in a single law, force that law to include reference to a dispositional variation across the equivalent situations.

Implications

Einstein's Baby. Bloom (2004) has proposed that human infants appear to be innately committed to a view that there are two kinds of lawfulness in the world: determinant laws of the physical bodies in the world and self-determinant (or indeterminant) laws of humans and possibly other animate beings that are affected by the intentions of their "soul". This causal dualism is aptly captured under Bloom¡'s metaphoric book title Descartes' Baby. Bloom is not arguing for the truth of such dualism, only that it appears to be an innate bias. To highlight the alternative view that I am proposing, I have introduced the above heading. Einstein's baby is a determinist, and by my reckoning, she will thereby be logically forced3 to attribute intentions (and other mental states) as dispositional properties of state to cope with the behavior of others. In my view, she does not need a second form of law (e.g., God's dice) to detect and cope with the mental states of others.

Consider the case of an infant learning to adapt to his mother's interactive style. If he is to gain even a small degree of control of the interaction sequence, he will need to anticipate her reactions to his behavior. Now if he views his mother's behavior as independent of external causes, then there is nothing to learn save perhaps her general tendencies of behaving one way or another. Alternatively, if he views her behavior from a determinist's stance, he is set to relate her behavior to potential causal events in her environment including his own behavior. His task will be to uncover the causal factors. He will discard any notion that she might be behaving indeterminately. He may use this stance to refine his classification of what he should record as equivalent behavior on his part and what he should record as equivalent behavior on his mother's part (see Watson, 2001). Once his judgment of equivalence is settled, he is in position to feel the logical pressure for attribution of mental state change in his mother. This will arise whenever he observes her behavior fulfilling the criterion of equi-origin: her behavior varying (nonequivalent) across instances of equivalent situations (equivalent in terms of the setting and his behavior).

As a simple concrete illustration, I will shorten an example I have used previously in a related discussion (Watson, 1995). Imagine an infant who enjoys extended interaction with his mom. Sometimes when she tucks him in and he smiles at her, she turns out the light and leaves. Sometimes, however, she tucks him in and if he smiles, she picks him up again and interacts with him for a while longer. Alternatively, if he pouts when she tucks him in, sometimes she leaves but other times she picks him up and interacts a little longer. As a determinist, our infant rejects the notion that his mother is behaving randomly. The fact that she might be caused to stay (or likewise caused to leave) by a smile or a pout presents no special problem, since lawfulness allows alternative causes for a specific kind of effect. However, the fact that his smile (and likewise his pout) appears to sometimes cause her leaving and sometimes cause her staying in an otherwise equivalent situation calls for remedy of the infant¡¯s formulation of the lawful efficacy of his behavior. He needs to posit a disposition of state in his mother in order to maintain his commitment to determinism. With additional experience on his part, we might imagine (as in Watson, 1995) that he will eventually uncover cues to her state variation that he has been logically forced to assume. If he is successful, then he will cope well with the fact that, in this example, his mother sometimes has rewarding days at work and sometimes has days that have depleted her reserves of nurturance. On rewarding days she views his smile as a sign he is content (so she leaves) but his pout as a sign that he needs additional comforting (so she picks him up). On her difficult days, she views his smile as a sunny reprieve (so she picks him up) but his pout as a sign that he is asking for unnecessary support that she is not prepared to give at this time.

Animacy as equi-origin. Animacy (self generated motion) has been proposed as a prime cue for eliciting the intentional stance and/or purposiveness (Baron-Cohen, 1994; Leslie, 1994; Premack,1990, Premack & Premack, 1995). Csibra, Gergely, Biro, Koos, and Brockbank (1999) provide evidence that

animacy is not necessary for an infant¡'s interpretation of behavior as purposive by the criteria cited above. Cues of equifinality appear to be sufficient. However, if the attribution of intentionality stands on some evidence of equi-origin as proposed above, then animacy may have a special relevance to that attribution. The standard display of agency introduces evidence of equi-origin because the focal object's behavior varies in an otherwise constant situation. The present argument for equi-origin as a logical elicitor of the intentional stance would suggest that if change in motion of an object where immediately preceded by a salient change in the situation, then the behavior would be resolved as a simple situational determinant law. By contrast, if the change in motion were not accompanied by a situational stimulus, as is usually the case in animacy displays, then the behavior would trigger the inference of a dispositional state change in the object and the essential basis for attribution of intention would be provided.

Equi-origin and theory of mind. If equi-origin is a logical trigger for the infant's attribution of dispositional states, then attribution of belief should be affected by this trigger. A study by Onishi & Baillargeon (manuscript under review) for assessing sensitivity to false belief in infants would lend itself nicely to testing the triggering effect of equi-origin. The false belief test in theory of mind research (Bartsch & Wellman, 1989; Dennett, 1971; Gopnik & Astington, 1988; Perner, Leekam, & Wimmer, 1987) was presented in a manner allowing for assessment of much younger subjects (15-month-old infants versus 4-5 year-olds). With the violation-of-expectation method, infants were shown an object being hidden within one of two boxes while a person was present. In the focal test sequence (versus a number of control sequences), the hidden object changed place from one box to the other while the person was absent. The person returned and reached into one or the other of the boxes. Extended fixation (the criterion of surprise) was observed when the person reached to the true hiding place. This is taken as evidence that the infant assumed the person held a false belief regarding the whereabouts of the hidden object and expected a reach to the wrong box.

Onishi and Baillargeon did not manipulate equi-origin and they used a person as their focal object. A small change in their method could assess the roll of equi-origin in infants' attribution of this dispositional/mental state to even novel non-human objects. As depicted in left panel of Fig. 5, an infant could be shown the focal object making a choice between two occluders. At the moment of choice, there is a stimulus equivalence of situation from one trial to the next (1b versus 2b and 3b versus 4b), the only difference being an historical reference to where the hiding took place. If the infant understands this choice as based on the focal object's dispositional state of goal representation, as forced by the equi-origin evidence, then a slight variation in procedure should interfere with the infant taking this intentional stance. The variation involves introducing a signal (e.g. appearance of a black oval or a white oval at the moment the focal object begins to move) that is consistent with where the object is hidden. This stimulus can be viewed as a simple external determinant of the focal object's behavior as depicted in the right panel of Fig. 5. In this case, evidence of equi-origin is not present. By the view I am proposing, one should expect the infant to show surprise when the object chooses the incorrect occluder (the opposite result from the standard task as shown in Fig. 6).

Conclusion

The infant's intentional stance may be derived logically from a primitive stance on causal determinism and evidence of equi-origin of behavior. If so, evolution need not have designed a dualistic view of lawfulness for coping with animate and inanimate objects in the world. Schulz and Gopnik (forthcoming) have recently and independently developed a similar analysis for how young children may detect "hidden causes" in the physical domain. As in the proposal of this paper, the assumptions that young humans are capable of sensing logical implication, at least unconsciously, and are committed to determinism are viewed as providing cognitive leverage for taking unobservable causes into account.

Notes

1) It is possible that my "determinist stance" is equivalent to what Dennett meant to convey with respect to his "physical stance", but I am not sure enough to borrow his term. If it is, it should yet be clear that Dennett does not propose that the intentional stance is logically derived from the physical stance.

2) Historically, there has been disagreement about whether a particular kind of effect should be

conceived as having the possibility of more than one kind of cause. William James (as cited by Copi, 1953, p 331) was apparently against this idea of a "plurality of causes" and made the counter claim that "every difference must make a difference". However, it is a more commonly held position that, in many causal sequences, the effect may carry no distinction as to alternative sufficient causes (e.g. the light comes on the same whether I through the switch with my fingers or my thumb.)

3) The idea that very young humans might respond to logical implication is not new. Piaget (1954/1937) framed the infant's capacity to pass stage IV of object permanence (around 8 months of age) in terms of underlying capacities for representation and deduction. The contemporary research on physical and social knowledge in infants (e.g. solidity, gravity, numerical object relations, implications of desire) that relies on manipulation of expectancy (e.g., use of the habituation/recovery method) is at least implicitly assuming logical processing on the part of the infant. Gopnik and her colleages (Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004) are explicit in the assumption of the young child's sensitivity to logical constraint in their studies of how "causal power" is inferred. Gergely and Csibra and their colleages argue for a view of young infants employing "principle-based" reasoning about rational action when engaged in the teleological or intentional stance (Csibra, Gergely, Biro, Koos, and Brockbank; 1999). When making this assumption of a young child's sensitivity to logical implication, I believe most of the cited theorists would concur with Gopnik et als (2004) explicit proposal that these logical inferences occur at an unconscious level.

References

Armstrong, D. M. (1968). A materialist theory of the mind. New York: Humanities Press.
Armstrong, D. M. (1969). Dispositions are causes. (Reply to Roger Squires). Analysis, 30, 23-26.
Armstrong, D. M. (1993). Reply to Martin. In J. Bacon, K. Campbell, and L. Reinhardt, (Eds.) Ontology, causality and mind. New York: Cambridge University Press. Pp. 186-194.
Baron-Cohen, S. (1994). How to build a bsaby tht can read minds: Cognitive mechanisms in mindreading. Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 13, 1-40.
Bartsch, K. and Wellman, H. (1989). Young children's attribution of action to beliefs and desires. Child Development, 60, 946-964.
Bloom, P. (2004). Descartes' Baby. London: William Heinamann.
Carnap, R. (1938). Logical foundations of the unity of science. In International encyclopedia of unified science (Vol. 1, Part I). Chicago: University of Chicago Press. Pp.42-62.
Copi, I. M. (1953) Introduction to Logic. New York: The MacMillan Company.
Csibra, G., and Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. Developmental Science, 1, 255-259.
Csibra, G., Gergely, G., Biro, S., Koos, O., and Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. Cognition, 72, 237-267.
Dennett, D C. (1971). Intentional systems. Journal of philosophy, 8, 87-106.
Dennett, D C. (1987). The Intentional Stance. Cambridge, MA.: MIT Press.
Gergely, G. and Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to Premack and Premack. Cognition, 63, 227-233.
Gergely, G., Nadasdy, Z., Csibra, G. and Biro, S. (1995). Taking the intentional stance at 12 months of age. Cognition, 56, 165-193.
Gopnik, A. and Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction Child Development, 59, 26-37.
Gopnik, A. Glymour, C., Sobel, D., Schulz, L., Kushnir, T., and Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review, 111, 1-30.
Heider, F. (1958). The Psychology of Interpersonal Relations. New York: John Wiley & Sons.
Heider, F. and Simmel, S. (1944). An experimental study of apparent behavior. American Journal of Psychology, 57, 243-259.
Hull, C. L. (1951). Essentials of Behavior. New Haven: Yale University Press.
Kelemen, D. (1999). Beliefs about purpose: On the origins of teleological thought. In Corballis, M. and Lea, S. E. G. (Eds.), The Descent of Mind:Psychological Perspectives on Hominid Evolution. (pp. 278-294).Oxford: Oxford University Press.
Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. Hirschfeld

and S. Gelman (Eds.), Mapping the mind: Domain specificity in cognition and culture. New York: Cambridge University Press. Pp. 119-148.
Lewis, M. (1990). The development of intentionality and the role of consciousness. Psychological Inquiry, 1, 231-247.
Onishi, K. H. and Baillargeon, R. (manuscript under review). 15-month-old infants understand false beliefs.
Perner, J.; Leekam, S. R.; and Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. British Journal of Developmental Psychology, 5, 125-137.
Piaget, J. (1954/1937). The construction of reality in the child. (M. Cook, Trans.). New York: Basic Books.
Leslie, A. M. (1995) A theory of agency. In D. Sperber, D. Premack, & A. J. Premack (Eds.), Causal Cognition: A Multidisciplinary Debate. Oxford: Clarendon Press. Pp. 121-149.
Premack, D. (1990). The infant¡'s theory of self-propelled objects. Cognition, 36, 1-16.
Premack, D. and Premack, A. J. (1995) Intention as psychological cause. In D.
Prior, E. (1985). Dispositions. Aberdeen: Aberdeen University Press.
Ryle, G. (1949). The concept of mind. London: Hutchinson.
Sarason, S. B., Lighthall, F. F., Waite, R. R., and Ruebush, B. K. (1960). Anxiety in elementary school children: A report of research. New York: Wiley.
Schulz, L. and Gopnik, A. (Forthcoming) Sight Unseen: Causal Determinism Allows Preschoolers to Learn about Unobserved Causes.
Watson, J. S. (2005) The elementary nature of purposive behavior: Evolving minimal neural structures that display intrinsic intentionality. Evolutionary Psychology, 3, 24-48." Link is at http://human-nature.com/ep/
Watson, J. S. (1995). Mother-infant interaction: Dispositional properties and mutual designs. In N. S. Thompson (Ed.) Perspectives in ethology, Vol. 11, Behavioral design. New York: Plenum Press. Pp. 189-210.
Watson, J. S. (2001). Contingency perception and misperception in infancy: Some potential implications for attachment. Bulletin of the Menninger Clinic, 65, 296-320.

Discussion

Causal thinking is infantile thinking
Walter Freeman
Apr 14, 2005 4:44 UT

John, you present compelling evidence and reasoning, with which I agree, that infants attribute intentional stance to others by distinguishing between objects and anima. Obviously this developmental stage precedes acquiring language and episodic memory. I believe your thesis supports my belief and contention that causal thinking is so deeply ingrained in everyone that it's almost impossible to question. Hilary Putnam wrote: "Perhaps the notion of causality is so primitive that the very notion of observation presupposes it?" [1990, p. 75]

Yet, it is possible. Galileo did so in experiments with gravity, asking not why objects fell but how? The inverse square law that emerged from his experiments (not with balls dropped from the tower in Pisa – the time intervals were too short for measurement – but with balls rolling down an inclined plane) is acausal. So are Newton's equations, in which time can run either way. Elsewhere I have made the case that every major scientific advance in the past four centuries has stemmed from thought processes that transcend causal thinking, which your article enables me to label as 'infantile'.

On broader grounds, John Dewey challenged causation: "Purposive behavior exists and is given as a fact of behavior; not as a psychical thing to be got at by introspection, nor as a physical movement to be got at by physical instruments. It is and it exists as movements having specific qualities characteristic of them. ... But to ascribe independent complete existence to the movement, to say that is deliberate behavior, behavior having meaningful or conscious quality, is a fallacy of precisely the same kind as ascribing complete and independent existence to purpose merely as a psychical state. And it is a fallacy that flourishes only in an atmosphere already created by the belief in consciousness ... ." [1896, pp. 510-511]

Dewey J (1896) The reflex arc concept in psychology. Psychological Review 3: 357-370. Putnam H (1990) Realism With a Human Face. Cambridge MA: Harvard University Press.

Reply To Freeman's reply

Robert Paul
Apr 27, 2005 21:49 UT

I think that part of Walter Freeman's reply to John's paper on 'causal logic' really passes it by. It's one thing to say that 'causal thinking' is somehow ingrained 'innate', hard-wired, or, as a neo-Kantian would say, one of the categories we must use to organize and to make sense of experience; it's another to say that Galileo's experiment with balls rolling down an inclined plane was not an attempt to show what moved them, but to discover a law describing the rate at which falling bodies fall. This of course is true, but it can be true whether or not infants employ 'causal thinking,' or whether anyone does. Thus, that science at its most refined and fundamental level doesn't deal with causation could hardly be a criticism of ordinary notions of bringing about or making happen. And that the laws of physics are atemporal says nothing about causality, one way or another.

In fact, Freeman recognizes this but, as I understand him, suggests that 'causal thinking' is 'infantile thinking,' not just from a developmental standpoint, but in a mildly pejorative sense: when you come right down to it, physical laws do not deal with causes—one might extrapolate from this that there could be a complete physical description of the world that did not mention causes, let alone the notoriously recalcitrant notion of causation. (Some have thought this would entail that explanations inferred from physical laws don't make use of such concepts either.) Newton's first law of motion is indifferent as to whether anything is in fact at motion or at rest, and is equally indifferent as to what sort of 'outside force' would disturb a moving body or set a resting one in motion. So, when the physicist is writing equations and formulas on the board she will not be making use of the notions of cause and causation at all; yet when she is sitting at her desk and her computer screen goes blank, she might wonder, in a perfectly ordinary way, what made this happen. It is surely this sense of bringing about or making happen that John was trying to capture.

If humans had no notion, innate, or learned of one thing's bringing about another, they could not make sense of the world. (The notion of 'making happen' must be different from and richer than that of merely following, even though the former may somehow develop from the latter, but this is a separate issue.) So, one might argue that because humans can make sense of the world they must have some notion of one thing's bringing about another: how fine-tuned and fine-grained this is is something else again. To say, e.g., that while Aristotle might have known that staring at a bright candle flame for a moment and then looking away would produce an after image his knowledge was nevertheless incomplete because he didn't know what we now know about the physiology of the eye seems tempting; yet perhaps it is a temptation that should be resisted.

Causal thinking needs research

John Watson
Apr 15, 2005 5:20 UT

Walter's comments are more far reaching than the comparatively narrow focus of my paper. I will try to respond to his points, albeit in reverse order.

I am enlightened by the quote from Dewey. I think my conceptual model is perfectly compatible with the position he espoused in that passage (if I understand it correctly). The first portion of the quote points to the observable quality of purposive behavior and the second portion appears to disparage of viewing purposive behavior as arising de novo (as in free will of consciousness).

Walter's seeing in my paper support for his view that a causal (determinist) stance can now be called 'infantile' depresses my sense of having made the contribution I had intended. Since I have not read Walter's case against causal thinking, I am on thin ice in resisting his general claim that Galileo and

Newton were formulating “acausal” laws. My guess is that any conceptual discord between us will hinge on our respective concepts of causality. My perspective is that if a stated law implies that, were I to intervene on a variable in the law, I would gain a predictive advantage regarding the next state of the world, then I am dealing with a causal law. Such laws can be formulated without our capacity to intervene, of course (e.g. as in astronomy or evolutionary history).

Walter’s opening paragraph highlights an important issue for developmental psychologists. The idea that the causal perception of the world may be in place essentially at birth (an innate bias) has been seeping back into developmental theory in recent years. My proposal of the determinist stance is one such example. I and others have been more cautious in the past (Watson, 1984). Early in the last century, behaviorists held, at best, a Humean view of causal perception/conception. Piaget introduced a neo-Kantian view in which causal conception, while not innate, was a normative construction (Piaget, 1954). Our recent move to assuming greater preparedness than Piaget envisioned is surely in need of greater empirical support than it has a present. I am particularly concerned with our need to support ongoing assumptions (including mine) regarding early sensitivity to logical implication.

Piaget, J. (1954) The Construction of Reality in the Child. New York: Basic Books

Watson, J. S. (1984) Bases of causal inference in infancy. In L. P. Lipsitt and C. Rovee-Collier (Eds.), Advances in Infancy Research, Vol. 3. Norwood, N. J.: Ablex Publishing, pp. 157-176.

Determinism, equi-origin and the intentional stance
Anne Reboul
Apr 18, 2005 9:36 UT

I would like to begin by a marginal comment on the discussion between Walter and John: I may be mistaken (and if I am, I can only say that I hope that Walter will put me right), but my own guess is that Walter's view of causality takes as essential to it temporal asymetry of cause and effect as well as explanation. The first criterion allows him to say that Newton's equations, being temporally reversible are not causal, and the second one allows him to say that Galileo's inverse square law was not causal because it was not explanatory. To come to the main topic of my comment, I think that in fact John's position about the intentional stance being triggered by a combination of a basic (and possibly innate) determinist stance and equi-origin receives some support from the fact that both Dennett in his book on the intentional stance and Dickinson have advocated the intentional stance in animal studies (i.e., attributing first order intentionality to an animal) depending on whether, in a given situation, the animal seems to have a choice between several different behavioral reactions (hence the interest of the audience effect on the production of various types of calls — alarm calls or food calls — in species as different as vervet monkeys or chickens). Unless I am mistaken, and I would very much appreciate John's reaction, this is exactly what he means by equi-origin.

> **Equi-origin and apparent choice**
> John Watson
> Apr 18, 2005 20:32 UT
>
> If, in Anne's example, the animals display different behavioral reactions over instances that they are "in a given situation," then this would count as cases of equi-origin. That an individual "seems to have a choice" raises a separate philosophical issue, perhaps. Strickly speaking, from the determinist stance, the variation of behavior across instances of the same situation does not provide evidence of choice (as in free choice), but rather evidence of a state contrast in the individual (e.g. a contrast in intentional state) that in combination with the situation provides a causal account of the behavior in question.

Retrospective causality and modus tollendo ponens
Anne Reboul
Apr 20, 2005 13:45 UT

I would like to go back to a comment that John Watson made to my paper and to which I couldn't make a reply then because it came too late. You'll find the comment under http://www.interdisciplines.org/causality/papers/1/9#_9. In that paper, John agreed on the importance of retrospection in species comparisons, and refers to a paper of his (and colleagues) in which children and dogs were tested as too whether they used modus tollendo ponens (A or B, not A, thus B) or association to find an hidden object. Though this is not said in the paper, John says in his comment that going through logic in such a problem entails a capacity for retrospection because "the event "Not A" only carries a useful behavioral cue by reference back to the event establishing "A or B"" (this is in relation to the discussion of Call's paper in which chimpanzees also had to use modus tollendo ponens but which I said did not entail retrospective causal cognition). I should add that John says it in support of my global view. To be candid, I can't quite see why applying modus tollendo ponens should entail retrospection. Or if it does, then so should modus ponens (If A, then B; A; therefore B) or modus tollendo tollens (If A, then B; not B; therefore not A). If John means that to draw a conclusion, you need to go back to the first premise(s), then that is presumably true of all logical rules (though I think that one could see it rather as incrementing relevant knowledge by adding new premises: surely we can hold two premises in mind at a given time). If John means that if you don't already know modus tollendo tollens, you can learn it by going back to what the experimenter did (i.e. he went behind screen A, but he also went behind screen B and C), I suppose it's possible, but then this wouldn't be a standard case of applying a logical rule, but rather a process of discovering a logical rule, rather a different problem. So I'd like to have John's more detailed explanation on the question. Why should retrospection be involved in modus tollendo ponens?

> **Retrospection and logical thought**
> John Watson
> Apr 20, 2005 22:42 UT
>
> Let me try to clarify what I was trying to say in my response to Anne's comment on Walter's paper. I agree completely with Anne's statement that, from my proposal, syllogistic reasoning, in general, fits the retrospective orientation—not just the case of modus tollendo ponens. I was not saying that retrospection entails logical thought, however. Indeed, my distinction of retrospective and prospective conditional probability analysis is not a claim about logical analysis. It is (and has been) only a claim about how infants (and possibly other animals) detect contingencies in their experience (Watson, 1997). The choice of using modus tollendo ponens as an experimental assessment of potential logical analysis in dogs and children was because this syllogistic frame (versus modus ponens or modus tollendo tollens) readily lends itself to an operational distinction from what one would expect or predict from an associative learning perspective.
>
> In my view, contingency analysis and logical analysis are separable cognitive capacities--though in a review of theoretical mechanisms of contingency detection (Watson, 1997), I note that one of the four historical proposals I found in the literature is based on logical implication (Bower, 1989). Contingency analysis can proceed without logical analysis. Contingency analysis can proceed without retrospection--contiguity and correlation are mechanisms that do not depend on it (Watson, 1997). But I am proposing that logical analysis does depend on retrospection. Anne's claim that "surely we can hold two premises in mind at a given time," if correct, would be negative evidence for my proposal. While I believe I can see a compound stimulus in a single moment, my introspective phenomenology is not in accord with hers regarding the perception of and relating of two propositions in a single moment. Has anyone tested this question? If so, how did they test it? Finally, as to whether I was focused on applying versus discovering a logical rule, I was focused on applying it. However, I do believe that we will eventually find that a sense of implication may precede and play a role in the developing individual's recognition of valid versus invalid syllogistic form (just as it appears to have in the discovery of fallacy in the history of logic, e.g. Geach's (1958) history of the boy/girl fallacy).

I would conclude this response with the hope that my speculations regarding retrospection will not occlude the relatively independent speculation of my paper regarding the possible role of a determinist stance in the eliciting of mental state attribution. I tried to include examples of testing the latter speculation, while I have no clear ideas about testing the former at this time.

Bower, T. G. R. (1989) The rational infant: Learning in infancy. New York: W. H. Freeman.

Geach, P. (1958) History of a fallacy. Journal of the Philosophical Association (Bombay), 5, no. 19-20, July-October.

Watson, J. S. (1997) Contingency and its two indices within conditional probability analysis. Behavior Analyst, 20, 129-140.

Scientific vs. Animistic Thinking
Walter Freeman
Apr 22, 2005 3:35 UT

John, I do apologize that my use of "infantile" depressed rather than enlightened you. I sensed the pejorative tone even as I selected it, but it was so perfect for my intended meaning that I used it anyway.

There is no difference between our views concerning the importance of the early development of intuitive understanding of cause-and-effect in prelingual children, as the basis for acceptance of responsibility for personal actions, which is why I expressed full support for your discourse examining the process by which children establish a working knowledge of causal logic through trial-and-error in the use of their bodies to achieve their goals through their "intentional stance".

As both you and Anne surmised, my target for understanding was and is the tacit assumption that underlies your comment concerning the advantage of prediction by causal laws: "Such laws can be formulated without our capacity to intervene, of course ..." [my italics]. Anne is correct in her interpretation of my view that causal laws require 'temporal asymmetry' and can explain only 'how?' and not 'why?', but that interpretation doesn't go far enough. My target is the attribution of agency in linear causality, by which we seek for 'movers', ultimately Aristotle's Prime Mover or the Big Bang. In that logical process, free will is seriously questioned, even denied in Universal Determinism, which for me is a fine example of disproof by reductio ad absurdem. So I want to distinguish between causal logic and relational logic. Anyone who describes the world with differential equations, or with statistical risk factors, or with Riemannian and Einsteinian hypergeometries is going beyond primitive, animistic thinking into strange territory.

There is nothing new in these thoughts thus far. What I bring to the table is the neurobiology of intentionality and the action-perception cycle, by which all knowledge is acquired and exercised. What I claim is that this neurobiological foundation explains why we humans rely so heavily on causal logic, why it is acquired so early, and why we have so great difficulty in transcending it in our approach to the physical world, as distinct from the psychological, social, and legal worlds.

Equi-origin, the intentional stance and developmental evidence
Anne Reboul
Apr 22, 2005 9:19 UT

I'm sympathetic to John's clever account of the intentional stance being triggered by the evidence of equi-origin in a determinist stance on events. However, the difficulty with the hypothesis is that, were it true, you'd expect that the intentional stance would appear at different ages in different children (depending on their experience), which does not seem to be true. To give some background to the problem I want to

raise, the standard false belief tests (i.e., displacement or so-called Sally-Ann tests) of theory of mind (the ability underlying the intentional stance) all converged on the finding that success at false belief occurs at around four years of age. This has been diversely interpreted though a common idea is, roughly, that children before four years have a proto-concept of belief which is of such a nature that it can support true belief but not false belief. I've also had occasion to read Onishi & Baillargeon's paper (which, by the way, has just been published). The false belief tests were mainly verbal tests and the novelty of Onishi & Baillargeon's work is to propose versions which are accessible to preschoolers, aiming to show that the age of success at false belief is much earlier than was previously thought (15 months). This would support the idea that in fact the difficulty with verbal tests does not come from passing from a proto- to an "adult" concept of belief but from the excessive computational, linguistic, or memory task demands. There is however an objection to such an account, which has been recently (and cogently) in my view, formulated by Richard Brehenny (to appear): if indeed, failure at verbal false belief tests was an artefact of these experiments (whether linguistic, computational or memory difficulties are involved), you'd expect chance results for children below four. This, however, is not what is found: what is found is that the great majority of children below four will give the wrong answer, while the great majority of children over four will give the correct answer. So, the question is: if equi-origin considered in a determinist stance is the trigger for the intentional stance, why do children before four overwhelmingly give the wrong answer? And, if, as seems to be shown by Onishi & Baillargeon's results, much younger children succeed at nonverbal version of the tests, suggesting that the late age of success in verbal versions is due to linguistic or other factors, why are the results of such tests before four not at chance level, as would be expected?

REFERENCES Breheny, R. (to appear), "Communication and folk psychology", Mind & Language (Available at http://www.phon.ucl.ac.uk/home/richardb/papers.html).

Onishi, K.H. & Baillargeon, R. (2005), "Do 15-month-old infants understand false beliefs?", Science, 308, 255-258.

The intentional stance and false-belief data

John Watson
Apr 22, 2005 23:17 UT

The points raised by Anne are intriguing and challenging, particularly so for consideration of cognitive underpinnings of children's capacity to pass the false belief test. I have two defensive (perhaps overly defensive) responses regarding the relevance of the cited data (from Breheny) for my proposal. First, the intentional stance (at least as proposed by Dennett) encompasses a variety of mental state attributions (e.g. desire, intent, memory, belief). My proposal about the role of the determinist stance and evidence of equi-origin is an attempt to explain the circumstances in which a child would be moved to make a mental state attribution (this explanation being an alternative to existing proposals that the intentional stance is cued by stimulus features). Dennett's suggestion of the false belief test as a means of confirming an attribution of belief was seminal for developmental psychologists. However, the test is not a test of the full range of intentionality (e.g. the attribution of desire or intent). I referred to the Onishi & Baillargeon study because it is a startling example of preverbal assessment and apparent success with false belief. I thought that if my proposal were valid, then this success should be vulnerable to a simple manipulation of the setting. In sum, I am resisting Anne's apparent equating of the intentional stance with the capacity to pass the false-belief test. I would also resist equating the intentional stance with the child's capacity to co-ordinate knowledge and intention or co-ordinate knowledge, desire, and intention (as discussed by Breheny). Second, I am not quite clear as to what the consistency of error in the meta-study data cited by Breheny can tell us about the young child's cognitive system. Without knowing the full experimental paradigms I am surely groping in the dark here, but I can imagine many ways in which a physical and linguistic setting might lead to non-random behavior in tests that are not understood and thus not passed by children. But let me add quickly that I am biased to a view that some mental states (intentional states) are very likely more difficult to conceive and attribute than others. In near tautology, I assume that simpler mental states will be attributed more easily and earlier. I am sympathetic to the idea that proto-forms of mental states may precede fully formed attributions.

However, in light of my paper's proposal, I assume that the determinist stance and evidence of equi-origin will be important in leading the child to make the attribution of any mental state.

Do young children possess dinstinct causalities for the three core domains of thought?

Kayoko Inagaki (Professor of Faculty of Education of Chiba University)
Giyoo Hatano (Professor of Psychology, University of the Air, Japan)
(Date of publication: 26 April 2005)

Abstract: A growing number of developmentalists agree that even young children possess coherent bodies of knowledge about a few important aspects of the world and that naive physics, psychology and biology are surely included among them. The acquisition of these core domains of thought is believed to be early, easy, and almost universal. How early do young children possess these naive theories of the world? Do they acquire three distinct causalities for these domains approximately at the same time? In the first half of this paper we discuss these issues. Our tentative conclusion is: whereas even three-year-olds have surely acquired and use differentially causalities for naive physics and psychology, a causal explanatory framework unique for biological properties and processes is established a little later. In the last half of the paper we speculate how young children can acquire these characteristic causal devices.

Introduction

A growing number of developmentalists (e.g., Keil, 1998) agree that human beings are genetically predisposed, prior to any experience, to attend to some events (e.g., an entity's movement) rather than others and to attribute a particular event to some preceding events rather than others (e.g., bodily ailment to unfamiliar food). Coherent bodies of knowledge about important aspects of the world are then built on these bases, and many researchers assume that naive physics, psychology, and biology are surely included among the core domains of thought. The acquisition of these core domains is early, easy, and almost universal in the sense that most if not all children possess domains of physics, psychology and biology before or without schooling or any other systematic teaching.

Each of the knowledge systems of these core domains supposedly includes a characteristic set of causal device. Thus, it is implied, even young children possess (1) physical or mechanical causality, by which they can predict and sometimes even explain, the motion, trajectory, and speed of solid objects in terms of 'force'; (2) psychological or intentional causality attributing (human) goal-directed behaviors to mental states including desires, beliefs, emotions and intentions; and (3) biological or teleo-vitalistic causality attributing bodily processes and properties to their functional significance and/or the working of vital power or energy within them. To put it differently, young children can make coherent and reasonable predications for representative physical, psychological, and biological phenomena. They can even offer a relevant causal explanation, or at the least, choose such an explanation out of a few alternatives. Furthermore, it is often claimed, they can apply these three different types of causalities flexibly and appropriately to specific situations.

How well are these claims about the early possession and differential application of causalities supported? How can young children recognize the proper causality for and adjust it to a variety of situations? How can they acquire, in the first place, these three causalities though adults seem not to teach them seriously? In this paper we will discuss these issues.

Young children possess three different kinds of causalities and apply them differentially

One of the prototypical situations in naïve physics is about the transmission of force from an object to another by contact. A somewhat earlier but ingenious study by Bullock, Gelman, and Baillargeon (1982), using a domino-like apparatus consisting of a rod pushing through a post, five standing wooden blocks and a rabbit doll falling from a platform, provided with us the evidence for young children's understanding of physical causality. Children aged 3-4 years, after viewing the complete "domino" sequence, were asked to predict whether the expected final event would occur when some modifications were done for the initial event (e.g., rod) or intermediary events (e.g., blocks). There were two types of modifications: one was relevant (e.g., a rod too short to hit the first block) and the other, irrelevant (e.g., the color of the rod). The

children were asked to give 23 predictions in all. Results indicated that 78-91% and 70-100% for the 3- and 4-year-olds' predictions were correct.

Shulz (1980) reported that children aged 2-6 years understand physical causality even when the transmission of an effect is mediated by wind. Each child observed events such as the following: a shield was placed around a candle with the open side facing the child, the candle was lighted, one blower was switched on, and after five seconds, the shield was removed. When the child was asked which blower made the candle go out, he or she correctly identified the responsible blower.

As well known, studies on theory of mind have shown that preschool children understand that human goal-directed actions are explained in terms of intentional causality, more specifically, his or her mental states, such as intentions, desires, emotions, and beliefs. A recent meta-analysis of the false belief task performances by Wellman, Cross and Watson (2001) clearly shows that by about age four children come to understand that actions are based on the actor's representation of the world, not the world itself. In other words, intentional causality is acquired as early as mechanical causality.

Some recent studies have indicated that even infants may be sensitive to the ontological category of entities. In particular, they seem to have different expectations for humans versus nonliving solid objects. Preschool children likewise apply different causal explanations for animate entities and inanimate ones. For example, Massey and Gelman (1988), using unfamiliar animals (e.g., echidna) and artifacts (e.g., wheeled vehicles, rigid objects, etc.) as target objects, found that 3- and 4-year-olds can correctly judge whether animals and artifacts have a capacity for self-initiated movement or not. The children were correct on about 85% of their first yes/no answers, indicating that they answered that animals could go up and down a hill by themselves, while inanimate objects, even if they looked like animals, couldn't. Analyses of explanations that the children gave spontaneously or in response to a request for justification of their yes/no responses suggested that these children tended to change kinds of explanations depending on the types of the objects. When talking about an animal, children often focused on body parts that enable the target to move such as "It can move because it has feet," whereas for the wheeled vehicles or rigid objects, they referred to an agent needed to move the object down and up (e.g., "It needs a push and then it goes," or "You have to carry it down."

A causal explanatory framework unique for biological properties and processes, however, may be acquired a little later than psychology and physics, because understanding bodily processes presupposes the awareness of the processes (Inagaki & Hatano, 2002), and because a biological mode of causal reasoning presupposes the construction of an integrated category of living things including animals and plants, which appear so different. Nevertheless, studies to date have indicated that children as young as 5 years can apply biological, or teleo-vitalistic causality to biological phenomena. When asked to choose one from three explanations presented for bodily phenomena such as digestion or respiration, young children preferred vitalistic explanations most often (Inagaki & Hatano, 1993, Morris, Taplin & Gelman, 2000). For a question, "Why do we eat food every day?", for example, the children preferred a vitalistic explanation ("Because our stomach takes in vital power from the food") to either an intentional explanation ("Because we want to eat tasty food") or a mechanical explanation ("Because we take the food into our body after its forms is changed in the stomach and bowels").

The evidence so far concerns that preschool children can apply different causalities to different entities or events. The real complexity in the enterprise of differentiating explanations lies in applying different types of causal reasoning for different behaviors of the same entities. As pointed out by Wellman, Hickling, and Schult (1997), humans are not only psychological beings but also biological organisms, and physical entities as well. In other words, human actions can be caused not only by psychological states (e.g., desires and beliefs) but also by physical forces (e.g., gravity) or biological processes (e.g., reflexes).

Thus, Wellman et al. took up young children's explanation of human actions that might induce psychological, physical, or biological reasoning, and examined whether children could apply types of causal reasoning best fit the target phenomena. In a series of their experiments, 3- and 4-year-olds heard four to nine stories about a protagonist who wanted and intended to do something and could or could not do so. For a simple, spontaneous actions (e.g., pours milk on the cereal) a protagonist can do so as he wants, but for impossible physical actions (e.g., floats in the air) or for impossible biological actions (e.g.,

hangs on a branch forever) a protagonist cannot do his intended actions because of physical or biological constraints. After each story, the children were asked to explain, "Why did that happen? Why did the protagonist do that?"

Wellman et al. found that the 4-year-olds consistently gave different explanations for the three kinds of human actions; they gave psychological explanations most often for the psychological actions, such as, "He thought it was milk"; biological explanations most frequently for the biologically impossible actions, for example, "His arms got hurting"; and physical explanations most often for th

The 3-year-olds consistently gave psychological explanations for the psychologically caused human actions and physical explanations for the impossible physical actions, but for the impossible biological actions they gave as many psychological explanations as biological ones. However, when asked to judge whether the desired actions were possible (before offering explanations), 3-year-olds predicted as correctly as the 4-year-olds, saying that psychological actions were possible but that the physically-constrained or biologically-constrained actions were not possible. This indicates that the 3-year-olds, like the 4-year-olds, clearly made different predictions for human actions with psychological and biological impetus.

Based on these laboratory data and further analyses of everyday conversation, Wellman et al. concluded that "children evidence at least three basic everyday reasoning systems as young as 2 years -- physical and psychological reasoning surely, and even biological reasoning in a rudimentary form," and apply them in flexible and sensible ways. However, the scrutiny of their laboratory data reveals that it is not until at 4 years of age that children understand well the psychology-biology contrast in causality.

Inagaki & Hatano (2002) dealt with the biology-psychology distinction. Here children aged 4-6 years were presented with a pair of drawings of two boys, who were allegedly different in terms of biological/bodily (e.g., imbalanced diet or insufficient fresh air) or psychological/social factors (e.g., misbehaviors) in their daily activity, and asked which of the two was more likely to catch cold. Results indicated that a majority of the children in each age group chose often, as being more likely to catch cold, the boy who engaged in biologically bad activities. Although the 4- and 5-year-olds could give few reasons for their choices, about half of the 6-year-olds justified their responses to the items of eating a little or eating few vegetables, such as, "(A boy who has) little nutriment does not have energy, so germs easily enter his body" or "When this boy X eats a lot, his throat is full of nutriment. This boy Y eats little, so his throat is not full of nutriment, and so the coughing can pass through his throat." Since the choice patterns of the 4- and 5-year-olds were very similar to those of the 6-year-olds, the 4- and 5-year-olds, like the 6-year-olds, would consider that a lack of energy or vital power is likely to make children susceptible to illness. Their results also indicated that these children believed that social/psychological factors would influence susceptibility to illness. However, when forced to choose one between the biological and psychological factors, the children evaluated the former factor more important than the latter for determining susceptibility to illness.

The above experimental data strongly suggests that by about 5 years of age children come to appropriately apply different causalities to psychological, physical and biological phenomena.

Differential applications of causal devices

How can young children apply causal devices so skillfully? The knowledge systems of the core domains that include characteristic causal devices are a prerequisite, but there must be a few additional conditions. Let us propose two of them below. First, even young children have some intuition about the appropriateness of causal devices that seems similar to lay adults'. In other words, their choice of the domain to which a given phenomenon is assigned is more or less pertinent, if somewhat unstable or inaccurate. In fact, they are fairly good at clustering pieces of knowledge (Lutz & Keil, 2002). Thus, when asked to predict or explain, they can promptly activate a few causal devices as promising candidates. The living-nonliving distinction is acquired especially early, and thus they seldom explain behaviors of living and nonliving entities in similar ways. Moreover, young children can aptly switch their explanations, taking contexts into consideration. For example, for apparently the same human action of drinking, a large majority of 6-year-olds gave a biological explanation when it was "inevitable" (i.e., drinking water after running for a while) but the psychological one when it was "optional" (taking a glass of juice when

beverages were offered), as illustrated as follows: one 6-year-old boy answered for the inevitable behavior, "Because the protagonist was thirsty", whereas he explained, for the optional behavior, "'Because the protagonist liked banana juice".

Second, though young children's repertory of possible causal devices must be limited, they can ingeniously process them by relying on analogy and metaphor. For example, when they see half-withered grass become greeny and healthy after a shower, they can infer that being given water produces the change. They may apply the core biological causality of "vital power taken from food/water makes a human active and lively" with a slight modification and indicate that "vital power taken from a shower makes grass healthy." Children are good analogists and are particularly good at exploiting usable partial analogical relationships, particularly often in naive biology (Inagaki & Hatano, 2003). Similarly, children may metaphorically expand the notion of vital power from the material or energy taken from food to the mental or social one such as refreshing air or sympathetic support.

Multiple origins of causalities

Preschool children have already acquired the ability to make coherent and reasonable predications for typical physical, psychological, and biological phenomena, and they can sometimes even offer a relevant causal explanation. When does this ability emerge? How is it prepared? Before concluding our paper, let us speculate about this issue, extrapolating from what we have seen for preschool children. We would like to rely on Piaget's ideas and observations rather than assuming forms of core causal knowledge to be "innate." First, we accept his general idea (1950) that intellectual devices including the core causalities are prepared during the sensorimotor period of intelligence, and are transferred to the verbal form later when children acquire the symbolic function and language. Second, we adopt his specific idea (1954) that infants' understanding of causality starts with the case in which they themselves are agents (i.e., his protocausality). As well known, Piaget was concerned with the long and gradual process of 'objectivization' of protocausality and the emergence of 'true' causality. We instead emphasize his starting point. In other words, we assume that the initial form of causality is the sense of agency that one's action consistently produces a desired change in an entity if and only if the action is executed appropriately. However, we believe that we have to expand Piaget's formulation to give a due place to the origin of psychological and biological causalities. In other words, in spite of his great insight, we are afraid, Piaget ignored the domain-specificity of causality and treated physical causality as the sole origin of forms of causalities. As Mandler (2004, p. 101) conjectures, a notion of causal force can be derived from perceptual analysis of the transfer of motion between two objects combined with "bodily experiences of pushing against resistance and being pushed," but it is only part of the whole story. Although very few episodes were reported by Piaget as to the infants' actions and reactions to caretakers and other humans, we can just assume that infants all over the world have to seek a caregiver's help to satisfy their physiological needs. They soon find that they can have the caregiver come by communication, such as moving their body, arms, hands or vocalizing. No direct contact is required. Again, their grasp of the causal role of communication in inducing the approaching behavior of the caregiver is "protocausal," that is, limited to when they are serving the causal agent.

The case of biological causality is a little different, primarily because biological processes are slower to proceed than the physical or mental ones. The situation in which infants are likely to recognize the effect of his attempt to operate regularly is probably the removal of hunger and lassitude by being fed, and this is consistent with the eating-centered nature of young children's naive biology. Infants may not suffer from severe hunger in the technologically advanced society, but they become hungry surprisingly fast. Although feeding may not be at infants' disposal, we assume their persistent request is usually met promptly by the caregiver. Thus they learn that something from milk or another food, which might later be conceptualized as 'vital power,' makes them full of energy. In conclusion, the core causalities are prepared during the sensorimotor period of intelligence as hypothesized by Piaget. However, unlike his assumption, there must be multiple origins and multiple pathways of development for physical, psychological, and biological causalities.

References

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), The developmental psychology of time. New York: Academic Press.
Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. Child Development, 64, 1534-1549.
Inagaki, K., & Hatano, G. (2002). Young children's naive thinking about the biological world. New York: Psychology Press.
Inagaki, K. & Hatano, G. (2003). Conceptual and linguistic factors in inductive projection: How do young children recognize commonalities between animals and plants? In D. Gentner, & S. Godin-Meadow (Eds.), Language in mind. The MIT Press.
Keil, F. C. (1998). Cognitive science and the origins of thought and knowledge. In W. Damon (Ed.), Handbook of child psychology, 5th ed., Vol. 1: R. M. Lerner (Ed.), Theoretical models of human development. New York: Wiley.
Lutz, D. J. & Keil, F. C. (2002). Early understanding of the division of cognitive labor. Child Development, 73, 1073-1084.
Mandler, J.M. (2004) The foundations of mind. New York: Oxford University Press.
Massey, C. M., & Gelman, R. (1988). Preschooler's ability to decide whether a photographed unfamiliar object can move itself. Developmental Psychology, 24, 307-317.
Morris, S. C., Taplin, J. E., & Gelman,< S. A. (2000). Vitalism in naive biological thinking. Developmental Psychology, 36, 582-613.
Piaget, J. (1950). The psychology of intelligence. Routledge and Kegan Paul.
Piaget, J. (1954). The construction of reality in the child. New York: Ballantine books.
Shultz, T. R. (1980). Rules of causal attribution. Monographs of the Society for Research in Child Development, Vol. 47, No. 1.
Wellman, H. M., Hickling, A. K., & Schult, C. A. (1997). Young children's psychological, physical and biological explanations. In H. M. Wellman & K. Inagaki (Eds.), The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena . San Francisco: Jossey-Bass.
Wellman, H. M., Cross, D. & Watson, J. (2001). Meta-analysis of theory-of-mind develop

Discussion

Specific core domains: innate or acquired
Anne Reboul
May 2, 2005 8:53 UT

I'm sympathetic to Kayoko and Gyoo's partly Piagetian hypothesis that different kinds of causal knowledge corresponding to the three core systems of (folk) physics, biology and psychology are acquired rather than innate. I was however slightly surprised to see that the experiments they quote in the physical domain all concern preschoolers but neither infants nor todlers. Yet there are quite a lot of experiments usually using the habituation-dishabituation paradigm purporting to show that infants much younger than one year have physical knowledge. So my first question is why do Kayoko and Gyoo ignore these experiments? Do they share the doubts about these paradigms that have recently been expressed by Aslin & Fiser (2005)? Or, if not, why don't they rely on these experiments? There are a few additional worries: if I understood the paper correctly, the developmental chronology would be first physical causal knowledge, then psychological knowledge and finally biological knowledge. There are, however, a few worries about this: infants are very little mobile before 3 to 4 months of age and are certainly not agents before that age. However, the inner sensations (hunger, etc) which Kayoko and Gyoo describe as being the basis for biological causal learning are present from the birth on. Additionally, the primitive communicative abilities (crying, etc. to induce caretaker's help) are also present from the birth on, eventhough they become progressively more refined with time, an important milestone being smile which appears at around two months (on infant communication, see Owings and Zeifman 2004), before any possible agentive act. So,

on the face of it, if the acquisition approach is right, one would expect either biological or psychological causal knowledge to precede physical causal knowledge rather than the reverse. Though this is sheer speculation, couldn't one assume a mixture of innateness and acquisition which would ensure learning bias (presumably perception based), which would come into action through maturation (not learning), but which would need experience (including experience of a linguistic knid in some cases) to yield normal cognitive abilities (for a similar account on songbird's learning, see Marler 1999)?

REFERENCES Aslin, R.N. & Fiser, J. (2005), "Methodological challenges for understanding cognitive development in infants", in TICS 9/3, 92-103. Marler, P. (1999), "On innateness: are sparrow songs "learned" or "innate"?", in Hauser, M. & Konishi, M. (eds), The design of animal communication, Cambridge, MA, The MIT Press. Owings, D.H. & Zeifman, D.M. (2004): "Human crying as an animal communication system: insights from an assessment/management approach", in Oller, D.K. & Griebel, U. (eds): Evolution of communication systems: a comparative approach, Cambridge, MA, The MIT Press.

Physical causality in human infants

Susan Hespos (Cognitive Studies and Developmental Area,Vaderbilt Univeristy)

(Date of publication: 10 May 2005)

Abstract: I would like to focus on whether there is specificity in causal cognition and how language influences this ability. Data from preverbal infants suggest that there is a set of perceptual and conceptual capacities that are common to everyone and rich enough to capture the core meanings expressed by any language. Language learning develops by linking linguistic forms to universal, pre-existing representations of meaning.

Causality is the cognitive basis for the acquisition and use of categories and concepts in children. The previous papers in this conference have discussed whether causality has a single underlying mechanism for this ability (Watson), two systems (Reboul), or three domains (Inagaki & Hatano). Watson suggests that a single mechanism could be used to reason about objects and people and that there is no need for these to be divided. These views are in line with work by Onishi and Baillargeon (2005) which will be discussed below. Reboul highlights comparisons between human and non-human causal reasoning and suggests that the onset of language might be a critical difference between arbitrary causal knowledge and natural causal knowledge. Some of my work coincides with this view and I will describe data that relate to these points. Finally, the core domains of physics, psychology, and biology presented by Inagaki and Hatano present compelling data describing the different signatures that each of these domains present. Their views overlap with Spelke's (2000) core domains, with the caveat that the domain names are slightly different and there may be more than three.

In this paper I take a developmental approach and investigate the origins of physical causality in infancy. By looking at change and continuities in infants' abilities it may be possible to gain a better understanding of the fundamental aspects of causal reasoning. Another advantage of looking at young infants is that we can gain insight to the nature of representation abilities prior to the onset of language. If we can characterize the prelinguistic state of causal categories and compare it to adult abilities, then we can better understand the impact that language makes on our cognitive processes. I propose that physical causality is evident early in infancy and it develops in a category-specific pattern. This ability is likely to be universal among human and non-human primates and language alters our conceptual categories.

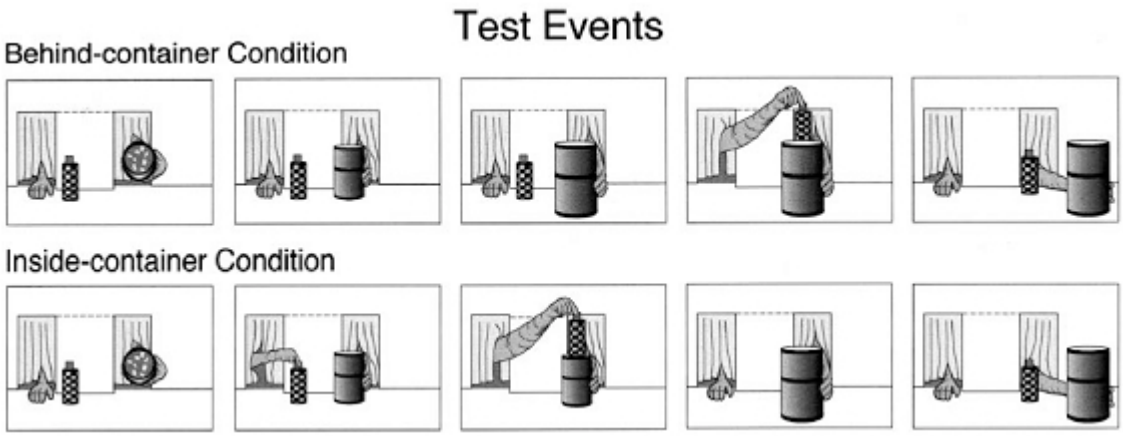
Knowledge about physical causality - defined as the way that objects behave and interact - is a central issue in cognitive development (Haith & Benson, 1998; Piaget, 1952, 1954; Spelke & Newport, 1998). The classic view of object representation was Piaget's (1952, 1954) depiction of object permanence developing over the first 18 months of life. Qualitative shifts in cognitive concepts about objects were central to Piaget's theory (e.g., out of sight is out of mind to the A not B error). More recently, theories of continuity and elaboration are used to explain developmental changes in object representation. Mandler describes a dual representation model where perceptual and conceptual capacities develop in parallel (Mandler, 2004). Baillargeon (2004) suggests that infants have some innate principles and over the first year they identify category-specific variables that allow them to predict the outcome of events more accurately. Spelke (2000) suggests that signature characteristics in object representation are ontogenetic - the characteristics of early object representation are still observable in adult object representation.

Experiments that test knowledge in preverbal infants rely on infants' tendency to habituate to repeated events and look longer at events that they perceive as novel or unexpected. Baillargeon and her colleagues have used looking time methods for over twenty years to map the developmental trajectory of infants' knowledge about physical causality. In this time we have amassed data that describe what infants know at different ages. More recently the agenda has switched to finding out how infants learn about objects (Baillargeon, 2004).

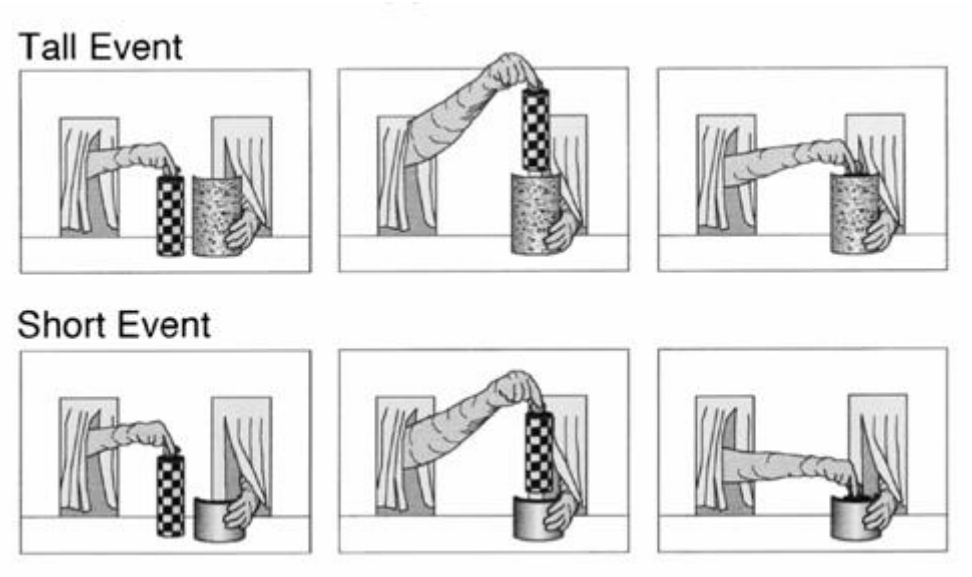
If the task before the infant is to learn about how objects behave and interact. The solution that infants seem to use is to divide the world into smaller categories of events and learn within each of these categories. This strategy implies that knowledge is bounded and doesn't generalize to other categories. Baillargeon (2004) has revealed this pattern of learning for a wide variety of physical events including,

support, collision, occlusion, covering, and containment.

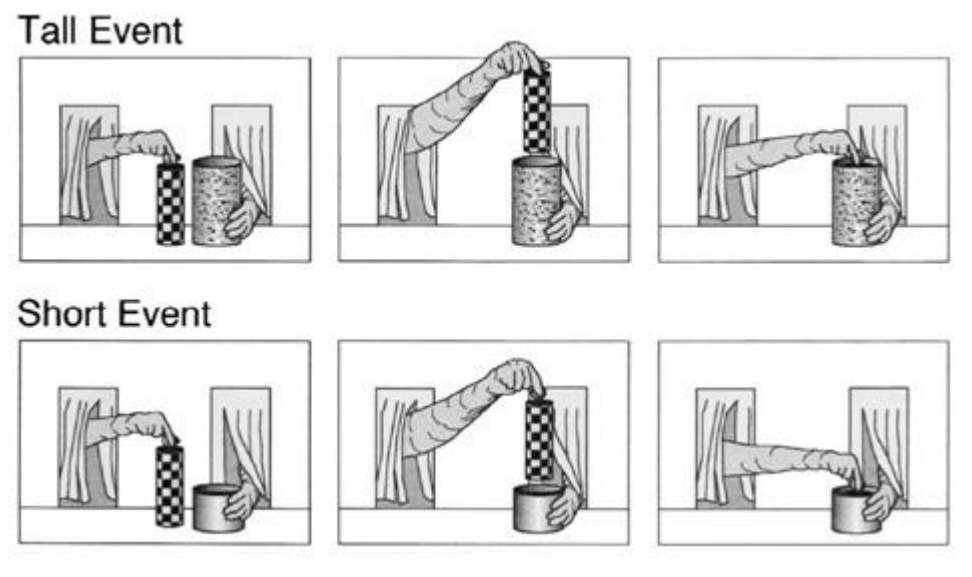
To illustrate I will describe the developmental trajectory of containment events in infancy. Our first study investigated whether infants had different expectations about occlusion and containment events. To test this we recorded infants looking times to the two displays depicted below. The perceptual features of the two displays were very similar the only difference between them was whether the checkered object was lowered behind the container (an occlusion event) or inside the container (a containment event). There was an important conceptual difference across the displays. The container event has and unexpected outcome because the checkered object appears to pass through the side of the container. If infants detect this difference they should look significantly longer at the containment event compared to the occlusion event. Two-month-old infants looked significantly longer at the containment compared to the occlusion event suggesting that very young infants have specific expectations about these events (Hespos & Baillargeon, 2001b).



Further experiments investigated how these initial expectations change over time. Baillargeon's (2004) model predicts that infants identify variables that allow them to predict the outcomes of events more accurately over time. For example, by 4 months of age infants have expectations about how much of an object should become hidden when it is lowered behind an occluder (Hespos & Baillargeon, 2001a). When infants were shown the events depicted below, they looked significantly longer at the short compared to the tall occluder event.



To investigate whether infants generalize their expectations to perceptually similar containment events we tested a new group of infants in conditions where the occluders were replaced with containers (see below). We tested 4-month-old infants in the container condition and they looked equal amounts of time at both events suggesting that they did not discriminate the short event as more unexpected than the tall event. Next we tested new groups of 5- and 6-month-old infants and got the same result. It wasn't until the infants were 7.5-months old that they looked significantly longer at the short compared to tall containment event. These findings suggest that infants' knowledge about physical causality is category specific and does not generalize broadly across perceptually similar events. Further evidence of context-specific limitations were found for categories of containment vs. covers and covers vs. tubes (Wang, Baillargeon, & Paterson, 2005). In addition, the same developmental patterns have been revealed in reaching tasks testing knowledge of occlusion, containment, and support events (Hespos & Baillargeon, 2005, in prep).

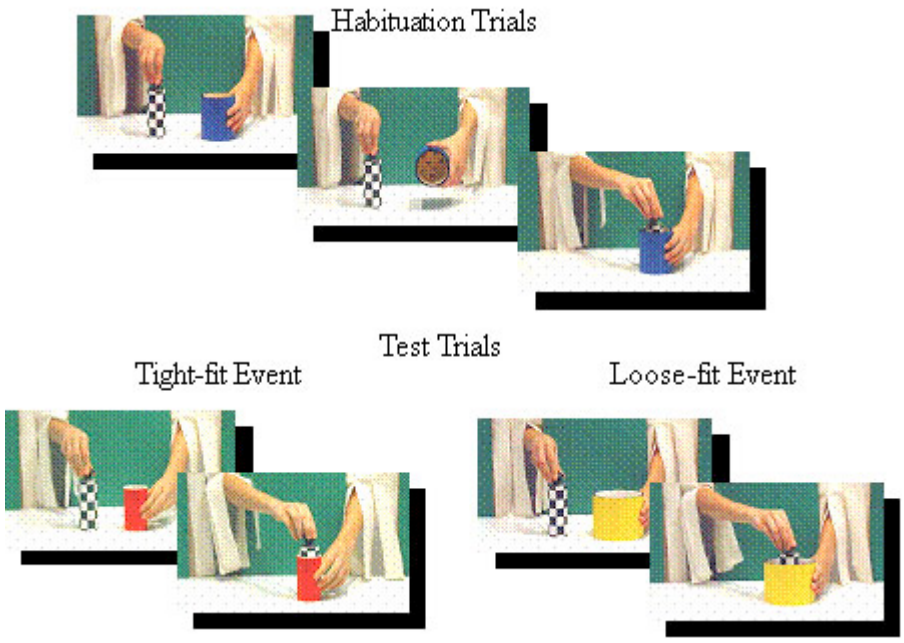


The work of Baillargeon and her colleagues has revealed a reliable pattern in infants' knowledge about physical causality for categories of occlusion, containment, support, and covering events. These findings beg the question of what constitutes and event category boundary? One possible answer begins with the

observation that each of the events studied has a word in English that describes the spatial relationship (occlusion - behind, containment - in, support - on, covering – under). This is not a novel idea to linguists, who know that languages vary in how they describe spatial relationships(Bowerman, 1996; Levinson, 1996; Sinha & Jensen de Lopez, 2000). One well-studied contrast is between English and Korean. For example, when Koreans say that one object joins another, they specify whether the objects touch tightly or loosely. English speakers, in contrast, say whether one object is in or on another. The book is on the table in English and is held loosely by the table in Korean.

Given the cross-linguistic differences, there are two possibilities for how these distinctions emerge. It is possible that infants do not have any categories until after they acquire language and their language creates the initial categories. Alternatively, languages may selectively enhance or diminish distinctions that are already there. Given the vast array of evidence that Baillargeon presents demonstrating that infants have expectations about physical causality as early as 2 months of age, it seemed unlikely that infants wait for language to create the initial categories. We decided to test this empirically by looking at preverbal infants and their knowledge about containment events.

The experiment used 5-month-old infants that were from a monolingual, English-speaking environment. We investigated infants' ability to discriminate a categorical boundary that is captured in Korean but not English. More specifically, we tested infants' categorization of tight-fitting versus loose-fitting containment relations using a habituation-dishabituation paradigm. First, infants saw a narrow cylindrical object lowered into a series of loose-fitting, medium sized containers on a series of trials until their looking time declined (see below). Next, the infants were presented with 6 test trials in which the same cylindrical object was lowered, in alternation, into a wide container (1.5 times wider, hence also a loose fit) and into a narrower container (1.5 times more narrow, a tight fit). If infants made a language-independent categorical distinction between tight- and loose-fitting containment events, then they were expected to look significantly longer at the tight-fit trials. Our results confirmed this prediction (Hespos & Spelke, 2004). Additional conditions replaced the checkered object in habituation trials with a wider object that was a tight-fit with the container and the reverse pattern of looking was revealed in test trials. We concluded that sensitivity to this distinction develops in the absence of any relevant linguistic experience.



A remaining question is whether speakers lose sensitivity to the conceptual distinctions that are not captured by the lexical semantics of their native language. To begin to address this question, we presented the same containment events to English-speaking adults and asked them to rate the similarity between the habituation and test events. In contrast to the infants' patterns of preferential looking, the

adults rated the two test events as equally similar to the habituation event. The adults, therefore, did not appear to make the same categorical distinction as the infants.

If the categories are not coming from linguisitic input, the question is where do these event categories come from? One possibility is that they come from knowledge about physical causality – the mechanical principles that govern how objects behave and interact. Because tight- and loose-fitting containment place different constraints on the motions of objects, it is possible that the principles governing infants' representations of objects and their motions could also lead infants to categorize these spatial relationships differently. In the first experiment, we used a paradigm similar to Baillargeon's experiments described above. First, infants saw a narrow cylindrical object lowered into a wide container until their looking time declined. Next, the infants were presented with 6 test trials that alternated between a move-separately event and a move-together event. In the move-separately event, the cylindrical object was lowered inside the wide container and then the container remained stationary and the object moved back and forth inside the container (expected event). In the move-together event, the cylindrical object was lowered inside the wide container and then both the object and the container moved horizontally as a unitary object (unexpected event). If infants expected the loose-fitting container to allow the object to move with some independence then they were expected to look longer at the move-together event. Our results confirmed this prediction.

In a second condition, we similarly tested infants' expectations for the effects of motion on tight-fitting containment relations. Infants saw the same cylindrical object lowered into a narrow container during the habituation and test trials. In the test trials, infants saw the object inside the container move back and forth horizontally. On alternative trials, the object and container moved together (expected events) or separately (unexpected events). If infants appreciated that the tight-fitting container more strongly constrains the motion of its contained object, then infants were expected to show the opposite looking preference from those in the loose-fitting condition and look longer at the move-separately event compared to the move together event. The results confirmed this prediction, 5-month-old infants have different expectations for horizontal movement in tight- and loose-fit containment.

In summary, this series of experiments suggest an interaction between language and physical causality. Five-month-old infants parse a continuum of the spatial variation into categories of spatial relationships between objects. Infants are sensitive to spatial distinctions that are lexicalized in non-native languages. These findings stress the theme of human universals that underneath all the things that vary across humans, (e.g, the language we speak, the meanings we convey) are a set of perceptual and conceptual capacities that are common to everyone and rich enough to capture the core meanings expressed by any language. The developmental trajectory is one where infants have more flexibility than adults because language has not influenced their causal categories. Taken together, these findings suggest that there is ontogenetic continuity in the development of physical causality but language may alter the category boundaries.

The evidence for ontogenetic continuity complements evidence for phylogenetic continuity in the capacity to represent objects and physical causality. In particular, monkeys represent objects similarly to human infants both in preferential looking and object search tasks (Antinucci, 1990; Hauser, MacNeilage, & Ware, 1996). In fact monkeys progress through Piaget's stage sequences more rapidly than human infants. These findings fit well with the view that basic mechanisms of object representation are consistent over much of evolution and ontogeny, and that their expression depends in part on the developing precision of representations and that this development occurs at different rates for different species (Spelke, 2000).

This paper has focused on evidence for physical causality in human infants. Could the same mechanisms be applied to other domains of knowledge? Most of the infant data pertains to physical causality and investigates expectations about inert objects. However, there is a growing body of research about psychological causality in infants. Psychological causality is distinct from physical causality in that it involves agents (e.g., people or other living things) and the fact that these creatures can have goals and intentions that guide their behavior.

Woodward and her colleagues have a variety of studies that show infants' expectations about people are different than their expectations about objects (Guajardo & Woodward, 2004; Woodward, 1998, 2003). For

example, infants were habituated to a human hand grasping a bear on a stage. The bear was on one side of the stage and a doll was on the other side. In test trials, the position of the bear and the doll were reversed. They measured infants' looking time to events where the goal (grabbing the bear) was the same or the action (grabbing the object in a specific location) was the same. Infants looked significantly longer when the human's goal changed. Interestingly, when the experiment was repeated with a mechanical arm instead of the human arm the looking pattern was the reversed, suggesting that infants expect humans to have goal-directed actions but mechanical arms do not have goal-directed actions (Woodward, 1998). Further experiments demonstrate that infants will link goals with human actions at 7 months of age but that infants do not link eye gaze with goals until 12 months of age (Woodward, 2003). Recent evidence that slightly old infants can use eye gaze information comes from a paper that demonstrates that 15-month-old infants succeed at a modified theory of mind task (Onishi & Baillargeon, 2005).

In conclusion I have given a brief tour of evidence for physical causality in infancy demonstrating that as early as 2 months of age infants have expectations about the way that objects behave and interact. In addition there are context-specific patterns in the acquisition of physical causality. These patterns can help us better understand the fundamental aspects of causal reasoning. The current evidence suggests that there is ontogenetic continuity in our causal categories, this ability may be shared with non-human primates, and language may alter our category boundaries. There is a growing literature on the development of psychological causality. In future research it will be interesting to compare the developmental characteristics of physical and psychological reasoning abilities.

References

Antinucci, F. (1990). The comparative study of cognitive ontogeny in four primate species. In K. R. Gibson, & Parker, Sue Taylor (Ed.), "Language" and intelligence in monkeys and apes: Comparative developmental perspectives. (pp. 157-171). New York, NY: Cambridge University Press.

Baillargeon, R. (2004). Infants' Physical World. Current Directions in Psychological Science, 13(3), 89-94.

Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In (1996). Peterson, Mary A (Ed), et al. Bloom, Paul (Ed), Language and space. (pp.385 436). Cambridge, MA, US: The MIT Press.

Guajardo, J. J., & Woodward, A. L. (2004). Is Agency Skin Deep? Surface Attributes Influence Infants' Sensitivity to Goal-Directed Action. Infancy, 6(3), 361-384.

Haith, M. M., & Benson, J. B. (1998). Infant Cognition. In W. Damon, Kuhn, D., & Siegler, R. S. (Ed.), Handbook of Child Psychology: Cognition, perception, and language (5 ed., Vol. 2). New York: Wiley.

Hauser, M., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. Paper presented at the Proceedings of the National Academy of Sciences of the United States of America.

Hespos, & Baillargeon, R. (2005). Décalage in infants' knowledge about occlusion and contaiment events: Converging evidence from action tasks. Cognition, Manuscript accepted for publication.

Hespos, & Baillargeon, R. (in prep). "Which toy can I get?": Converging evidence from action tasks for violations-of-expectation findings.

Hespos, S., & Spelke, E. (2004). Conceptual precursors to language. Nature, 430, 453 - 456.

Hespos, S. J., & Baillargeon, R. (2001a). Infants' knowledge about occlusion and containment events: A surprising discrepancy. Psychological Science, 121(2), 141-147.

Hespos, S. J., & Baillargeon, R. (2001b). Reasoning about containment events in very young infants. Cognition, 78(3), 207-245.

Levinson, S. C. (1996). Relativity in spatial conception and description. In S. C. Levinson & J. J. Gumperz (Eds.), Rethinking linguistic relativity. New York: Cambridge University Press.

Mandler, J. M. (2004). The foundation of mind: Origins of conceptual thought. New York: Oxford University Press.

Onishi, K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? Science, 308, 255-258.

Piaget, J. (1952). The origins of intelligence in children: Oxford, England: International Universities Press. (1952).

Piaget, J. (1954). The construction of reality in the child. Oxford, England: Basic Books.

Sinha, C., & Jensen de Lopez, K. (2000). Language, culture and the embodiment of spatial cognition. Cognitive Linguistics, 11(1-2), 17-41.

Spelke, E., & Newport, E. L. (1998). Nativism, empiricism, and the development of knowledge. In R. E.

Learner (Ed.), Handbook of child psychology: Theoretical models of human development (Vol. 5th Ed Vol. 1). New York: Wiley.
Spelke, E. S. (2000). Core knowledge. American Psychologist, 55(11), 1233-1243.
Wang, S. h., Baillargeon, R., & Paterson, S. (2005). Detecting continuity violations in infancy: A new account and new evidence from covering and tube events. Cognition, 95(2), 129-173.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition, 69(1), 1-34.
Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. Developmental Science, 6(3), 297-311.

Discussion

An evolved modular program of thought
Eric Baum
May 11, 2005 22:00 UT

The data in Hespos's excellent paper fits nicely within the computational picture discussed in What is Thought?(MIT Press 2004). WIT? models thought as the execution of a hierarchic program built on modules that exploit underlying structure of the world (for example causality) to reason and act. Evolution encoded into the genome programs that interact with sensory data to build such modules. These modules then provide enormous inductive bias that allows further elaboration of the program, building new programs that call previous modules as subprograms. Language mainly involves labelling of preexisting computational modules that exploit real (Platonic) structure.For example, a word like "on" labels/invokes code that has certain properties because it exploits certain real (roughly Platonic) structure. Language, by virtue of allowing newly discovered programs to be communicated, allows cumulative program construction over generations.While apes can discover new meaningful programs over a lifetime, humanity has made cumulative progress in discovering and refining more powerful modules built on top of the submodules coded in the genome.

The results Hespos surveys are consistent with this picture, indicating that low level modules exploiting causality are essentially innate, and the same flow of development is shared with apes. It's interesting, and completely consistent, that some modules for computing and exploiting containment develop in humans a few months slower than modules for computing occlusion.

There is no clear divide between learned and programmed behavior. Consider the "development" of visual cortex. For stereopsis, the width between the eyes must be "learned". It's easy to imagine how the DNA program for this might evolve. The genome mutates, and a program that is effective in the chemical environment of the cells is selected as fit. The chemical environment within neurons, however, depends on the sensory environment. To be fit the DNA must program development so that, in the sensory environment, the circuitry that arises adjusts to the changing width between the eyes. But similar mechanisms can be imagined for "learning" more complex things, such as social behavior, languages, and causal reasoning.

The program underlying thought must be fairly complex, and the big problem is how it can be discovered. If low level modules are genetically encoded in such a way that later meaningful modules can be short chunks of code that call appropriate low level modules, it becomes much easier to discover them, to the point at which the learning may be so biased in that it becomes reliable and fast.

One characteristic that often signals genomic involvement is a critical period, during which the relevant stimulus must be presented if the skill is to be learned, as have been observed for celestial navigation by birds, social behavior in monkeys, and grammar learning in humans. This raises the following (potentially answerable) questions: would a monkey raised from birth in a sensory deprivation tank progress through Piagettian stages, and normal development of understanding of occlusion and containment? If not, would it be able to learn them after being removed at an advanced age? Eric Baum

Could language widen as well as restrain categorization?
Anne Reboul
May 12, 2005 16:22 UT

In her extremely interesting paper, Susan proposes that, though physical causal cognition is universal among humans, some categories present in infancy might disappear or at least be considerably weakened if they're not supported by the language that the child acquires, i.e., if there is no corresponding lexical item in that language. This is an intriguing idea which agrees rather well with observations made about phonological categorical perception in infants: though infants seem to be able to discriminate between any human phonemes, whether or not they belong to their mother tongue, this ability disappears in the process of the acquisition of the phonology of their language. As an example for her hypothesis, Susan gives the case of spatial cognition, the spatial lexicon being quite diversified between different languages (her example is between Korean and English). I would like to propose that language can also do the reverse, i.e. broaden categorization or enhance cognitive abilities, staying in the domain of spatial cognition. In Levinson's work, purporting to show that cognition is linguistically constrained (i.e. not universal), he analyzed the behaviour of speakers of different languages in which the expression of spatial relations rest on different frames of reference (ego-centered, object-centered and absolute = based on the cardinal points, mainly North and South). More precisely, he compared speakers of languages with only the first two frames of reference (e.g. Dutch speakers) to speakers of languages with only the last frame of reference. He found out that speakers of languages with a single, absolute, frame of reference performed very differently from speakers of languages with no such frame of reference at spatial orientation tasks. He claimed that this shows that cognition is constrained by language. Bloom et al., who edited the volume in which the paper was published, pointed out in their conclusion, rightly in my view, that it only showed that when one uses a cognitive ability more or less all the time, that cognitive ability gots more developped. This, of course, agrees with Susan's hypothesis (it does not suppose that the group of speakers of "absolute" languages acquire special abilities, just that some universal ability gets more developped in them that in other groups whose language does not make such demands). However, it is not clear that this would constitute anything like the loss of an ability in parallel with the categorization that Susan discusses. It seems unlikely that infants are very good at cardinal orientation. It might be that, in addition to pruning out unecessary categorization, some languages oblige their speakers to develop capacities — which anyone might develop — which are just not developped by speakers of other languages.

> **Maybe, let's test it!**
> Susan Hespos
> May 17, 2005 22:43 UT
>
> I think that the analogy between phonological and semantic development is rich. There are 4 comparisons that I believe are worthwhile noting. First, our studies suggest that infants can parse a continuum of the spatial variation into categories of spatial relationships between objects. This is similar to the phonological experiments that show infants parse a continuum of acoustic variation into categories of speech sounds (Werker, 1989). Second, our studies suggest that infants are sensitive to spatial distinctions that are lexicalized in non-native languages. Similar to the findings that infants are sensitive to the phonological distinction of non-native languages (Kuhl et al., 1992); Werker, 1991). The third comparison is that speakers seem to lose sensitivity to the spatial categories that are not captured by their language, shown by the fact that adults do not make the same categorical distinction as the infants. The fourth point is to highlight a difference. Intuition suggests a difference between mature auditory and conceptual capacities. In studies of speech perception, adults' recognition of non-native phonological categories may improve with training but rarely attains native facility. In contrast, mature English speakers have little difficulty distinguishing tight-fit from loose-fit categories once these are pointed out, and many English speakers discover the categories on their own.
>
> With this fourth distinction in mind, I concede to Reboul's point. I think it is possible that language (or is it culture?) could create categories that would not emerge otherwise. One example that was described to me by some students at University of Singapore is a Chinese classifier that makes little sense to non-native Chinese speakers. There are animal classifiers 'zhi' and 'tiao'. Apparently tiao

applies to animals with long bodies, but there is no reason why dogs are in the tiao category and cats are in the zhi category when their body proportions are similar.

Reboul's example about cardinal orientation presents an interesting and potentially testable idea. It would be interesting to find out the developmental trajectory of children's use of words referring to the cardinal directions. I predict that given the abstract nature of the concept it would develop later than more perceptually concrete terms (e.g., in and on). I think that if we could come up with a non-verbal test for cardinal direction that children have the concept prior to having the word for it. If anyone can come up with a way of testing this, I'll happily collaborate and test it in my lab.

We are currently testing a similar idea with respect to on/in events. Data from Gentner and Bowerman suggest that there is a developmental decalage in the production of Dutch prepositions for 'in/on' spatial relationships. In Dutch there are three prepositions (op, aan, in). In production aan emerges later than the other two terms (Gentner, personal communication). Currently we are testing the physical knowledge that is captured by these concepts. It will be interesting to see if the developmental order of infants' causal knowledge about these events is parallel to the developmental pattern of production.

Once again we arrive at the question about the relative contributions of language and thought. Research on infants can make a new contribution by characterizing preverbal knowledge to shed new light on an old debate.

Phonology and pruning
Anne Reboul
May 20, 2005 9:07 UT

Following up on the analogy between phonology and pruning, Hauser (1996) quotes studies in phonology which seem to show that if you're exposed to a second language (with significant phonological differences from your mother tongue) during your first year, eventhough you never again hear it afterwards and indeed never get to speak it, you'll keep in adult age (even forty years later) the ability to discriminate between the phonemes of that language, just as you do for the phonemes of your mother tongue, which control subjects which speak only your mother tongue and haven't had such exposure can't do. Now I remember a study by Bowerman and Choi (I think) in the Bloom et al. collection on space to the effect that young Korean children will look longer at some spatial situations which are expressed in their language than do English children in whose language these spatial situations are not encoded. So the question is: could anything similar to the phonological discriminative abilities which remain active described above be found in the spatial domain? For instance, would English-speaking children exposed to both English and Korean during the relevant "sensitive period", but being afterwards raised in a Korean-less environment be able to discriminate without prompting this spatial situation in later years or even adult age?

Hauser, M. (1996) Evolution of communication, Cambridge, MA, the MIT Press. Bloom, P., Peterson, M.A., Nadel L., Garrett, M.F. (1996), Language and Space, Cambridge, MA, The MIT Press.

Agent-based origin of causality
Giyoo Hatano & Kayoko Inagaki
May 20, 2005 14:57 UT

We are pleased that the paper by Susan Hespos discusses infants' naive physics, based on a number of her elegant experiments, particularly because our paper doesn't refer to the experiments dealing with infants' physical knowledge. Causal cognition in naive physics is often represented by pioneering studies led by Spelke and Baillargeon (e.g., Sperber et al., 1995), and these studies are especially informative about the origins, as well indicated by Susan's paper. We would like to take this opportunity to clarify the possible differences in the conception of causality between Susan or infant researchers in general and we investigators of early childhood.

Susan's experiments have clearly demonstrated that infants recognize some events to be 'novel or unexpected.' However, this recognition does not always guarantee that infants' reactions are based on causality, because atypical or unusual events in terms of frequency can also be novel or unexpected. How to differentiate making coherent, reasonable and differentiated predications based on a proper causal device from similarly coherent predictions that are not based on a causal device (or based on an arbitrary cause) is a methodologically hard problem, but one possible way to go is varying the preceding event (PE) in a variety of ways and observing how children's predictions or expectations of the subsequent event (SE) change -- if the prediction for SE stays the same when PE is varied on irrelevant dimensions but the prediction for SE varies when PE is varied on relevant dimensions, we may conclude that the respondent relies on a (natural) cause. The experiments with preschoolers we refer to in our paper surely satisfy this criterion, but how can Susan claim that her experimental findings reveal infants' knowledge about physical causality? Another question concerns the source of physical causality. Susan seems to believe that infants' causal understanding emerges as early as two months, long before they become mobile. Does this mean that physical causality originates only from perceptual experience? We assume that human infants/toddlers come to grasp 'protocausality' when they are serving as a causal agent, in other words, their varied actions produce the correspondingly differentiated changes in spatially connected and temporarily following external entities/events. To put it differently, the subjective basis of causality is the cognition/emotion that PE and SE are intrinsically related, even though it is not possible to specify the relationship. At the very least, as Mandler (2004, p. 101) conjectures, a notion of causal force can be derived not only from perceptual analysis of the transfer of motion between two objects but also from "bodily experiences of pushing against resistance and being pushed." Some studies (e.g., Dan & Omori, 2002) seem to show that infants' motor development leads to the elaboration of physical causality.

References Dan, N. & Omori, T. (2002) Self-sitters know that objects should fall straight down. Paper presented at the International Conference on Infant Studies, Toronto. Mandler, J.M. (2004) The foundations of mind. Oxford University Press. Sperber, D., Premack, D. & Premack, A. J. (Eds.), Causal cognition . Clarendon Press.

Causality in Non-Humans

Jennifer Vonk (Cognitive psychologist, University of Louisiana, Lafayette)

(Date of publication: 23 May 2005)

Abstract : We have recently argued that one fundamental difference between the conceptual systems of humans and other primates may be that humans alone are capable of reasoning about 'imperceptibles', defined as abstract theoretical constructs that can not be directly perceived through the senses. One particular class of 'imperceptibles' is the class of constructs that indexes causal forces. Although non-humans are extremely keen observers of perceptual features in the environment, and appear to understand the role of observable contingencies in the outcome of various events, we argue that they do not posit the existence of causal forces underlying these observable behaviors and events. Non-humans may not reason about 'imperceptibles', (including causal forces), because of a second key distinction between human and non-human conceptual systems; humans alone strive to explain as well as to predict events. This second distinction, which is inextricably tied to the first, may inspire a fruitful avenue of research in comparative psychology.

The discussion presented so far in this conference has nicely articulated what are, in my view, the critical issues surrounding the study of causal understanding in non-verbal animals. Anne Reboul's contribution, in particular, is consistent with my own approach to the problem. The "Unobservability Hypothesis" (Bering & Povinelli, 2003; Povinelli, 2000, 2004; Vonk & Povinelli, in press) addresses a key difference between human and non-human cognition in general that is also relevant for differences between human and non-human causal understanding in particular. If humans alone are capable of reasoning about abstract theoretical entities that cannot be directly perceived through the senses, then non-humans will fail to demonstrate causal understanding because causal forces are unobservable theoretical entities.

The preceding discussion, however, has raised the question of whether all causal forces are necessarily unobservable [1]. If not, it is possible, even if the Unobservability Hypothesis is correct, that some animals might be capable of causal understanding in cases in which the cause of an event can be directly perceived. If we accept this tenuous assertion that causal forces can be both observable and unobservable, a clear prediction can be made from the Unobservability Hypothesis; animals (and probably humans) should perform better on tasks requiring them to reason about the former, compared to the latter, and may not be able to reason about the latter at all. But, is it ever the case that the causal force itself is directly observable? It is clear that we need to carefully define what we mean by 'causal force' and 'causal understanding'. If 'causal force' is defined as a mediating variable between a cause and an effect (c.f. Tomasello, 1998, it is most probably of an unobservable nature. For example, one can directly observe the effect of pouring liquid on to a dry object. It seems reasonable to attribute causal understanding to an individual who appreciates that the liquid causes the dampening of the object, but it is debatable whether the process by which this happens, (diffusion), is considered observable or unobservable. In the case of Call's (2004) experiment, the presence of food is directly observable as the cause of the noise produced by shaking the cup.

Much of human knowledge about causal forces is derived from explicit teachings. If one knows that an ice cube will melt when it is hot does that mean that one understands that the heat causes the ice cube to melt? In order to demonstrate causal understanding, would one need to understand precisely how the heat causes the ice to melt? If that is the case, then humans often fail to demonstrate true causal understanding. How many of us truly understand the principles of gravity or force transfer? How many fewer of us would respond affirmatively to that question without the benefit of explicit tutoring on these subjects? So perhaps simply understanding the direction of the relation between the cause and the effect (it is the heat that causes the ice to melt) should be sufficient to constitute causal understanding. But clearly the representation has to be more than an association between the two (when it is hot, ice melts; when it is cold, ice does not melt). How do we distinguish between these two representations of the event – one that relies on reasoning about underlying causal forces, and one that relies on associating a cause and effect- in a nonverbal being?

It may be impossible to make this distinction both in cases in which the causal force is observable and

unobservable. This is true because it will be exceedingly difficult, if not impossible, to distinguish a representation of the causal forces at work, from an ability to predict effects from causes where the two have been observed to occur together. Even unobservable forces invariably result in observable regularities. These forces and their observable manifestations will rarely lead to contrasting predictions as they are necessarily correlated. So we return to the question posed at the end of the preceding paragraph; how do we disentangle the two possible representations of events? We, as scientists, can create situations in which the two events (the cause and the effect) have not been previously observed to occur together, but should be logically expected to co-occur if one appreciates the causal force at work. The difficulty of course is in providing subjects with enough information that they could conceivably make the logical inference about the causal force.

Elsewhere we have attempted to clearly articulate this empirical challenge with the goal of encouraging the development of clever, new, more diagnostic methods for testing our hypothesis. We stressed the danger of confounding predictions based on observable correlates of underlying causes with predictions based on the cause itself. In addition, we called for an overhaul of the current research paradigms being used to address these issues specifically in the context of theory of mind research (Povinelli & Vonk, 2003, 2004; Vonk & Povinelli, in press). Mental states are only one particular type of unobservable causal forces. In our previous writings we did not outline a clear strategy for tackling the same problems within the context of testing non-humans for their understanding of physical causality. But, perhaps too subtlety, we acknowledged the strength of focusing on explanatory versus predictive paradigms because we believe that one key distinction between humans and non-humans, as relates to causality, is that humans alone strive to explain as well as to predict events (see also Andrews, in press; Bering & Povinelli, 2003; Reboul, this conference, Vonk & Povinelli, in press). If this is true, then non-humans should fail at tasks requiring them to posit explanations for observed events[2]. However, if they show no interest in positing explanations for events, would this disinterest constitute evidence that they are not capable of doing so? A vast array of situations, in which a solution to a problem depends upon an ability to explain effects, must be called upon to address this question.

Clearly animals can reason from observed events (causes) to resulting events (effects), in other words, to make predictions based on observable regularities in their environments. But do they also reason backward from effects to causes? Very little empirical research has directly addressed this question (see Povinelli & Dunphy-Lelii, 2001, for one exception). Indeed, what has been lacking to this point in this conference, although it has certainly been touched upon, is a clear articulation of a program of empirical research designed specifically to address such questions. In the remainder of this brief paper, I would like to suggest some approaches for testing the hypotheses that have influenced the participants of this conference.

Here are the critical questions as I see them:

1. Do humans alone strive to explain as well as to predict the behavior of objects and other beings[3]?

2. Are humans the only species capable of reasoning about strictly theoretical entities (i.e. The Unobservability Hypothesis)?

3. Do all causal forces belong to the category of strictly theoretical, unobservable constructs?

4. If the answer to the first two questions turns out to be affirmative, why did this come to be the case? To what extent do these abilities rely upon the human capacity for language?

5. What advantages are conferred to a species that a) reasons about unobservables, and b) has an explanatory drive?

The latter three questions are theoretical and can not be adequately addressed in this brief exposition (and in the absence of answers to the first two questions). The first two questions are empirical and should be tested. We are pleased that others have embraced the Unobservability Hypothesis as a valid model for framing the uniqueness of human cognition, but we would like to encourage tests of our hypothesis. We would also like to emphasize the utility of explanatory versus predictive paradigms for elucidating non-

human understanding of the causes underlying events. I will focus on the latter hypothesis as it more specifically addresses the question of representations of causality in non-humans.

Researchers have attempted to address non-humans' understanding of causality a number of ways. Most notably, they have tested other species' (mostly primates') use of tools, and examined whether tool use reveals an understanding of the causal structure of a task. As with many areas of research, there is some dispute as to the extent to which apes, but not monkeys, grasp the causal significance of various attributes of the given tasks (Boesch & Boesch, 1990; Cacchione & Krist, 2004; Fujita, Kuroshima & Asai, 2003; Hauser, 1997; Hauser, Kralik & Botto-Mahan, 1999; Hauser, Pearson & Seelig, 2002; Hauser, Santos, Spaepan & Pearson, 2002; Kralik & Hauser, 2002; Kohler, 1925; Limongelli, Boysen & Visalberghi, 1995; Matsuzawa, 1996, 2001; Povinelli, 2000; Santos & Hauser, 2002; Santos, Miller & Hauser, 2003; Visalberghi, 1997, 2002; Visalberghi, Fragaszy & Savage-Rumbaugh, 1995; Visalberghi & Limongelli, 1994, 1996; Visalberghi & Tomasello, 1998; Visalberghi & Trinca, 1989). However, an extensive research program in our own laboratory has given us reason to doubt that chimpanzees appreciate various causal forces underlying folk physics, such as physical connection, gravity, rigidity, etc. (Povinelli, 2000). In these tasks, chimpanzees were presented with problems that they should have been able to solve from the very first trial forward if they had an accurate representation of the causal forces at work in the task. For instance, if they understood the effect of gravity they would not attempt to retrieve a food reward by dragging it across an empty space in a "trap-table", and yet all but one subject did so just as often as they chose to drag the food across an unbroken surface (Povinelli, 2000, Chapter 5, Experiment 3).

Recently, Josep Call (2004, in press, and see his comments in response to Anne Reboul's paper) found that apes chose cups containing hidden food more accurately when cues as to the food's location were of a causal versus an arbitrary nature. (Although please see Anne Reboul's comments on why these experiments do not address whether apes are capable of diagnostic versus predictive learning). Let us conduct a thought experiment in order to illuminate some of the issues raised by this sort of paradigm. Suppose an ape is presented with two identical liquid containers, the contents of which are hidden. An experimenter moves the two containers to two new locations. The ape is required to point to the container that he wants the experimenter to give to him. Only one of the containers has left a water mark in its former resting place. If the ape points to the container that formerly held the position where now there is a water stain, is this because he has evoked an explanation for the stain – that it must be caused by condensation caused by the presence of liquid in that particular container? Or is he predicting that liquid will be found (or is more likely to be found) in the presence of containers that leave marks behind, versus those that do not? Of course the ape would have to have had the relevant experience with such containers – those that are full and leave stains, and those that are empty and do not, in order to make both the non-causal prediction and the causal inference/explanation.

In a related vein, others have sought to determine whether chimpanzees imitate arbitrary, irrelevant actions as well as causally relevant actions after watching a demonstrator retrieve food from a closed container (Horner & Whiten, 2005). The basic idea is that, if animals are privy to the mechanism in a mechanical task, they should learn faster, or at least, fail to imitate irrelevant actions, compared to conditions in which they are not privy to the mechanism involved. Although the experiment is clever, and the results intriguing, there are some flaws in the design that prevent us from accepting the authors' conclusions that access to causal knowledge allowed the apes to ignore irrelevant actions on the box. For instance, it is possible that the chimpanzees performed the relevant actions, because they were directed toward the food reward, and, in the 'causal' condition, they were aware of the food's location in the box. They may have failed to imitate irrelevant actions only because they were not directed towards the food's location. We do, however, believe that the experiment could be modified to answer the kinds of questions we have posed here.

Researchers have also made extensive use of the violation of expectancy paradigms first applied to the study of human infants (Baillargeon, Spelke & Wasserman, 1985; Leslie, 1982; Spelke, 1985). This paradigm can be exploited to determine what nonverbal subjects understand about the likelihood of various events, by measuring whether they look longer at events that violate natural laws (Gergely, Nadasdy, Csibra & Biro, 1995; Hauser, 1998; Hauser & Carey, 1998; O'Connell & Dunbar, 2005). If they look longer at outcomes that would only be surprising given an understanding of the causal force at work, then we might conclude that they make predictions based on an understanding of that causal force. Of

course it is possible that an event violates an individual's expectations, not because they have a deep understanding of the causal forces at work, but, instead because they have gained, through previous experience, an appreciation of the likelihood, or commonality of various events. This problem can be easily overcome by designing scenarios in which the unexpected events in a particular context are generally as common as the expected events. This paradigm could possibly be recruited to examine animals' diagnostic abilities. Imagine a procedure whereby subjects are shown first the final result of an event, and subsequently were shown the preceding action under conditions in which the action could and could not have resulted in the observed end result. For example, in a physical causality study, if a collision has just occurred; can the subjects guess as to which object must have been used to strike another, based on the resulting impact and end state of the objects involved? So, can an object's current resting position be used as a cue as to which event previously occurred in both collision and deformation paradigms?

Only one study, to our knowledge, has directly tested whether non-humans seek causal explanations (Povinelli & Dunphi-Lelii, 2001). This study was briefly described in Anne Reboul's paper so will not be reiterated here. We recommend more experiments along these lines. These experiments will be most informative when they measure subjects' attempts to seek explanations for unexpected events, especially when an event violates expectations based on presumed causal understanding. For instance, our lab has begun to investigate chimpanzees' reactions to apparent "magical" (arbitrary) causality. For example, chimpanzees might be trained to associate arbitrary cues with the functionality of a tool. Once they have learned to associate such cues with success in a food retrieval task we could reverse the contingencies such that the cues previously associated with the 'correct' or 'functional' end of a tool are now associated with the 'incorrect' end and vice versa. If chimpanzees are surprised by this unexpected change one might expect them to examine the tools to look for some anomaly that would explain their failure in this task.

Some other contexts in which this experimental approach might be appropriate are as follows: subjects have been misled by cues provided by experimenters as to the location of hidden food – do they check back to make sure they read the cue correctly? Do they look for signs of an event that would explain another's actions? For example, do they search for a predator or another alarming object if another animal has just displayed a fearful reaction? Can animals guess what another animal has just done from cues as to a preceding act as easily as they can predict what another animal will do? In other words, one could contrast subjects' ability to explain versus predict another's actions. Do animals question why another individual reacted in the manner that they did, particularly when this reaction was directed towards the subject in question? Humans are very sensitive to behaviors directed towards themselves. If someone appears to be angry or fearful towards us we wonder why. We often analyze our own behavior in an attempt to understand whether we are to blame for this unexplained reaction. If we recognize some component of our behavior that may have precipitated the aversive behavior of another, we typically attempt to change our own behavior in an effort to avoid the aversive reaction of the other individual. Do animals adjust their own behaviors accordingly? If they do so, can we conclude that they have "explained" the source of the aversive behavior? If the individual's changed behavior is a direct result of the realization that his own behavior caused the behavior of another, we can attribute causal understanding to that individual. But how can we determine whether the change came about as a result of reasoning about the cause of the behavior, as opposed to reasoning in a predictive manner about how others will react to one's own behavior? Can we construct an empirical test of this question?

Imagine that two chimpanzees are placed in adjacent enclosures. One chimpanzee (the actor) has the opportunity to choose between two food trays – however, he can not see the contents of either of the trays prior to making his choice. Prior to the choice, the chimpanzees have been trained to understand that only one tray contains a food reward on every trial. The first tray chosen by the chimpanzee will be offered to the neighboring chimpanzee (the partner). The actor will not be able to see what was in the tray given to the partner, but will be able to see the partner's reaction. Based on the reaction of the partner the chimpanzee should be able to determine whether the chosen tray contained a food reward, and to deduce that, if the partner received a reward, the remaining tray must be empty. Therefore there would be no point to making an additional choice. However, if the partner does not appear to have received a reward, the baited tray must remain available. Therefore it would benefit the actor to make a second selection, because the actor is always offered the reward, if present, from the second tray chosen. The frequency of second selections should vary according to whether the partner expresses pleasure or displeasure. If the

actors consistently choose a second tray only when the partners do not receive a food reward, does this mean that they have deduced the contents of the chosen tray from the partner's expression? I suspect, based on Anne's summary of Cheney and Seyfarth's (1990) findings that vervets do not infer the presence of predators from observable evidence as to their whereabouts (for example, the tracks of a python, or the carcass from a cheetah's kill etc.) that monkeys will not demonstrate evidence for seeking causal explanations. It is possible that apes will differ on these measures, although Povinelli & Dunphy-Lelii's (2001) findings suggest otherwise.

Data from the numerous animal language projects might get us closer to this goal of evaluating animals' diagnostic abilities. Because there is some evidence that sign-language-trained chimpanzees, at least, answer "wh" questions appropriately (Van Cantford, Gardner & Gardner, 1989; Gardner, Van Cantford & Gardner, 1992), these chimpanzees might to some degree understand the meaning of the signs for such questions. If there is evidence that they utilize these same signs to ask questions of others, particularly "how" or "why" questions, we would be very interested. Of course such a finding should not be taken to imply that only language-trained animals are capable of, or interested in, searching for explanations, or understanding causality[4]. It may be that only language allows a creature to express such a drive in a manner that we, as humans, can interpret as such. This circularity is precisely why we need to invent diagnostic tests for studying causal understanding in creatures who can not express themselves through language.

Within all of these experiments, it is possible to contrast cases in which the cause can be described as emanating from an unobservable process (such as transfer of force, gravity, beliefs, intentions) versus an observable event (a baited cup makes a sound, an unbaited cup does not, as in Call, 2004). It should be clear from these examples that there is much work to be done before we can draw any final conclusions as to which other species might share our desire to explain events, and which are capable of reasoning about causal forces, either observable or not, in order to do so. Based on the data so far, it appears that only humans reason about causal forces that are not directly observable, while other species appear to reason solely about the observable properties of events and object interactions. It also seems to be the case that non-humans reason about causes only insofar as to predict the actions of others and objects in their environments, and are not driven to explain events that have already occurred.

References

Andrews, K. (in press). Chimpanzee theory of mind: Looking in all the wrong places? Mind and Language.

Baillargeon, R., Spelke, E.S., & Wasserman, S. (1985). Object permanence in five- month-old infants. Cognition, 20, 191-208.

Bering, J.M. & Povinelli, D.J. (2003). Comparing cognitive development. In D. Maestripieri, Ed. pp. 205-233. Primate psychology: Bridging the gap between the mind and behavior of human and nonhuman primates. Cambridge, MA: Harvard University Press.

Call, J. (2004). Inferences about the location of food in the great apes. Journal of Comparative Psychology, 118, 232-241.

Call, J. (In press). Descartes' two errors: Reason and reflection in the great apes. In S.
Hurley and M. Nudds (Eds.) Rational Animals? OxfordUniversityPress.

Cheney, D.L., & Seyfarth, R.M. (1990). How Monkeys See the World: Inside the Mind of Another Species.Universityof ChicagoPress, Chicago.

Boesch, C. & Boesch, H. (1990). Tool use and tool making in wild chimpanzees. Folia Primatologica, 54, 86-99.

Cacchione, T. & Krist, H. (2004). Recognizing impossible object relations: Intuitions about support in Chimpanzees (Pan troglodytes). Journal of Comparative Psychology, 118, 140-148.

Fujita, K., Kuroshima, H. & Asai, S. (2003). How do tufted capuchin monkeys (Cebus apella) understand causality involved in tool use? Journal of Experimental Psychology: Animal Behavior Processes, 19, 233-242.

Gardner, R. A., Van Cantford, T.E., & Gardner, B.T. (1992). Categorical replies to categorical questions by cross-fostered chimpanzees. American Journal of Psychology, 105, 27-57.

Gergely, G. Nadasdy, Z., Csibra, G. & Biro, S. (1995). Taking the intentional stance at 12 months of age. Cognition, 56, 165-193.

Hauser, M.D. (1997). Artifactual kinds and functional design features: What a primate understands without language, Cognition, 64, 285-308.
Hauser, M.D. (1998). A nonhuman primate's expectations about object motion and destination: The importance of self-propelled movement and animacy. Developmental Science, 1, 31-37.
Hauser, M.D. & Carey, S (1998). Building a cognitive creature from a set of primitives: Evolutionary and developmental insights. In D. Cummins & C. Allen (Eds.) The Evolution of Mind (pp. 52-106). Oxford, England, OxfordUniversityPress.
Hauser, M.D., Kralik, J. & Botto-Mahan, C. (1999). Problem solving and functional design features: Experiments on cotton-top tamarins (Saguinus oedipus). Animal Behaviour, 57, 565-582.
Hauser, M.D., Pearson, H.M. & Seelig, D. (2002). Ontogeny of tool-use in cotton-top tamarins (Saguinus oedipus): Innate recognition of functionally relevant features. Animal Behaviour, 64, 299-311.
Hauser, M.D., Santos, L.R., Spaepen, G.M. & Pearson, H.E. (2002). Problem solving, inhibition and domain-specific experience: Experiments on cottontop tamarins, (Saguinus oedipus). Animal Behaviour, 64, 387-396.
Horner, V. & Whiten, A. (in press). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). Animal Cognition. Published online Nov. 11, 2004. DOI: 10.1007/s10071-004-0239-6
Kohler, W. (1925). The Mentality of Apes. Liveright, New York.
Kralik, J.D. & Hauser, M.D. (2002). A nonhuman primates' perception of object relations: Experiments on cottontop tamarins, Saguinus oedipus. Animal Behaviour, 63, 419-435.
Leslie, A.M. (1982). The perception of causality in infants. Perception, 11, 173-186.
Limongelli, L., Boysen, S.T. & Visalberghi, E. (1995). Comprehension of cause-effect relations in a tool-using task by chimpanzees (Pan troglodytes). Journal of ComparativePsychology, 109, 18-26.
Matsuzawa, T. (2001). Primate foundations of human intelligence: A view of tool use in non-human primates and fossil hominids. In: Primate Origins of Human Cognition and Behavior (Ed. by T. Matsuzawa). Pp. 3-25. Tokyo, Springer-Verlag.
O'Connnell, & Dunbar, R.I.M. (2005). The perception of causality in chimpanzees (Pan spp.). Animal Cognition, 8, 60-66.
Povinelli, D.J. (2000). Folk physics for apes: The chimpanzee's theory of how the world works. Oxford: OxfordUniversity Press [Reprinted with revisions, 2003].
Povinelli, D.J. (2004, Winter). Behind the ape's appearance: Escaping anthropocentrism in the study of other minds, Daedalus, 29- 41.
Povinelli, D.J. & Dunphy-Lelii, S. (2001). Do chimpanzees seek explanations?Preliminary comparative investigations. Canadian Journal of Experimental Psychology, 55,93-101.
Povinelli, D.J. and Vonk, J. (2003). Chimpanzee minds: Suspiciously human? Trends in Cognitive Science, 7, 157-160.
Povinelli, D.J. and Vonk, J. (2004). We don't need a microscope to explore the chimpanzee mind. Mind and Language, 19, 1-28.
Santos, L.R. & Hauser, M.D. (2002). A non-human primate's understanding of solidity:
Dissociations between seeing and acting. Developmental Science, 5, F1-F7.
Santos, L.R., Miller, C.T. & Hauser, M.D. (2003). Representing tools: How two non-human primate species distinguish between the functionally relevant and irrelevant features of a tool. Animal Cognition, 6, 269-281.
Spelke, E.S. (1985). Preferential looking methods as tools for the study of cognition in infancy. In G.Gottlieb & N. Krasnegor(Eds.) Measurement of audition and vision in the first year of postnatal life (pp. 85-168). Norwood: Ablex.
Tomasello, M. (1998). Uniquely primate, uniquely human. Developmental Science, 1, 1-16.
Van Cantford, T., Gardner, B.T. & Gardner, R.A. (1989). Developmental trends in replies
To Wh-questions by Children and Chimpanzees. In R.A. Gardner, B.T. Gardner, & T. E. Van Cantford (Eds.) Teaching sign language to chimpanzees, (pp. 198-239). Albany, NY: SUNY Press.
Visalberghi, E. (1997). Success and understanding in cognitive tasks: A comparison between Cebus apella and Pan troglodytes. International Journal of Primatology, 18, 811-830.
Visalberghi, E. (2002). Insight from capuchin monkey studies: Ingredients of recipes for, and flaws in capuchins' success. In Bekoff, M. & Allen, C. (Eds.) The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition. MIT Press, Cambridge, MA, pp. 405-411.
Visalberghi, E., Fragaszy, D.M. & Savage-Rumbaugh, S. (1995). Performance in a tool-using task by common chimpanzees (Pan troglodytes), Bonobos (Pan paniscus), an Orangutan (Pongo pygmaeus),

and capuchin monkeys (Cebus apella). Journal of Comparative Psychology, 109,52-60.
Visalberghi, E., & Limongelli, L. (1994). Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (Cebus apella). Journal of Comparative Psychology, 108, 15-22.
Visalberghi, E., & Limongelli, L. (1996). Acting and understanding: Tool use revisited through the minds of capuchin monkeys. In Russon, A.E, Bard, K.A. (Eds), Reaching Into Thought: The Minds of the Great Apes. New York, Cambridge University Press.
Visalberghi, E. & Tomasello, M. (1998). Primate causal understanding in the physical and psychological domains. Behavioral Processes, 42, 189-203.
Visalberghi, E., & Trinca, L. (1989). Tool use in capuchin monkeys: Distinguishing between performing and understanding. Primates, 30, 511-521.
Vonk, J. & Povinelli, D.J. (in press). Similarity and difference in the conceptual systems of primates: The Unobservability hypothesis. In E.Wasserman and T. Zentall (Eds.) Comparative Cognition: Experimental Explorations of Animal Intelligence.

---

[1]In addition, of course, not all unobservables are causal forces.

[2]Although see comments by Hatano & Inagaki, on March 8, as a response to Anne Reboul. They point out that subjects might express through prediction, an understanding of the causal attributes of a task, which might not be expressed if they are required to explain the task.

[3]Related to this point, humans may be the only species to seek explanations for their own thoughts and behaviors; perhaps by virtue of being the only species endowed with metacognitive abilities (that is, the ability to reflect upon our own thoughts).

[4]Although we did suggest that the ability to represent unobservables may have co-evolved with natural language (Vonk & Povinelli, in press), we remain agnostic as to the extent that one depends upon the other.

Discussion

Causal / non-causal, or rather backwards / forwards?
Teresa Bejarano
May 25, 2005 10:28 UT

*Non-causal prediction vs causal explanation: Vonk focuses on the difference between these two sides. In my view, each side is too complex: Non-causal AND forwards, causal AND backwards. Which feature is the crucial one? I will suggest that forwards/backwards is the crucial one. *Causality ties events in the world: One event -effect- follows another -cause-. But animal learning ties perceptions. In animal learning, the relevant perception (outcome, 'unconditioned stimulus') follows the non-relevant perception (cue). This is the adaptively advantageous order. It guides the immediate behaviour. This is why predictive association -learned expectation- is enough for animals. *However, it is not sufficient for some tasks. Why can humans understand the track of pythons? They perceive a python. Later, when there is no python in that place, they see the track. The track can be interpreted because the first, outdated perception is retrieved. Let us compare Call's results and Vonk's thought experiment. Call's apes had learned that noise/absence of noise in a container precedes the presence/absence of content in the container. This is the typical animal order. I would accept that that previous experience involves causal understanding. But if so, this causal understanding would not require any high-level cognitive process: No outdated perception was retrieved. But let us analyse the necessary previous learning with Vonk's containers. One container is perceived; later, the container is seen in a different place, and a water mark appears in the former place. The water mark will be understood if and only if the outdated perception is retrieved. Apes would fail because they only pay attention to updated news. The former place has got out of apes' awareness. *Causal understanding would be either a too low-level, non-exclusively human, capacity, or (if we require

excessively stringent criterions: Hatano&Inagaki, Mar 8) a too high-level, non-universal, human capacity. What about outdated perceptions? Certainly, there are differences between them and false beliefs. Verbal past tense turns an outdated perception into a true sentence. However, outdated perceptions and own past false beliefs are on the same level (Riggs&Simpson, 2005, Bejarano, 2003). This is an adequately high level (Deceptive Box Task). Likewise, retrieval memory may be an exclusively human memory (Suddendorf&Busby, 2003). Thus, awareness of outdated perceptions may be an exclusively human resource. *This awareness is required for some tasks that don't involve any causal understanding. In Buytendijk's experiment, animals are faced with a row of containers and they see the bait being placed in the first one. Animals will rush to this container. But they are shown that the bait is in the second one, etc. Why are animals unable to discover the rule 'present container = previous container + 1'? They are apparently unable to single out the current container in terms of the outdated perception. *Causal understanding and forwards; non-causal understanding and backwards. Vonk's complex sides have been disentangled. Which feature is the crucial one?

Forwards/Backwards versus Causal/Non-Causal and Oudated Perceptions: Reply to Bejarano

Jennifer Vonk
Jun 3, 2005 2:51 UT

I do not agree that causality can be completely disentangled from the directionality of an association within the context of this discussion. Clearly there are “forwards” associations that are causal and those that are not, and the same is presumably true for “backwards” associations. Associations that are entirely arbitrary can become predictive, in both directions, based solely on temporal and spatial contiguity. The ability of non-verbal beings to learn these kinds of associations would not reveal much about their causal understanding. My point was that the ability of animals to seek causal explanations seems to be a necessary precursor to the ability to reason causally – for if they are not attempting to explain the “how” and “why” of events they can not be evoking causality, and might be simply forming associations about observable regularities. It is possible to argue for simple associations even in the case of backwards reasoning; however the presence of backwards reasoning might be a basic building block for the ability to seek causal explanations. Finding evidence for backwards reasoning is thus a first step in determining whether animals might be capable of causal reasoning, but it remains to disentangle the extent to which the reasoning evoked true causal forces rather than calculations regarding the likelihood of different temporal patterns of events.

I think that Teresa’s focus on outdated perceptions is an interesting one and one that could be very fruitful for future research. She states that “retrieval memory may be an exclusively human resource”. This is an empirical question, which I think has yet to be resolved. Indeed I think there is evidence to the contrary; clearly animals show savings in the learning of rules, categories, motor skills, social ranks etc. The question is - to what extent do they consciously reflect on what they have learned previously? Innovative studies by Smith and colleagues (Smith, Shields & Washburn, 2003), Inman and Shettleworth (1999), Son, & Kornell (in press) and Hampton (2001) have indicated some (arguably) metacognitive abilities in animals ranging from pigeons to dolphins to monkeys. Do these and other species have thoughts about things that no longer exist? Evidence from object displacement and hidden food tasks implies that they do. Also, studies by Menzel (1999) and Beran and Beran (2004) have indicated that apes report on the whereabouts of objects that are not currently visible. These findings imply that these apes do think about and consciously act out desires for items that are not currently being perceived. Again it is necessary to clearly distinguish between unobservables in the sense intended by Povinelli and myself from things that are not currently being perceived but have been perceived or could conceivably be perceived. We have never suggested that non-human primates are incapable of representing events that are not currently being perceived. I would agree that such representations are necessary for causal reasoning and I do not think that there is any evidence to suggest that non-humans lack the ability to hold such representations, thus this line of thinking does not preclude their ability to reason causally.

Second Reply: Memory and Causal Understanding

Jennifer Vonk
Jun 3, 2005 2:52 UT

Much research in the area of human memory has focused on the dissociation between conscious and automatic retrieval processes, or the extent to which memory retrieval is an explicit, consciously controlled process. The investigation of this topic has unfortunately not yet been extended to the study of animal memory. I assume that it is to this aspect of memory that Teresa refers when she states that retrieval memory is exclusive to humans. This is a question that I intend to pursue in the future.

Further, Teresa's commentary brings to mind another important distinction, that between memory for things that no longer exist versus thoughts or beliefs that are no longer held to be true. Do animals represent their own past thoughts? Is it necessary for them to do so to demonstrate causal understanding? I do not believe that it is. I think it would be sufficient for them to reason about past objects and events that they have perceived.

Lastly, as for Teresa's description of Buytendijk's experiment, I am not familiar with this study but I do not think that this result demonstrates that the animals do not hold the outdated perception. In contrast, I think that it demonstrates that, when the outdated perception exists, the animals actually perseverate on it. Their difficulties in this procedure seem likely to be problems of inhibition, not of memory.

Thus, I think an exploration of non-humans' abilities to maintain outdated perceptions may be critical for elucidating the extent to which animals reason causally, but I do not think the former ability would be sufficient for expression of the latter, and I do not think it is a foregone conclusion that it is beyond their ken.

Beran, M. J. & Beran, M. M. (2004). Chimpanzees Remember the Results of One-by- One Addition of Food Items to Sets Over Extended Time Periods. Psychological Science, 15, 94-99.

Hampton, R. R. (2001). Rhesus monkeys know when they remember. Proceedings of the National Academy of Science, 98, 5359-5362.

Inman, A. & Shettleworth, S. J. (1999) Detecting metamemory in nonverbal subjects: A test with pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 25, 389-395.

Menzel, C. R. (1999) Unprompted recall and reporting of hidden objects by a chimpanzee (Pan troglodytes) after extended delays. Journal of Comparative Psychology, 11, 426-434.

Smith, J. D, Shields, W. E. & Washburn, D. A. (2003) The comparative psychology of uncertainty monitoring and metacognition. Behavioral & Brain Sciences, 26, 317-373.

Son, L. K. & Kornell, N. (in press) Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In H. Terrace & J. Metcalfe (Eds.) The missing link in cognition: Origins of self-knowing consciousness. Oxford U. Press

Continuation. About inobservability, agency and outdated information

Teresa Bejarano
Jun 3, 2005 12:21 UT

1) Inobservability. We pour liquid on to a particular abject. What causes the dampening of the object? According to Vonk, the true cause is not the poured liquid on that particular object but the diffusion. In my view, both alternatives are descriptions of the same fact. More and more profound redescriptions can be achieved. In the end, the most profound one would involve the history of the

universe. But a more profound level does not go with a more true causality. 2) Agency. A relevant perception gives relevance to the immediately previous 'action and environment'. This is the basic loop of the animal learning. This circularity -in the first episode, diagnosis, and in the following ones, expectation- was highlighted by Piaget. In the evolutionary and developmental beginning, this diagnostic understanding, this sense of agency, is primarily egocentric. Later somebody else's agency can be understood. Apes may share this ability with humans. Call's containers are shaken by the experimenter. But, since animals would have previously learned that when they shaked a full (empty) container they caused (did not cause) a noise, they can transfer this knowledge to somebody else' agency. This ability, however, would reach a very particular intensity in humans. Heterochrony is probably a resource that strengthens this particular intensity. Without locomotion, the baby concentrates on the activities that he can do. He will look at people and will ask them for many things. So, alien agency will be given a lot of attention. 3) But what is the crucial difference between human and non-human causal understanding? In animal learning the relevant perception must follow the non-relevant one. If this order is reversed, the basic loop will not longer operate. In these circumstances, learning will arise if and only if the relevant (but outdated!) perception is artificially brought into the awareness. In my view, this artificial recovery -this retrieval- is as demanding as the ability of remembering own past false beliefs. An outdated perception, i.e. a perception that has been refuted by the following ones, is very similar to a past false belief. Without the adequate analysis in each case, outdated perceptions and past false beliefs may be hardly distinguishable. Consequently, some well-known conclusions of ToM can explain the exclusively human type of causal understanding. Animals are able to return from a relevant situation to that action of theirs that has caused the relevant situation. They return "from effects to causes". This is a diagnostic causal understanding. But animals cannot retrieve an outdated perception or a frustrated expectation of theirs. When "one block cannot be made to stand", perceptual updating replaces the frustrated expectation. When an information has been refuted, it immediately goes out of animal awareness. Humans, on the other hand, can handle outdated, refuted information in addition to their current awareness.

The bibliographic references of my previous message
Teresa Bejarano
Jun 8, 2005 6:22 UT

The Buytendijk Task. This old experiment can be read in Luria, 1981 (English translation), chapter 1. In my previous unskilful description, it might sound like the A-not-B error. But it is radically different. Subjects are faced with a row of containers and they see the bait being placed in the first one. Later on, without the subject knowing, the bait is placed in the second container. Of course, the subject will go the first container in search of the bait. After his failure to find it, the subject is shown that the bait is in the second container. Put again to the test, the subject looks for the bait in the second container. Once again, the subject fails and is shown the bait in the third one. This is carried out a number of times, always placing the bait in the next container in the row. While animals fail again and again, 3-year-old children discover the rule soon (and with laughter, according to my experience). 'Search for the reward in the container next to the one where you found it last time': What does this discovery involve? How has the child got it? When, after some failure of his, the child is shown that the reward is in a container, he must reformulate this container in terms of his immediately past belief or frustrated expectation. 'My immediately past false belief + 1= The currently full container'. No causal understanding is required in this task. The awareness of outdated perceptions or past false beliefs is the key element in the Buytendijk task. So, some questions arise. Do animals lack that awareness? If so, can this unawareness explain animal inability to understand a particular type of causal understanding?. Bejarano, T. (2003). Metarepresentation and human capacities. Pragmatics and cognition, 93-140. Luria, A. R. (1981). Language and Cognition. New York. J. Wiley. Riggs, K. & Simpson, A. (2005). Young children have difficulty ascribing true beliefs. Developmental Science, F27-30.

Thanks for Info!
Jennifer Vonk
Jun 9, 2005 14:28 UT

Thanks Teresa for the additional information on this very interesting study! We are currently very interested in the rules that animals can learn, by observation or otherwise. This seems like a very valuable research design to explore!

Unobservability and causality
Anne Reboul
May 30, 2005 9:38 UT

I think that the unobservability hypothesis is a major advance for making sense of the cognitive differences between nonhuman animals (more specifically primates) and humans. Thus I am in complete agreement with Jennifer and am interested in her discussion of how to test the hypothesis. I would just like to shortly discuss the suggestion at the end of Jennifer's paper that data from animal language projects could help "evaluating animals' diagnostic abilities". Though this may be the case, I think that the data presently available is not encouraging and I will quickly say why I think it is not. One of the main differences between human and chimpanzee use of language seems to lie in displacement, the ability of speaking of something which not only is not directly perceived but may even not exist and to do so not in the context of a request but in the context of an assertion (indeed, the production of assertions by "talking" chimpanzees seems to be extremely restricted even when compared with very young children). Now, I think that there is a strong link between displacement and the conception of unobservables, mainly because displacement, which is a divorce between directly available perceptual information and the representation of non directly available perceptual information (including non perceivable individuals) allows ipso facto the representation of absent, inexistant or unobservable entities. Displacement is one of the hallmarks of linguistic use in humans but it is one of the central features of language (the other one being syntax) that great apes do not seem to master. Thus, if displacement is absent in nonhuman primates involved in animal language projects, it is doubtful that they could have diagnostic abilities in causal reasoning if causal reasoning includes reasoning about unobservables. What would be left would be causal reasoning about observables, but, as Jennifer says, it is not clear that causal reasoning stricto sensu ever is about observables.

Displacement and Outdated Perceptions: Reply to Anne Reboul
Jennifer Vonk
Jun 3, 2005 3:31 UT

Anne’s comment ties in nicely to the discussion raised by Teresa Bejarano. Anne points to the failing of “language-trained” apes to demonstrate displacement. I think it is interesting that apes do not comment about things not presently perceived because, as I indicated in my response to Teresa, I believe that animals are capable of representing objects and events not currently perceivable. It is possible that there is a dissociation between what they are capable of representing and what they choose to communicate about – although it is not immediately clear why this should be the case. So while I agree that the lack of displacement in animal communication is telling, I do not think that it is necessarily strong evidence for a failure of causal reasoning. However, even if animals were to communicate about events that they aren’t currently perceiving, this ability would not necessarily lay the foundation for representing or communicating about events that are not perceivable in the strict, theoretical sense. Thus, while representing and reasoning about past events and absent objects may be required for causal reasoning it may not be sufficient if causal forces are strictly unobservable. Therefore, even if there WAS evidence of displacement, it would not necessarily strongly affirm the presence of causal reasoning. However, I think if animals were asking “why” or “how” questions it would at least affirm an explanatory drive, which may lie at the root of causal understanding.

Displacement and Outdated Perceptions: Reply to Anne Reboul

Jennifer Vonk
Jun 3, 2005 20:22 UT

Anne's comment ties in nicely to the discussion raised by Teresa Bejarano. Anne points to the failing of "language-trained" apes to demonstrate displacement. I think it is interesting that apes do not comment about things not presently perceived because, as I indicated in my response to Teresa, I believe that animals are capable of representing objects and events not currently perceivable. It is possible that there is a dissociation between what they are capable of representing and what they choose to communicate about – although it is not immediately clear why this should be the case. So while I agree that the lack of displacement in animal communication is telling, I do not think that it is necessarily strong evidence for a failure of causal reasoning. However, even if animals were to communicate about events that they aren't currently perceiving, this ability would not necessarily lay the foundation for representing or communicating about events that are not perceivable in the strict, theoretical sense. Thus, while representing and reasoning about past events and absent objects may be required for causal reasoning it may not be sufficient if causal forces are strictly unobservable. Therefore, even if there WAS evidence of displacement, it would not necessarily strongly affirm the presence of causal reasoning. However, I think if animals were asking "why" or "how" questions it would at least affirm an explanatory drive, which may lie at the root of causal understanding.

Observable cues to unobservable causality

Leyre Castro
Jun 1, 2005 0:16 UT

Causal relationships cannot be directly observed—they must be inferred. This inference is always based on observable events. Hume listed several well-known primary cues to causality: 1) spatiotemporal contiguity, 2) priority, and 3) consistent conjunction. He also mentioned other cues that would better pinpoint causality: 4) where several different objects produce the same effect, it must be by means of some quality common amongst them, and 5) the difference in the effects of two objects must proceed from that in which they differ. Humans evaluate all of these cues to determine whether a causal-effect relationship actually exists. If animals' behavior is also affected by the same cues, then the most parsimonious interpretation is that animals are also capable of causal understanding. Contiguity, priority, and contingency correspond exactly to the factors known to affect classical conditioning (e.g., Shanks, 1985). Cues 4) and 5) correspond to what has been termed "cue competition" in the animal learning literature or "discounting" and "augmentation" in the causal attribution literature. Basically, organisms take into account not only how a potential cause covaries with the effect, but how this cause competes with rival explanations as well. Cue competition effects have largely been observed in animals such as rats: if a light is presented along with a tone followed by a shock, after the tone alone has been presented followed by the shock, then little fear is observed when the light is later presented alone (Kamin, 1968). Although the light has been followed by shock, the presence of the tone (an alternative cause) leads the animal to disregard the light as the cause of shock ("discounting"). On the other hand, if a light and a tone are first presented together followed by shock, rats exhibited moderate fear to each of them. If the tone is later presented without shock, enhanced fear of the light is observed (Kaufman & Bolles, 1981). When the light and the tone were experienced together, both of them could be the cause of the shock, but when the animals later learned that the light did not produce the shock, they "re-evalued" their previous knowledge and attributed more causal power to the tone ("augmentation"). Surely, these causal inferences are based on observing the occurrence or the nonoccurrence of events; but, how could it be otherwise? Jennifer Vonk herself admits that "even unobservable forces invariably result in observable regularities" (par. 4). Furthermore, Humean observable cues lead to the idea of causality. Whether or not there is a correspondence of subjective causal attributions to actual causal relations in the real world is something that we humans can never know for sure. In fact, humankind has constantly modified its explanations about how the world works in light of scientific (observable) evidence. There is widespread belief that the principles of causal understanding can be observed in the behavior of animals such as rats. There are, of course, many differences between human and nonhuman animals; but, the lack of causal understanding does not seem the most obvious illustration.

Causal Understanding versus Making Predictions Based on Observable Cues

Jennifer Vonk
Jun 3, 2005 20:46 UT

I think that Castro and Wasserman have a different definition of “causal understanding” than I do. I do not contest any of the findings they report as evidence that animals are well equipped to reason about the co-occurrence of events in the world. However, where we differ is with regards to whether making predictions about the temporal and spatial relations between observable events is sufficient to constitute causal understanding. I do not believe that it is, for reasons expressed by myself and others throughout this conference. I tried to stress that animals are very adept at using observable regularities to make predictions about the behaviors of objects and events, as well as other beings. What remains to be determined is whether they additionally invoke concepts of underlying unobservable causal forces to explain these events. Nothing that Castro and Wasserman point to indicates that they do. They suggest that “Whether or not there is a correspondence of subjective causal attributions to actual causal relations in the real world is something that we humans can never know for sure”, but this is exactly what I think we should be investigating. Certainly it is a challenge but it is not an impossible one. I tried to suggest some ways in which we could begin such explorations. I am sure that others will have better ideas.

I would like to make an additional point. Castro and Wasserman suggest that “If animals’ behavior is also affected by the same cues, then the most parsimonious interpretation is that animals are also capable of causal understanding”. This statement suffers from the argument by analogy which Povinelli and colleagues have discussed at length (Povinelli & Giambrone, 1999; Povinelli, Bering & Giambrone, 2001). One simply can not infer that similar behaviors arise as the result of identical cognitive mechanisms. It is quite possible that animals arrive at the same resulting actions for quite different reasons than we do. It is also quite possible that humans respond in ways indicative of causal understanding when in fact no such understanding is being employed. The argument of parsimony is often applied to explain animal data at the cost of acknowledging different etiologies for similar behaviors. Let’s consider convergent evolution as one example where the argument by analogy might be misapplied.

Lastly, Castro and Wasserman point out that humans modify their understanding of the world in light of new, scientific insights. If there is any evidence that animals also test hypotheses and modify existing behaviors in the light of new discoveries (other than relying on extinction and reversal of learning sets), it might indicate that they too are driven to seek out causal explanations.

Povinelli, D.J. & Giambrone, S. (1999). Inferring other minds: Failure of the argument by analogy. Philosophical Topics, 27, 167-201.

Povinelli, D.J., Bering, J., & Giambrone, S. (2001). Toward a science of other minds: Escaping the argument by analogy. Cognitive Science, 24, 509-541.

Can we "prove" the lack of causality in non-humans?

Giyoo Hatano & Kayoko Inagaki
Jun 5, 2005 8:06 UT

Vonk's paper offers a number of insightful methodological suggestions. She is extremely cautious about attributing causality in non-humans, so cautious that we have an impression that, in contrast with Susan Hespos, who is willing to recognize causality in preverbal infants, she implicitly assumes causal reasoning to be a uniquely human capability. As pointed out by a number of participants in this conference, differentiating predictions based on a proper causal device from those that are not (or based on an arbitrary cause) is a methodologically hard problem, but one possible way is varying the preceding event (PE) in a variety of ways and observing how the target respondent's predictions of the subsequent event (SE) change -- if the prediction for SE changes only when PE is varied on relevant dimensions, we may conclude that the respondent relies on a (natural) cause. Thus, if chimpanzees attempt to retrieve a piece

of food by dragging it across a table when its color, size, material quality, etc. are varied, but they do not do so when there is an empty space, it is strong evidence for their possession of a physical causal principle ("things fall unless supported," an intuitive version of the notion of gravity). However, if such differentiated patterns are not observed, can we claim that they do not understand the causal principle? We do not think so. As a number of conceptual development researchers (e.g., Keil, 1992) indicate, young children's, or even lay adults', naive theories (that surely include causal devices) are not for making novel predictions. Naive theories do not cover all phenomena they could, and their implications are not fully extended nor exploited.

It is not surprising that non-verbal animals are not motivated to explain events, but do they not interpret the observed events? In other words, non-human animals just learn associations between a large number of possible preceding events and the subsequent event? Considering that even rats tend to attribute their bodily disturbance to novel food eaten (Garcia, 1981), apes may select a particular PE as the cause for SE, at least in some contexts. As Jennifer indicates, what is needed is conducting many more ingenious experiments that are likely to induce apes' maximal capability for causal reasoning.

References Garcia, J. (1981) Tilting at the paper mills of academe. American Psychologist, 36, 149-158. Keil, F. C. (1992). The origins of an autonomous biology. In M. R. Gunnar & M. Maratsos (Eds.), Modularity and constraints in language and cognition. The Minnesota Symposia on Child Psychology (Vol. 25, pp.103-137). Hillsdale, NJ: Erlbaum.

Reply to Hatano & Inagaki

Jennifer Vonk

Jun 6, 2005 9:28 UT

Contrary to what Hatano and Inagaki suggest, I remain open-minded to the possibility that non-humans may well demonstrate causal understanding in some circumstances. I do think that these circumstances may be limited to those in which the causal force is directly observable. I am sure that many animals reason about the possible consequences of various actions for example chimpanzees surely know that, when they spit or throw feces at a person, that person will react in quite a stereotypical fashion. I am just not quite sure the extent to which such cases are truly representative of abstract causal forces. The answer depends upon how we define causal forces and I think, at this point, we have not settled on a clear definition. In addition, I think we have reason to believe, based on the evidence so far, that animals generally fail to reason about many causal forces in the same manner that humans do. For instance, chimpanzees persist in dragging food rewards (of various sizes, shapes and colours) across a hole in a table surface (Povinelli, 2000).

Let me add one caveat though I believe that humans often fail to demonstrate true causal understanding. For instance, many of us falsely believe that heavier objects fall faster than light objects. I think many of the arguments that I have applied to suggest that animals are not representing the true causal forces underlying a sequence of observable events also apply to humans in many cases. However, I think, (sometimes only with explicit teaching), we are able to gain awareness and understanding of many underlying unobservable causes mediating such events. There may be no good evidence as of yet that animals do the same, but I am more optimistic than Hatano and Inagaki regarding the possibility that we can test these claims. However, I do not think the fact that animals change their predictions about an SE based on changes in PEs (in the very general sense) proves anything. I would never argue that animals are incapable of making predictions based on what kinds of events follow other kinds of events. In order to use such a process to demonstrate causal understanding, there would have to be very specific cases whereby the changes in the PEs could be predictive only given a representation of the underlying causal force, in the absence of the opportunity to learn what events are likely to follow other events. I am not sure why Hatano & Inagaki feel that naïve folk theories are not designed to lead to accurate predictions in novel situations if this is true, what WOULD be the adaptive function of the ability to form theories?

Left with the association, naked as if it were? Ideas from honeybee learning

Martin Giurfa (Researcher, CNRS Centre de Recherche sur la Cognition Animale)

(Date of publication: 6 June 2005)

Abstract: The brain of a honeybee contains only 960 000 neurons and its volume represents only 1 mm3. However, it supports impressive behavioural capabilities. Honeybees are equipped with sophisticated sensory systems and have well developed learning and memory capacities, whose essential mechanisms do not differ drastically from those of vertebrates. But, which regularities can honeybees extract from their environment besides those underlying simple forms of associative learning? Here, I focus on non-elemental forms of learning by honeybees. I show that bees exhibit learning abilities that have been traditionally ascribed to a restricted portion of vertebrates, as they go beyond simple stimulus–stimulus or response–stimulus associations.

Anne Reboul's contribution to this conference has open the debate on causality in terms of the distinction between associative and explanatory behavior. It is argued that most animals (the example of Anne's dog is illustrative) are left with the association, "naked as if it were" while Humans capable of causal learning build explanations that bring them back from the effects to the causes, as a way to grasp the causal relationship between events. Retrospection would be therefore a critical aspect in causal knowledge (see also John Watson's comment on Anne Reboul's article). Moreover, the distinction between extracting the relationship between observable and unobservable entities is also invoked as an essential point to distinguish between associative and causal learning. It is argued that associative learning allows extracting predictive relationships between perceptible entities in the world while causal learning allows going beyond the observable and learning about stimuli that are not present.

I will concentrate here on experiments on non-elemental forms of learning in honeybees. My objective will be to provide experimental evidence allowing to discuss a statement that was at the origin of this conference, namely that causality is at the base of the acquisition and use of categories and concepts (see conference primer). I will focus on chosen examples of categorization and rule learning in honeybees. I maintain that different forms of cognitive processing underlie these two forms of learning despite experimental commonalities. While an elemental associative account can explain categorization performances, rule learning requires a different explanatory basis. I argue that prospective / predictive analysis applies to categorization but that retrospection may be involved in rule learning. As honeybees can solve both kinds of problems, demarcating between species on the basis of these capacities seems inappropriate.

Causality as the cognitive basis for the acquisition of categories?

Categorization refers to the classification of perceptual input into defined functional groups (Harnard 1987). It can be defined as the ability to group distinguishable objects or events on the basis of a common attribute or set of attributes, and therefore to respond similarly to them (Troje et al 1999; Delius et al 2000; Zentall et al 2002). Categorization deals, therefore, with the extraction of these defining attributes from objects of the animal's environment. Our use of the term categorization will be restricted to those cases in which animals transfer their choice to novel stimuli that they have never met before on the basis of common features shared with known stimuli.

The question of whether an insect brain can categorize visual objects in its environment has been recently answered affirmatively (Giurfa et al. 1996). Although this capacity can appear as surprising in the case of insects, considering that categorization obeys simple associative rules should allow demystifying this performance. We will restrict our examples to the visual modality as this is the modality in which categorization studies have been performed in the honeybee.

The honeybee constitutes a good model for addressing the question of visual categorization due to its remarkable learning and memory capabilities for visual stimuli (Menzel and Giurfa 2001; Giurfa 2003). Bees can be easily trained to fly towards a visual target on which a reward of sucrose solution is delivered by the experimenter. The associations built in this context link visual stimuli and reward, but also the

response of the animal (e.g. landing) and reward, i.e. bees learn that a given visual cue (e.g. a color) will be associated with a reward of sucrose solution and that they have to land on it to get the reward. Using this basic design in which procedural modifications can be introduced, several studies have shown recently the ability of visual categorization in honeybees trained to discriminate different patterns and shapes. As mentioned above, such a demonstration requires that bees are able to transfer appropriate responding to novel stimuli belonging to the trained category.

Such a transfer has been demonstrated for a variety of visual features. For instance, van Hateren et al (1990) trained bees to discriminate two given gratings presented vertically and differently oriented (e.g. 45° vs. 135°) by rewarding one of these gratings with sucrose solution and the other not. Each bee was trained with a changing succession of pairs of different gratings, one of which was always rewarded and the other not (Fig. 1). Despite the difference in pattern quality, all the rewarded patterns had the same edge orientation and all the non rewarded patterns had also a common orientation, perpendicular to the rewarded one. Under these circumstances, the bees had to extract and learn the orientation that was common to all rewarded patterns to solve the task. This was the only cue predicting reward delivery. In the tests, bees were presented with novel patterns, which they were never exposed to before, which were all non-rewarded, but which exhibited the same stripe orientations as the rewarding and non-rewarding patterns employed during the training. In such transfer tests, bees chose the appropriate orientation despite the novelty of the structural details of the stimuli. Thus, bees could categorize visual stimuli on the basis of their global orientation. This conclusion led to a model of orientation detection in the honeybee, based on the existence of three types of orientation detectors, with a defined preferred orientations and tuning (Srinivasan et al 1994), comparable to those available in the mammalian visual cortex (Hubel and Wiesel 1962). Such detectors were found later by means of electrophysiological recordings in the visual areas of the bee brain (Yang and Maddess 1997).

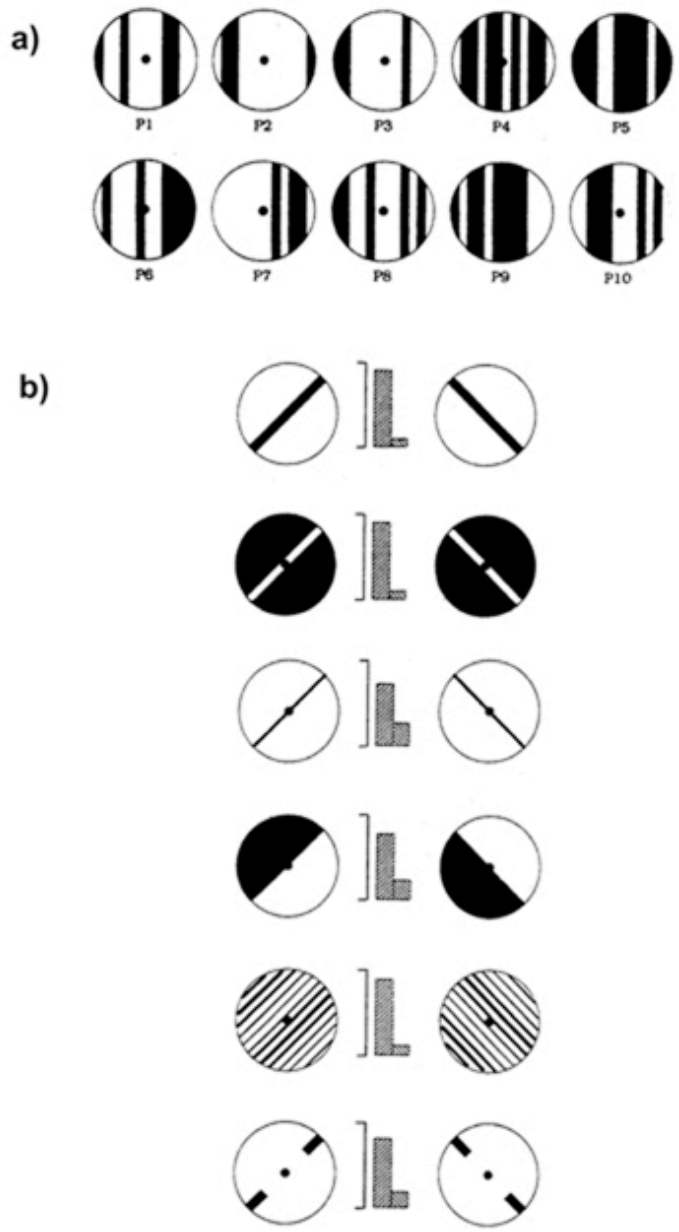


Figure 1: Categorization of edge orientation by honeybees. **(a)** Training stimuli (P1 to P10) used in van Hateren et al's experiments (1990). Pairs of stimuli were presented in a random succession to the bees. Within each pair, one was oriented at 45° and the other at 135°. In this case, gratings oriented at 45° were rewarded with sucrose solution while those at 135° were non rewarded. **b)** Tests performed with stimulus pairs not used during the training. In each case, there was a significant preference for the pattern presenting the orientation rewarded during the training. Bars indicate the proportion of choices for each stimulus. Bees transferred their choice from the known to the novel patterns and classified them according to their orientation (from van Hateren et al 1990).

Thus, honeybees show positive transfer of appropriate responding from a trained to a novel set of stimuli, and their performances are consistent with the definition of categorization. Visual stimulus categorization is not, therefore, a prerogative of certain vertebrates. However, is it so surprising and do causality judgments (sensu Reboul) underlie this ability, as stated at the origin of this conference? I don't think so. In my opinion, categorization does not reflect retrospective analysis of events but results from simple

associative learning. To explain this view, the neural mechanisms underlying categorization could be considered, in particular with respect to the organization of the bee brain.

If we admit that visual stimuli are categorized on the basis of specific features such as orientation or symmetry, the neural implementation of category recognition could be relatively simple. The feature(s) allowing stimulus classification would activate specific neuronal detectors in the optic lobes, the visual areas of the bee brain. Examples of such feature detectors are the orientation detectors whose tuning and orientation have been already characterized by means of electrophysiological recordings in the honeybee optic lobes (Yang and Maddess 1997; see above). Thus responding to different gratings having a common orientation of, say, 45°, is simple as all these gratings will elicit the same neural activation in the same set of orientation detectors despite their different structural quality. In the case of category acquisition, the activation of an additional neural element is needed. Such element would be necessary and sufficient to represent the reward (sucrose solution) and should contact and modulate the activity of the visual feature detectors in order to assign value to appropriate firing. This kind of neuron has been found in the honeybee brain as related to the olfactory circuit. VUMmx1 is a neuron present in the honeybee brain that receives its name from its localization (the name is the abbreviation of "ventral unpaired median neuron of the maxillary neuromere 1"). The dendrites of VUMmx1 arborize symmetrically in the brain and converge with the olfactory pathway at different sites (Hammer 1993). The essential property of VUMmx1 is that it responds to sucrose solution delivered both at the antennae and the proboscis of the bee with long lasting spike activity (Hammer 1993). Furthermore, the activity of this neuron constitutes the neuronal representation of reward in the case of olfactory learning as shown by the fact that bees can learn an olfactory stimulus which was paired with an artificial depolarization of VUMmx1 instead of sucrose reward (Hammer 1993). Other VUM neurons whose function is still unknown are present in the bee brain. It could be conceived that one of them (or more than one) contact the visual circuit to function as reinforcement in associative visual learning. Category learning could be thus reduced to the progressive reinforcement (through Hebbian rules, for instance) of an associative neural circuit relating visual-coding and reinforcement-coding neurons, similar to that underlying simple associative (e.g. Pavlovian) conditioning.

Thus, caution is needed before relating categorization to causality, i.e. to explanatory, retrospective behavior. One of the original statements of this conference, namely that "causality is the cognitive basis for the acquisition and the use of categories and concepts" may not hold, as indicated by experiments on honeybee visual categorization. Categorization, even if viewed as a higher-order cognitive performance, may simply rely on elemental links between conditioned and unconditioned stimuli. It may thus be based on prospective/predictive analysis and not on retrospection.

Causality as the cognitive basis for rule learning?

Like categorization, rule learning also presupposes positive transfer of an appropriate response from a known set to a novel set of stimuli. Despite this common experimental basis, I maintain that these processes do not rely on common mechanisms. In rule learning, the animal bases its choice, not on the perceptual similarity between the novel and the known stimuli which may not share, contrarily to categorization problems, any common feature, but on links that transcend the stimuli used to train it. Examples of such rules are "larger than", or "on top of", which may apply to stimuli which do not share any common feature but which can nevertheless be classified following the rule. I maintain that simple, elemental associative links cannot account for success in rule learning and that retrospective/diagnosis may be necessary to solve this kind of problem.

An example of rule learning is the learning of the so-called principles of sameness and of difference. These rules are usually uncovered through the delayed matching to sample (DMS) and the delayed non-matching to sample (DNMS) experiments, respectively. In DMS, animals are presented with a sample and then with a set of stimuli, one of which is identical to the sample and which is reinforced. As the sample is being changed regularly, they have to learn the sameness rule 'choose always what is shown to you (the sample), independently of what is shown to you'. In DNMS, the animal has to learn the opposite, i.e. 'choose always the opposite to what is shown to you (the sample)'. The interesting point concerning these protocols is that predictive analysis based on stimulus or feature generalization does not necessarily hold as the rule is ideally independent of the physical nature of the stimuli used. To discover the rule, the

animal has to operate on the set of examples known such that retrospection and different forms of heuristics can be applied to solve the problem. Neural accounts based on simple associative networks such as that proposed for visual categorization (see above) may not be valid in this case. Although reinforcement can still be represented by a specific neural pathway or element (such as the VUMx1 neuron or its equivalents; see above), the novel, differing sample (e.g.; a color) will not activate the same network components responding to a previous sample (e.g., an odor). Extracting the rule in a changing learning set means therefore going beyond stimulus modality and performing a form of retrospective or diagnostic analysis of the problem faced.

Is this kind of problem a good candidate for species demarcation? The answer is no. Honeybees foraging in a Y-maze learn to solve both DMS and DNMS rules (Giurfa et al. 2001). Bees were trained in a DMS problem in which they were presented with a changing non-rewarded sample (i.e. one of two different color disks or one of two different black-and-white gratings, vertical or horizontal) at the entrance of a maze. The bees were rewarded only if they chose the stimulus identical to the sample once within the maze. Bees trained with colors and presented in transfer tests with gratings that they have not experienced before solved the problem and chose the grating identical to the sample at the entrance of the maze. Similarly, bees trained with the gratings and tested with colors in transfer tests also solved the problem and chose the novel color corresponding to that of the sample grating at the maze entrance. Transfer was not limited to different kinds of modalities (pattern vs. color) within the visual domain but could also operate between drastically different domains such as olfaction and vision (Giurfa et al. 2001). Furthermore, bees also mastered a DNMS task, thus showing that they also learned a principle of difference between stimuli (Giurfa et al. 2001). In both DMS and DNMS, win-stay/ loose-shift (or win-shift / loose-stay) strategies could not account for the performances of the bees. These results document that bees learn rules relating stimuli in their environment. The capacity of honeybees to solve DMS tasks has been verified in other contexts (see, for instance Zhang et al. 2004, 2005). In particular, introducing longer delays between the offset of the sample and the onset of the comparison stimuli yielded a decay in matching performances and allowed to suggest that honeybees retrospectively code the samples in delayed matching-to-sample task (Zhang et al. 2005).

Retrospective revaluation in honeybees

Associative learning theories account for causal and predictive learning but face the problem of retrospective learning as for such theories, learning can only occur when a stimulus is present. In this sense, the case of backward blocking (Shanks 1985) is interesting as it implies training an animal with a compound stimulus AB reinforced (AB+) in a first phase, and then with A reinforced (A+) in a second phase; if backward blocking occurs, ratings of B in a third test phase are reduced retrospectively by experience with A in the second phase because A alone is enough to predict the outcome of AB. This experiment is therefore interesting because it can be related to the claim made in this conference on the distinction between associative and causal learning. As mentioned in the first paragraph, in Anne Reboul's contribution it is suggested that associative learning allows extracting predictive relationships between perceptible entities in the world while causal learning allows going beyond the observable and learning about stimuli that are not present. If this is the case, then backward blocking could be an interesting case as animals learn retrospectively about stimulus B.

Backward blocking has been recently studied in honeybees (Blaser et al. 2004). It was shown that responding to B after AB+training was less in animals that also had A+ training than in control animals that were equally often reinforced in the absence of A. Furthermore, responding to B was less after AB+followed by differential training A+C- than after AB+followed by C+A- training. In the first case (AB+, A+C-), retrospective revaluation would decrease the value of B as A is a reliable cause or predictor of the outcome during the compound training. In the second case (AB+, A-C+), retrospective revaluation would have the opposite effect, i.e. it would enhance the value of B as reliable cause or predictor of the outcome during the compound training.

Caution is nevertheless needed when analyzing these data in the light of a possible dichotomy between associative and causal learning. It is worth mentioning that associative accounts have been provided to explain retrospective revaluation. For instance, Van Hamme and Wasserman (1994) suggested that when A and B are paired with the outcome in Phase 1, a within-compound association is formed between them,

which then allows the presentation of A in Phase 2 to activate the representation of B. The predictive strength of an expected but absent cue decreases. Therefore, when A alone is followed by the outcome in Phase 2, the associative strength for A increases while the associative strength for the absent cue B simultaneously decreases. In this case, therefore, the proposed distinction between associative learning, which allows extracting predictive relationships between perceptible stimuli in the world, and causal learning, which allows learning about stimuli that are not present, may not be so straightforward.

Conclusions

Although honeybees, as Anne Reboul's dog, are sometimes left with the associations 'naked as if they were", they can also operate on associations between events in their environment in order to extract rules and retrospectively evaluate stimuli and their outcomes. It seems therefore that bees have expectations based on associative, predictive learning but that such learning is not the whole of honeybee cognition. Clearly, research articulated on categorization and rule learning may be useful to distinguish between different levels of complexity of cognitive processing but not to determine what is unique to Humans. This implies, therefore, that either diagnostic / retrospective learning is not unique to Humans, or that such uniqueness resides elsewhere, for instance in the existence of language.

References

Blaser RE, Couvillon PA Bitterman ME (2004) Backward blocking in honeybees. Q J Exp Psychol B 57: 349-60.

Giurfa M, Eichmann B, Menzel R (1996) Symmetry perception in an insect. Nature 382: 458-461

Giurfa M, Zhang S, Jenett A, Menzel R, Srinivasan MV(2001) The concepts of 'sameness' and 'difference' in an insect. Nature 410: 930-933

Giurfa M (2003) Cognitive neuroethology: dissecting non-elemental learning in a honeybee brain. Curr Opin Neurobiol 13: 726-735

Hammer M (1993) An identified neuron mediates the unconditioned stimulus in associative olfactory learning in honeybees.Nature 366: 59-63

Harnard S (1987) Categorical Perception. The Groundwork of Cognition. CambridgeUniversityPress, Cambridge

Hateren JH v, Srinivasan MV, Wait PB (1990) Pattern recognition in bees: orientation discrimination. J Comp Physiol A 197: 649-654

Hubel DH, WieselTN(1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol (London)160: 106-154

Menzel R, Giurfa M (2001) Cognitive architecture of a minibrain: the honeybee. Trends Cognit Sci 5: 62-71

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. Quart J Exp Psychol 37B: 1-21.

Srinivasan MV, Zhang SW, Witney K (1994) Visual discrimination of pattern orientation by honeybees: Performance and implications for "cortical" processing. Phil Trans Royal Soc Lond (B) 343: 199-210

Troje F, Huber L, Loidolt M, Aust U, Fieder M (1999) Categorical learning in pigeons: the role of texture and shape in complex static stimuli. VisRes 39: 353-366

Van Hamme LJ, Wasserman EA (1994) Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. Learn Motivation 25: 127-151

Yang EC, Maddess T (1997) Orientation-sensitive neurons in the brain of the honey bee (Apis mellifera). J Insect Physiol 43: 329-336

Zentall TR, Galizio M, Critchfield TS (2002) Categorization, concept learning and behavior analysis: an introduction. J Exp Anal Behav 78: 237-248

Zhang SW, Srinivasan MV, Zhu H, and Wong J (2004). Grouping of visual objects by honeybees. J Exp Biol 207: 3289-3298.

Zhang S, Bock F, Si A, Tautz J, Srinivasan MV (2005) Visual working memory in decision making by honey bees. PNAS 102:5250-5255.

Discussion

Association, causality and restropective-diagnostic reasoning
Anne Reboul
Jun 9, 2005 9:51 UT

I would like here to comment not only on Martin's excellent paper, but also on Leyre and Ed's comment over Jennifer's paper because all of them seem to home in very much the same bunch of ideas. Let me begin first with Leyre and Ed's view of the Humean approach to causality. As they quite rightly point out, Hume outlined a number of (perceptible) cues to causality. However, his account did not stop there: he also added that in the human conception of causality, there is a notion (which he at least sometimes took to be a mere artefact of human psychology) to the effect that there is something NON perceptible in the human view of causality, which he described rather vaguely as a necessary connection between the two events considered. Being an empiricist, he tended to see the postulation of such a connection as without any base because the connection itself is not perceptible. Leaving aside the metaphysical problem of whether such "connections" do or do not exist, I think it is fair to Hume to say that his cues to causality closely correspond to what has been discovered about association in animals and that, in that way at least, he was a great precursor (a point made by Leyre and Ed). However, this doesn't mean, that, for him, human causality was association (hence the notion of "connection"). The notion of unobservability which has been developped by Jennifer and Povinelli is not equivalent to the Humean "connection", but is more precise and can well be articulated differently in different causal domains (e.g., force in naive physics, belief in naive psychology, essence in naive biology). This is not to deny that association plays an important role, even in human causal cognition, but in humans, it is a basis for causal reasoning, not the end of the process and it is here, I think that Jennifer's ideas about unobservables step in. I should add that association can go a very long way, which is why animals manage so well, supposing them to be only capable of association. To come now to Martin's paper, there is a misunderstanding about what is meant by categories and concepts in the conference primer. There is, I think, no role for causality in the learning of categories of the kinds shown in Martin's paper. They have a purely perceptual basis and there doesn't seem to be any concept associated with them. The kind of categories alluded to in the primer were natural or artifactual kinds for whole objects, where knowledge seems to be strongly dependent on knowledge about other categories falling under the same (superordinate) concepts. There also seems to be a misunderstanding about what is meant by "retrospective". In my original paper, the concept was quite simply based on a comparison between two ways of reasoning about causality: supposing that there is a correlation between event A and (a slightly later) event B, there is an asymetry between them due to priority (A is before B) and the distinction between predictive and retrospective/diagnostic reasoning follows the asymetrical link between A and B, from A to B in predictive reasoning and from B to A in retrospective/diagnostic reasoning. This has nothing to do with revising pre-established causal links: it has to do with the ability to go in both directions, or in only one direction, i.e., from cause to effect. The claim was that the very notion of explanation entails retrospective/diagnostic abilities, and one hypothesis is that it is only humans who are able of that type of reasoning. But this is only an hypothesis, susceptible to empirical contradiction.

Inferring Causality and Making Predictions. Some Misconceptions in the Animal and Human Learning Literature

Helena Matute (Professor of Psychology, Universidad de Deusto, Spain) and
Miguel A. Vadillo (Graduate Student, Universidad de Deusto, Bilbao, Spain)
(Date of publication: 20 June 2005)

Abstract: A common assumption in studies on associative learning is that inferring causal relations, assessing predictive relations, and making predictions are closely related processes. In spite of the interdependencies between them, evidence (and simple intuition) indicates that people perform these processes differently and that they cannot be attributable to a common, single underlying mechanism. As we will show, this evidence has both methodological and theoretical implications.

Researchers may disagree in their theories of human and animal learning. But all of them agree at least in one thing: Learning is an adaptive tool that improves and organism's ability to survive in a changing environment. The process of learning provides an animal with (either explicit or implicit) knowledge that captures, among many other things, the statistical relationships between different significant events that occur in the environment. This process takes place, for example, in Pavlovian conditioning experiments, in which humans and other animals learn the statistical relationship between a conditioned stimulus (CS) and an unconditioned stimulus (US) and adapt their response to the CS as a result of this experience. A similar process can be observed in experiments in which human subjects have to learn the relationship between different cues and outcomes and are asked to rate the perceived strength of those relationships.

However, as we have just stated, there is little consensus regarding the mechanisms by which organisms acquire and use this information. And, what is more important in our opinion, there are some conceptual contradictions in the terms used by theorists that will necessarily make the consensus more and more unlikely. As we will argue below, many researchers use common concepts related to associative learning inconsistently. Specifically, some authors implicitly assume that responding to the predictive value of a cue is the same as predicting an outcome, and a similar lack of clear-cut definitions is also evident when it is assumed that animals learn predictive relations between events or that they learn causal relations. Most theories of human and animal learning use terms such as prediction, predictive value and causality as if they were synonymous.

Normative analyses and experiments with humans

Regardless of whether people and animals ordinarily make this distinction in their daily lives, making a prediction is clearly different from assessing the predictive value of a cue. Imagine for example that a study shows that people living in your country have a probability of 10% of suffering skin cancer. Should you protect your skin? Of course you should, because there is a moderate likelihood that you will suffer skin cancer otherwise. Moreover, you should do so regardless of the probability of suffering skin cancer in other countries. It does not matter whether people living abroad suffer cancer with a probability of 0%, 10%, or 20%. You should always be equally careful because you are likely to be affected by this disease. Your prediction of the likelihood of suffering skin cancer (and your preparatory behavior to avoid it) should be unaffected by what happens in circumstances different from yours.

This knowledge of what happens in other countries is, however, extremely important if you have to assess whether or not living in your country is a good predictor of the likelihood of suffering skin cancer. If the probability of suffering cancer is 0.10 both in your country and in other countries, you cannot say that living in your country is a good predictor of developing skin cancer. Knowing that one person lives in your country does not help you decide whether he or she will be more likely to suffer cancer. Therefore, one thing is to make a prediction (how likely it is to suffer cancer) and a different one to assess the predictive value of a cue (whether or not living in your country increases the odds of being affected by skin cancer). Whereas predicting an outcome in a given situation requires taking into account only the probability of that outcome given the cues that define such situation [i.e., p(outcome|cue)], assessing the predictive value of those cues requires taking into account whether the probability of the outcome is greater or smaller in the presence than in the absence of those cues. In other words, in order to assess the predictive value of a

cue, you need to know whether the probability of the outcome is different in the presence than in the absence of that cue. This predictiveness or cue-outcome contingency is usually measured by the statistical index Äp, which is equal to the probability of the outcome given the cue minus the probability of the outcome in the absence of the cue [i.e., Äp = p(outcome|cue) – p(outcome|no cue)].

Data gathered in our laboratory shows that the distinction between making predictions and assessing predictiveness and causal relations is a distinction that people draw spontaneously, and not a merely theoretical distinction that people are potentially able to understand but that they don't use in their daily lives (Matute, Vegas, & De Marez, 2002; Vadillo, Miller, & Matute, 2005). For example, if people are told that 50% of the patients taking a medicine develop and allergic reaction and they are asked to assess how likely it is that a new, unknown patient taking the medicine will suffer the allergy, their response is close to the objective 50% regardless of the probability of developing the allergy in patients not taking the medicine. However, if they are asked to say whether they think that taking the medicine is a good predictor of developing the allergy, then they take into account what happens with patients not taking the medicine.

Similarly, there is a difference between assessing whether a cue is a good predictor of an outcome and assessing whether a cue is a cause of the outcome. This difference can be easily understood with the aid of a real-world example related to the Simpson's paradox. During the 70's it was discovered that women applying to study in Berkeley University were more likely to be rejected than men. Although this evidence was taken as evidence of discrimination against female applicants, a closer look at the data showed that women were more likely to be rejected because they tended to apply for more selective programs with higher rejection rates. In this example, knowing that a woman has applied for Berkeley is a datum that provides us with important predictive information: If we know that a given woman has submitted an application to Berkeley, then we know that the probability that she will be rejected is higher than it would have been if the applicant had been a man. However, there is no causal relation between being a woman and being rejected. The reason why sex is a good predictor of rejection is that sex covaries with a factor that causes rejection. Thus, in order to infer a causal relation it is not enough to pay attention to the predictive value of the cue, it is also necessary to check that there are no confounding variables that could artificially increase the covariation between the cue and the outcome (see Cheng, 1997; Cheng & Novick, 1992).

In spite of these normative and descriptive differences between predicting, assessing predictive value and estimating the strength of causal relations, researchers often use these concepts, either explicitly or implicitly, as synonymous. For example, researchers studying causal learning have sometimes used in their experiments instructions or test questions suggesting scenarios where the predictive value of the cues is more important than their causal status. Similarly, theories proposed to account for causal learning have been assumed to be adequate to account for participants predictions and for judgments of cue-outcome predictiveness.

Predictions and predictive value in animal conditioning experiments

So far, we have shown that making a prediction, assessing the predictive value of a cue and assessing the strength of a causal relation are different things, in spite of what learning theories posit. And that there is also evidence showing that people make these distinctions spontaneously. But the next question we could ask is: Is this distinction relevant for animal learning researchers? What do animals do when they give a conditioned response? Are they predicting the unconditioned stimulus? Or are they responding to the value of the conditioned stimulus as a predictor of the unconditioned stimulus? These two ideas are often confounded in the animal literature.

Many published reports on Pavlovian conditioning start their introductions by remarking how associative learning allows animals to prepare for future events adaptively. From this point of view, one would expect animals in a classical conditioning experiment to be preparing themselves for the occurrence of the US after they experience the CS. Therefore, it is predicting the US what should be important for them. However, there is some evidence inconsistent with this perspective.

Imagine that a rat receives a footshock after a light in 50% of the trials in which the light is presented. In

principle the rat should be moderately afraid after the presentation of the light, because this light is followed by the aversive US in half of the occasions. Moreover, if the rat were simply predicting the US, then, as we have discussed above, the probability of the US in the absence of the light should be completely irrelevant. If the footshock follows the light with a probability of 0.50, then the rat should be afraid, no matter whether the footshock never occurs in the absence of the light or whether it occurs constantly in the absence of the light. In other words, when predicting whether a US will follow a CS, the predictive value of the CS should not be important or, at least, not as important as the probability of the US after the CS.

Famous experiments performed by Rescorla during the 60's seem to be inconsistent with this perspective. Rescorla (1968) found that for a given probability of the US given the CS, the conditioned response (CR) was negatively correlated with the probability of the US in the absence of the CS. That is, a rat shows stronger fear to a light that is followed by a US in 50% of the trials if the US never occurs in the absence of the light than if it occurs in the absence of the light with a probability of 0.50. This seems to indicate that an animal's CR does not reflect a prediction of the US. The animal, on the contrary, seems to be assessing the CS's predictiveness or predictive value.

The Rescorla-Wagner model of animal conditioning: Predictive value or prediction?

A few years after those experiments were published, Rescorla and Wagner (1972) proposed their famous model of animal conditioning which explained the sensitivity to the CS-US contingency. According to this model, the strengthening of the CS-US association in a given trial depends on the extent to which the US is unpredicted after the presentation of the CS. In each trial, the change in the strength of the association, ΔV, is given by the equation

$$\Delta V = \alpha \cdot \beta \cdot (\lambda - V_T)$$

where α and β are learning rate parameters dependent on the salience of the CS and the US respectively, λ is the maximum associative strength that the US can support, and V_T is the associative strength of all the stimuli (CS, experimental context, etc.) that are present in that trial. The term $(\lambda - V_T)$ measures the degree of surprise produced by the presentation of the US: The difference between the US that is actually presented (λ) and the US expected on the basis of the stimuli presented in that trial (V_T).

Apparently, this model was able to explain Rescorla's (1968) results in a quite simple manner. According to the model, when the US occurs frequently in a given context, both in the presence and in the absence of the CS, the experimental context becomes strongly associated with the US given that there are many context-US pairings. Because of this strong context-US association, the animal is able to predict the US in that context and therefore the US is no longer surprising. In other words, the term $(\lambda - V_T)$ gradually approaches zero. Thus, eventually, the aforementioned equation will yield a value close to zero when the CS is presented, which means that the animal will not develop a strong CS-US association. This model predicts, therefore, that a low CS-US contingency should result in only weak conditioning to the CS.

However, the model also predicts that there should be a strong CR to the context in these situations. A strong context-US association would allow an animal to predict the US in that context. According to the Rescorla-Wagner model, if a light is followed by a footshock in 50% of the trials, and this footshock is also present with a probability of 0.50 in the absence of the light, the conditioned animal should be afraid at all times while being in that context because of fear being conditioned to the context. In other words, if an animal exposed to a null CS-US contingency is tested in the same context were it was conditioned, its total level of CR should not reflect CS-US predictiveness, but a prediction of the US in that situation. As we have shown, this is contrary to the pattern of results reported by Rescorla (1968), which showed an absence of fear in that situation.

In order to make sure that these are the actual predictions of the model, we have conducted some simulations of it. Below we show the results of some of them. These simulations illustrate how this model would expect animals to make accurate predictions of the USs if they are tested in the context where they were originally trained.
Figure 1 shows the associative strengths that a CS would accrue under different CS-US contingencies. As

there can be seen, the associative strength of the CS is asymptotically equal to the programmed CS-US contingency [1] and, therefore, this is not the information an animal should use when its goal is to predict the US.

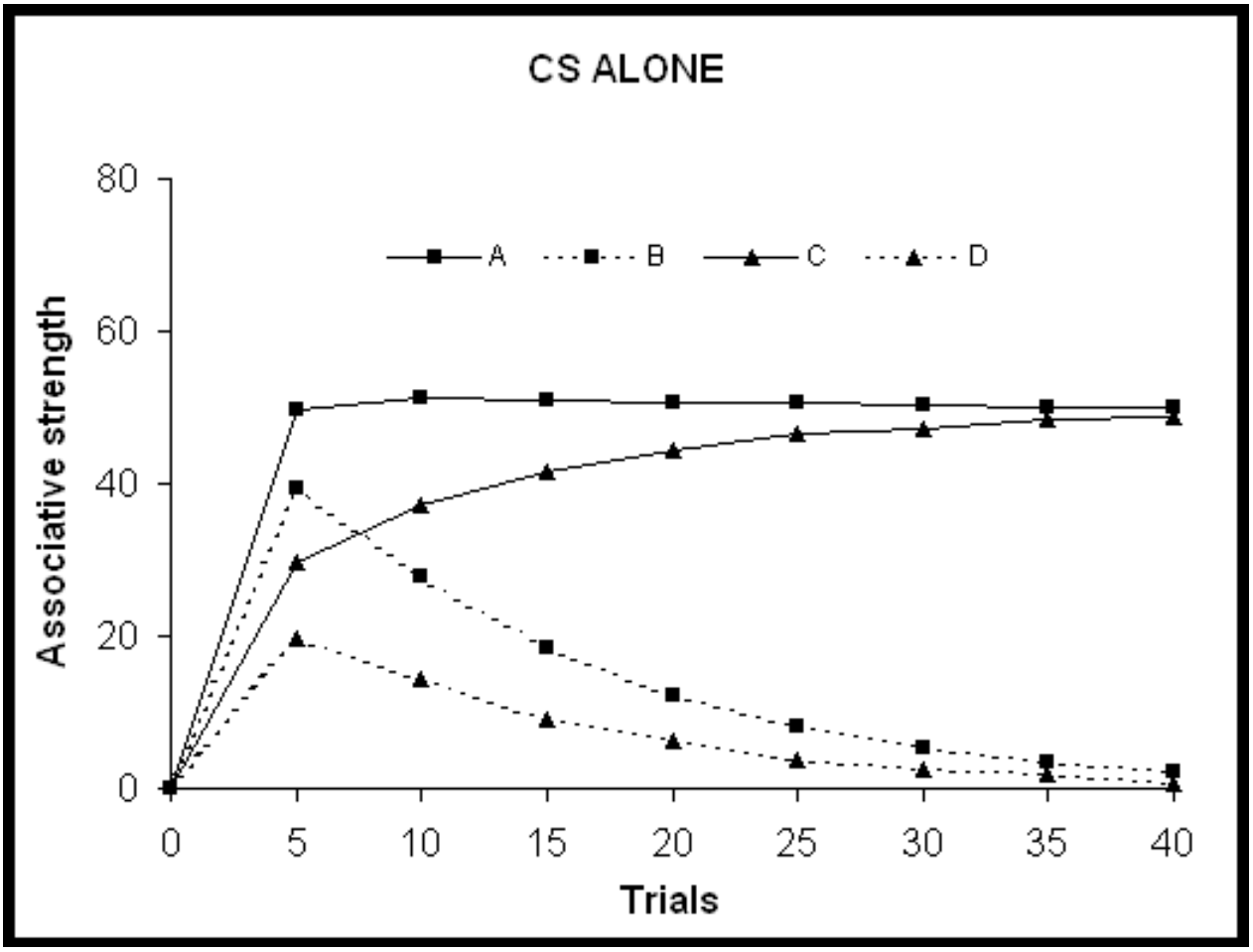


Figure 1. Simulation of the Rescorla-Wagner model showing the associative strength of the CS under several CS-US contingencies. The A series show the simulation in a conditioning situation where p(US|CS) = 1.00, p(US|~CS) = 0.50, Δp = 0.50. In the B condition p(US|CS) = 1.00, p(US|~CS) = 1.00, Δp = 0.00. In C, p(US|CS) = 0.50, p(US|~CS) = 0.00, Δp = 0.50. In D, p(US|CS) = 0.50, p(US|~CS) = 0.50, Δp = 0.00. Learning rate parameters were assigned the following values: αCue = 0.8, αContext = 0.5, βOutcome = 0.6, and βNoOutcome = 0.6. For each condition, 10,000 iterations with randomized trial orders were performed.

Figure 2 shows the sum of the associative strength of the CS and the associative strength of the context. As there can be seen, this sum of associative strengths always tends to be equal to the probability of the US given the CS [p(US|CS)][2], regardless of cue-outcome contingency.

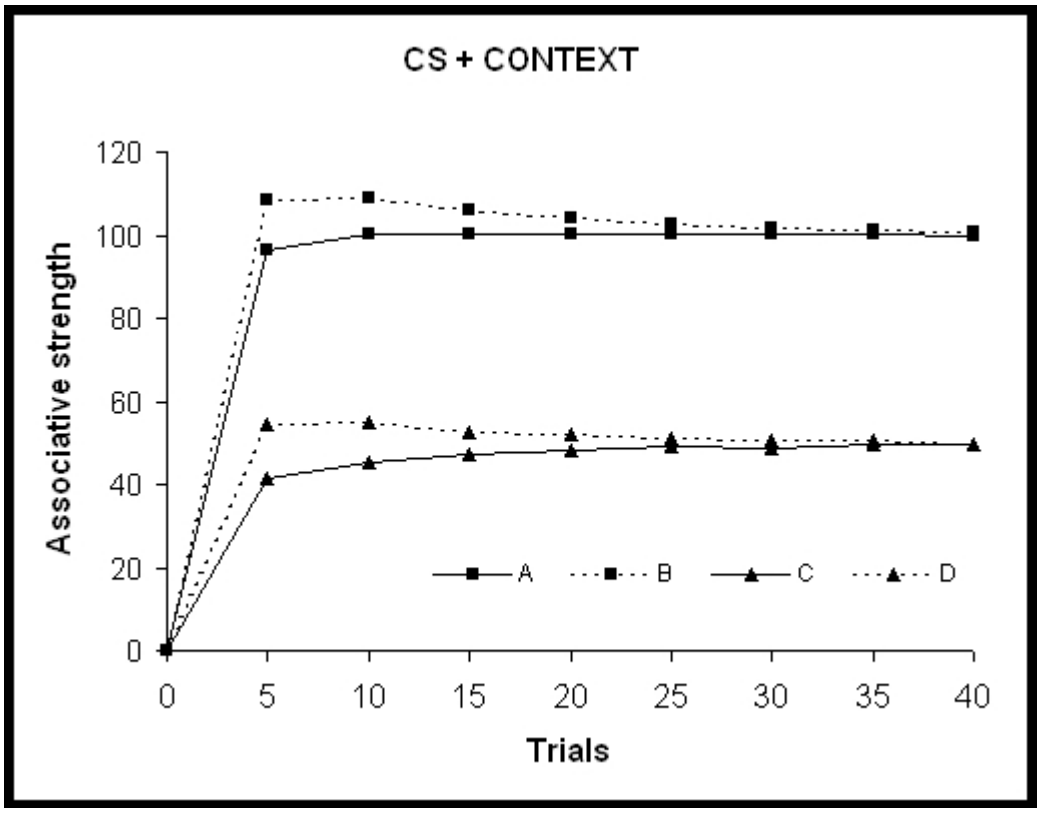


Figure 2. Simulation of the Rescorla-Wagner model showing the sum of the associative strength of the CS and the associative strength of the context under several CS-US contingencies. As in Figure 1, the A series show the simulation in a conditioning situation where p(US|CS) = 1.00, p(US|~CS) = 0.50, Δp = 0.50. In the B condition p(US|CS) = 1.00, p(US|~CS) = 1.00, Äp = 0.00. In C, p(US|CS) = 0.50, p(US|~CS) = 0.00, Δp = 0.50. In D, p(US|CS) = 0.50, p(US|~CS) = 0.50, Δp = 0.00. Learning rate parameters and number of iterations were set to the same values as in the simulations reported in Figure 1.

According to the Rescorla-Wagner model, the CR in a given situation depends on the associative strength of all the stimuli present in that situation (including the context). Therefore, this model predicts that the total strength of the CR when the CS is presented in the training context should be dependent on the probability of the US given the CS. The model, therefore, expects animals to give responses that accurately prepares them for the potential subsequent US.

Resolving the paradox

With the preceding paragraphs we are not trying to conclude that the Rescorla-Wagner model is wrong, but simply that there is an obvious contradiction between the experiments performed by Rescorla (1968) and the Rescorla-Wagner model, and that this contradiction has not been made explicit in the animal or the human learning literature, among other things, because researchers have not drawn an appropriate distinction between predicting a US and assessing the CS-US predictiveness.

The origin of this contradiction might perhaps be found in the measure of CR that Rescorla (1968) used in his experiments. In experiments were aversive stimuli are used as USs, researchers often measure the animal's CR by using a conditioned suppression paradigm. In this preparation the animal is first taught to press a bar to obtain food pellets. Then the animal is exposed to several pairings of the CS and the aversive US. Finally, in the test phase the animal is again allowed to bar-press for food and after it has spent some time pressing the bar, the CS is presented. If the animal is scared of the CS, then it will

probably freeze and give fewer bar-press responses during the CS than in the immediately preceding pre-CS interval. Researchers can thus calculate a suppression ratio as a measure of the amount of fear the animal is showing. What the suppression ratio measures is to what extent the number of bar-press responses is lower during the CS than during the pre-CS period. Thus, the suppression ration does not provide a direct and absolute measure of how scared the rat is during the CS. It is only measuring whether the rat is more or less scared when the CS is presented than it was previously in that context when the CS was not present. Imagine, for example, that the animal had been receiving footshocks both in the presence and in the absence of the CS and that there is a null CS-US contingency. In this situation, the rat might be equally scared and freezing both when the CS is off and when it is on. If the rat is equally scared during the pre-CS and the CS periods, then the suppression ratio would yield a null CR. However, this would not mean that the animal was not predicting the US, but simply that it was not predicting it more strongly during the CS than before the CS.

Therefore, the contradiction between the Rescorla-Wagner model and Rescorla's (1968) experiments could be due to the dependent variable used by Rescorla (1968). In order to decide to what extent Rescorla's (1968) experiments are actually contradictory with the predictions of the Rescorla-Wagner model one would need a measure of CR that directly assessed the intensity of the CR without making reference to the pre-CS level of responding.

Concluding comments

With the preceding paragraphs we have tried to show how the lack of distinction between concepts such as prediction and assessment of predictive value introduced confusion in the study of animal conditioning and human learning. Classical conditioning was supposed to provide animals with a means to successfully predict significant events. Rescorla's (1968) experiments showed that animals were actually assessing CS-US predictiveness. This meant that they were not predicting the US (at least not optimally), but this implication went unnoticed. Moreover, when the Rescorla-Wagner model was proposed, it was thought that it provided an accurate explanation for the sensitivity of CR to the CS-US predictiveness (or contingency). But the model was actually predicting that animals tested in the same context where they had been conditioned should show conditioned responses that reflected an accurate prediction of the US; not CS-US predictiveness. Again, the lack of a clear conceptual distinction between predicting and assessing predictive value (or predictiveness) obscured this feature of the Rescorla-Wagner model.

It is unfortunate that after so many years since the Rescorla-Wagner model was published and after so many experiments testing the model there is still no empirical evidence with animals that would help resolve this paradox. However, as we have shown in the introduction of this paper, there is a growing body of literature showing that at least humans are able to flexibly make predictions, assess predictiveness or assess causal value as a function of what they believe will be more adequate at each time. A closer relationship between researchers investigating animal conditioning and those investigating human learning might help us detect and solve some of these conceptual problems in the future.

References

Cheng, P. W. (1997). From covariation to causation: a causal power theory. Psychological Review, 104, 367-405.
Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. Psychological Review, 99,365-382.
Matute, H., Vegas, S., & De Marez, P. J. (2002). Flexible use of recent information in causal and predictive judgments. Journal of Experimental Psychology: Learning, Memory, & Cognition, 28, 714-725.
Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. Journal of Comparative and Physiological Psychology, 66, 1-5.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appelton-Century-Crofts.
Vadillo, M. A., Miller, R. R., & Matute, M. (2005). Causal and predictive-value judgments, but not predictions, are based on cue-outcome contingency. Learning & Behavior, 33, 172-183.

[1]This only holds if certain parameter values are used. If β is allowed to have different values when the US is present and when it is absent, then the associative strength of the CS would no longer be equal to Δp. However, it would still be dependent on this statistical index (i.e., a greater Δp would give rise to greater associative strength).

[2]Again, this only holds if β is assumed to have the same value in all trials.

Discussion

A question concerning 'predictive value'.
Walter Freeman
Jul 5, 2005 19:46 UT

I am grateful to Matute and Vadillo for their clarification of the difference between predicting a US and assessing the predictive value of a CS. In terms of their long-overdue analysis of the well-known Rescorla-Wagner model, I would like to ask them whether they would accept a restructuring of the notion of assessment of predictive value in respect particularly to interpretation of responses of animals under classical conditioning.

A brain model based on neurodynamics of intentionality holds that prediction of some future state is prerequisite for construction of a plan of action that is accompanied by preafference, that is, the prediction of the sensory consequences of the embodiment of the planned action. More specifically in classical conditioning the prediction is the action to be taken on the basis of whatever may happen, no matter what happens. In a subject that has been conditioned with intermittent CSs and USs the context is thrown into sharper relief than is the case with invariant pairing. In the expectant state the subject is preparing for action based on its available information, which is the context up until the onset of the CS. Therefore the prediction is for the protective action to be taken with or without the CS, which is why subjects commonly curl their pads off the grids.

In this interpretation the concept of 'predictive value' appears to me to be as superfluous as is the concept of 'causality' for interpreting the brain mechanisms of animal behavior. Both concepts invoke a level of abstraction that is inaccessible to animals (and humans) that do not have language. Animals clearly do have the capacity for abstraction, such that they can extract whatever cues they need to apprehend and categorize their contexts (cues that are available to researchers only by inference). What I request from the authors is a brief description of their conception that underlies their distinction between research on animals and humans, which it appears to me may be something more than the capability for language.

Walter J Freeman, M.D. Professor of the Graduate School Division of Neuroscience, Donner 101 Dep't Molecular & Cell Biology University of California Berkeley CA 94720-3206 USA tel 1-510-642-4220 fax 1-510-643-9290 drwjfiii@berkeley.edu http://sulcus.berkeley.edu

Humans are animals, but not only
Juan Rosas
Jul 6, 2005 8:35 UT

I agree with the authors on the need of a correct use of the language in human and nonhuman learning literature. Certainly, terms as prediction and causation are used often as if they were synonymous, and that has lead to what it might be inappropriate confrontations between models to explain the results of the research. Associative models have been used to explain causal learning while, judging from many results in this area, this approach might be inappropriate. Human causal learning might include associative processes. There is little doubt, at least from my line of research, about the similarities between human and nonhuman animal predictive behaviour. Most likely, confrontation between associative and computational theories of learning should be circumscribed to those predictive situations, where they both can make testable predictions. However, humans count with the use of language and reasoning

processes to explain reality. To give a simple example, human behaviour depends on the kind of instructions participants receive during the task or, what it is the same, on the kind of question participants are requested to answer (note that the type of question it is itself a type of instruction, as it guides the participant about what information he or she should use from all the information presented –Vila, 2004, personal communication), something the authors of this paper have clearly shown in their previous work. This kind of result implies that simple associations cannot explain causal learning and, likely, neither they can explain the predictive value of the cue. This is a field that lies away from associative theory. From its point of view, the discussion might be whether causal models are at all necessary to explain human behaviour. But there is not seem to be question about inferential and reasoning processes in some aspects of human learning. That is not to say that human learning does not imply associative processes. There are grounds in the literature (and in life) that suggest that simple associative processes may overcome higher order reasoning processes. Think about a person that takes the subway in a non-busy hour and it is assaulted. Imagine now that the next time this person has to take the subway is a busy hour. Assault is now impossible, given that the amount of people in the train makes prevents movement, and this person "knows" it. However, this person cannot overcome the anxiety that the situation (known as safe) produces, and leaves the train. This description of an actual situation clearly suggest that reasoning processes are not enough to explain human behaviour, as well as associative processes are not enough either. From my point of view, research in nonhuman and human animal learning should focus on those aspects that we share as much as in those aspects that we do not share, but to be able to do so it will be necessary first to agree on how the concepts of causality and prediction should be used in the literature, as the authors of this paper reasonably suggest.

Causality, prediction and modulation
Javier Vila
Jul 7, 2005 1:07 UT

I agree with the authors in that the necessary competencies to predict an event B from an event A are different from the ones that are employed for establish a causal relation among them. And that confusion is a conceptual error in which two different behaviors are confused. It is a good supposition, if also we consider the methodology employed in causal learning studies, the different procedures and instructions are presented and mixed arbitrary this will aggravate the confusion even more.

Nevertheless, if we consider the experiments in causal learning in which participants observe trials repeatedly in which A precedes B. We can think about what is learned in this situation; a) learn to make a prediction of B when A occurs, b) the causal relation from A and B, or both possibilities. The studies cited by the authors (Matute and cols, 2002; Vadillo and cols, 2005), are some examples of procedures in which the humans can learn a causal relation between A and B and they declare it differently according to the final instructions given in question form.

But this fact remits us to the importance of the modulation of the causal or predictive behavior. Finally that is what in last instance determined the occurrence of one or another. It's a clear point that a causal or predictive question is a final instruction that tells us how to respond correctly and this modulation permits us to finish with ambiguous situations in which is not clear how we should behave.

Then, if prediction and causal attribution are two different forms to declare the learned relations, or well two different behaviors both are conditional in their occurrence to the type of final question. And this modulation does not clarify which would be the necessary competences so that a person themselves behaves in a predictive or causal way. And do not tell us if predictive and causal learning are different learnings or different demonstrations from a unique learning process.

We can suppose that this modulation occurs from a basic learning process of causal induction susceptible to rational processes that modify it subsequently in causal or predictive behavior. Nevertheless, because not thinking about the contrary situation, rational processes modulate a basic associative process of learning.

Personally I consider that inside the study of the basic learning processes there are examples of modulation in animals. Thus a symbolic relation between two events can be a guide for a correct response choice in pigeons. Nevertheless if similarities between the modulation of the causal behavior by a final instruction in the shape of question and it observed in animals this even by being studied.

Prediction, predictiviness and adaptive advantage
Teresa Bejarano
Jul 7, 2005 10:34 UT

1) "An animal's CR does not reflect a prediction of the US. The animal, on the contrary, seems to be assessing the CS's predictiviness or predictive value". In my opinion, this makes evolutionary sense. Passive, non-comparative, predictions (i.e., "predictions" in Matute & Vadillo) aren't adaptively useful resources. Animals, on the contrary, cannot live without predictiviness (I have interpreted that predictiviness means 'prediction that is tied to a behavioural plan'). Predictiviness, or comparative prediction, allows animals to choose a particular expectation; then, this expectation drives the behavioural plan that will fulfil it. Certainly, in Pavlovian conditioning experiments, the behavioural plan is very reduced. Animals will choose to stay where they are. In addition, they will not look for the reward. However, this might be an extreme case. Animal abilities can perform more difficult and useful tasks. That is why those abilities were selected.

2)What about women in Berkeley and causality? I think, firstly, that a similar process might be found in animals, and, secondly, that this type of processes does not need a true understanding of causality. Let us suppose that red round things often precede the reward. Has this learning to be transferred to red things, or rather to round things? Progressive experience will drive these changes. Although this is similar to the double cue -"sex, selective program"- in Berkeley, it seems more close to animal abilities. But do these processes involve an understanding of causality? Certainly, human beings will understand that the right cue is the cause. However, in my opinion, this understanding is not a necessary requisite in order to learn the right cue.

Thinking About Action. The Logic of Intervention in Reasoning and Decision Making

Steven Sloman (Cognitive Psychologist, Brown University)

(Date of publication: 5 September 2005)

Abstract: The fundamental idea of the causal modeling framework is that people represent causal systems by decomposing them into autonomous mechanisms that support intervention, both actual interventions on the world and counterfactual interventions in imagination. I will discuss the adequacy of this view as a description of human behavior in one or more of three domains: reasoning, decision making, and learning.

One of the key constituencies of causal reasoning is the effect of action. Action is important, both because it concerns behavior in the actual, physical world and because it is involved when we think about other worlds as we do when we fantasize, imagine, or predict. Thinking about an alternative world involves acting on our model of the world to change it in whatever way our fantasy or image dictates. More fundamentally, causal relations are by definition those relations that support action in the form of intervention. This is the upshot of some compelling recent philosophy (Woodward, 2003, offers a comprehensive review). Very roughly, A is a cause of B if intervening on A would change the value of B, if other things were held constant in the right way. The gravitational pull of the Earth causes the moon's orbit in the sense that if the gravitational pull could somehow be changed, without changing other things, the moon's orbit would change too.

The logic of intervention is unique and does not fall naturally out of conventional propositional logics, even probabilistic logics. Consider an event or property whose value is normally determined by some set of causes. For instance, the height of a blade of grass is normally determined by the amount of sunshine and rainfall, the quality of its soil, etc. Normally, knowing the value of a variable is diagnostic of the value of its causes. If the blade is tall, that suggests more rainfall and richer soil conditions than if the blade is short. But this diagnostic relation fails under intervention. If someone intervenes on the event or property by setting it to some value, the value reveals nothing about its normal causes. If I cut the blade of grass so that it's very short, you should not infer from its height that there's been little rainfall. I call this an undoing effect because the causal linkage to a variable being intervened on from its normal causes must be cut for the sake of inference. A formal discussion of inference from intervention is offered by Pearl(2000) who introduces the DO operator as a mathematical tool for representing intervention. Spirtes, Glymour, and Schienes (1993) present a more general analysis of intervention.

Both common sense and psychological data suggest that people are sensitive to the logic of intervention. But other data demonstrate that people use their own actions as signals, often as signals to themselves about their own nature and dispositions, signals that violate the undoing prescription of interventional logic. My purpose here is to discuss where people's reasoning about intervention goes right and where it goes wrong in hopes of revealing something about the nature of human causal reasoning.

The psycho-logic of intervention

Evidence that people are appropriately sensitive to the logic of intervention when reasoning comes from experiments in which people are given a causal structure and then asked to reason about the causes of an effect that has been intervened on. For instance, in an intervention condition, Sloman and Lagnado (2005) gave people the following causal scenario:

All rocketships have two components, A and B. Movement of component A causes component B to move. In other words, if A, then B. Both are moving.

And then asked them the following questions:

i. Suppose Component A were prevented from moving, would Component B still be moving?

ii. Suppose Component B were prevented from moving, would Component A still be moving?

The simple causal model underlying this scenario looks like this:

Movement of Component A ----------> Movement of Component B

and both components are moving.

To answer question (i), one should mentally intervene on Component A by imagining its movement stopped and note that Component B would then be stopped. The vast majority of Sloman and Lagnado's (2005) participants followed this logic and concluded that "no, Component B would not be moving."

But question (ii) is different. Here, one must imagine the intervention downstream, on Component B. This requires simplifying the causal model by removing any links into the prevented component because the reasoner is setting its value, its normal causes are not. This has no effect when A is prevented because A has no normal causes. But when B is prevented, we must disconnect it from A:

Movement of Component A Movement of Component B

As there is no longer any linkage between the components, it is apparent that Component A's movement is now independent of B's movement and therefore the logic of intervention predicts the undoing effect, that component B's lack of movement is not diagnostic of A. In other words, people should infer that A would still be moving and respond "yes" to the second question. Again, this is what the vast majority of people said.

In contrast, when Components A or B are observed to not be moving, as opposed to being intervened on, lack of movement in B does suggest that A is not moving just as A's lack of movement suggests that B is not moving. In other words, when asked what to expect after observing no movement:

i. Suppose Component A were observed to not be moving, would Component B still be moving?

ii. Suppose Component B were observed to not be moving, would Component A still be moving?

participants should respond "no" to both questions, and that's just what the great majority did. Corroborating data can be found in Waldmann and Hagmayer (2005).

People also show sensitivity to the logic of intervention when making decisions (Hagmayer & Sloman, 2005). Imagine an action that has some desirable causal consequence (e.g., doing the chores improves health by providing exercise). This knowledge should and does increase people's willingness to do the chores. But you might be told instead that the action and consequence are correlated by virtue of a common cause with no direct causal relation between them (e.g., doing the chores and good health are correlated because both are consequences of a caring attitude found in only some people). In this case, doing the chores does not increase the chance of good health. In fact, the choice to do the chores is an intervention that renders the chores independent of one's degree of health. So this causal model should not affect one's willingness to do the chores and, in line with the logic of intervention, it had little influence on Hagmayer and Sloman's participants. A general review of studies demonstrating sensitivity to the logic of intervention can be found in Hagmayer et al. (in press).

There is also strong evidence that intervention facilitates learning relative to mere observation both in children and adults (for a review, see Lagnado et al., in press). The advantage of intervention in learning may be in part a result of implicit temporal cuing or attentional cues rather than the model-changing informational value of intervention due to undoing (Lagnado & Sloman, 2004).

Failures of interventional logic: Signaling

The undoing effect is not consistently observed in human inference. Violations are seen in the form of instances of signaling (Bodner & Prelec, 2002). Quattrone and Tversky (1984) asked a group of students to hold their arms in very cold water for as long as they could. Half of the group was told that people can tolerate cold water for longer if they have a healthy type of heart, while the other half was told that the healthy heart causes lower tolerance. The first group held their arms in the water for longer than the second even though both groups claimed that they were not influenced by knowledge of the link to heart quality. In other words, they used their tolerance for cold water as a signal that they were healthy by not being aware of or denying the influence of the heart hypothesis on their action. They acted as if they had not intervened when in fact they had. On the assumption that participants were making every effort to be honest, and really did not believe that the hypothesis influenced them, signaling in this case was a form of self-deception. People deceived themselves into believing they had not intervened in order to use their action as a signal that they were healthy. The signal may have been directed only toward participants themselves; there may have been no one else they cared to convince.

Shafir and Tversky (1992) report another violation of interventional logic. They gave participants a real-life version of Newcomb's paradox (Nozick, 1969). In brief, people were given a choice between two prizes where the second prize was always larger than the first (the second dominated the first), although the amount of the prizes was uncertain. They were told that a computer had used earlier choices they had made as a basis for predicting their choice between the first and second prize and if it predicted that they would choose the small prize, there would be much more money available for both prizes than if it predicted that they would choose the larger prize. Participants always chose after the prediction was made so their choice could not influence the prediction. Nevertheless, 65% of participants chose the small prize, giving up money in the hope that their action would render a prediction of their choice accurate. Choice is an intervention in the sense that it is an action that renders a selection independent of its normal causes, in this case rendering the amount chosen independent of the prediction. So participants acted as if their choice signaled a prediction that it could not possibly have influenced.

The rationality of signaling

Is signaling an error by virtue of violating interventional logic? Clearly, people are behaving non-optimally in the examples just cited. The Quattrone and Tversky (1984) example is a case of an action performed merely for its signaling value (to convince the participant, experimenter, or both that the participant has a good heart) and yet the very act reduces the value of the signal (because it is designed to convince rather than reveal the participant's actual heart type). The Shafir and Tversky (1992) study demonstrates how decisions made for their signal value can lead to nonconsequentialism. People gave up money in an obviously futile effort to gain more money.

But a couple of examples don't condemn all cases of signaling. A nickel has negligible value yet stealing a nickel remains an outrageous act, not because of its consequences but because of its meaning: It signals (to oneself if no one else knows about it) that one is a thief. The symbolic value of certain acts provides sufficient justification for those acts even in the absence of desirable causal consequences of the action (see, e.g., Nozick, 1995).

Moreover, our decisions can give us information about ourselves. I might not be aware that I prefer one type of beverage or location or partner, yet I might notice that I am consistent in my choices, and this tells me something about myself. Some people consistently choose partners who dominate them and learn by this about their own desire to be submissive. People's choices can have multiple determinants and one can sometimes learn about those determinants by observing one's own choices. Self-signaling is not an error when it tells us something about ourselves that we do not already know.

Nevertheless signaling remains problematic. It can even lead to paradox when an action is performed merely for the sake of its signaling value (Campbell & Sowden, 1985). Consider someone who is not altruistic yet wishes they were. They might donate money to charity to signal that they are altruistic, yet their action is not a result of altruism but the desire to signal altruism. Of course, the act itself is altruistic, so how should we judge the person? In this case, there's no causal link from the individual's conscience

to their action, although the diagnostic value of the act suggests such a link. As Bodner and Prelec (2002) point out, examples of self-handicapping have a parallel logic. Someone might not study for an exam, or might dress inappropriately for an important meeting, in order to have a ready excuse when they do not meet expectations.

Conclusion

These examples do not invalidate the logic of intervention. But they do force some care in deciding what is and isn't an intervention. Alternatively, we can distinguish different types of intervention. To see which interventions should and which need not follow the logic of intervention, we need to distinguish the process of decision making (a cognitive activity) from the process of choice (determining the state of the world via the action of selecting an option from a set). The process of decision making takes information from the world and analyzes it to select a course of action. Choice is an action that is either fully determined by the decision-making process or else affected by other factors as well.

Let's first consider choices that are fully governed by the decision-making process, for example, my choice on a multiple-choice multiplication test. I do my calculations and they determine my choice, nothing else affects it. In such a case, the process of choice cannot teach us anything that we do not already know. I chose my answer because I did my calculations and decided that was the right answer. I cannot learn anything new from my answer about what the question was or how I answered it. I knew as much about those things before I chose my answer as I do immediately afterward. Call cases where the decision-making process fully determines choice deliberate intervention. Such cases are fully consistent with the logic of intervention -- with undoing -- inasmuch as the choice itself does not reveal anything new about its determinants.

But not all choices are completely deliberate. My willingness to donate to charity might not be fully governed by my decision-making process. There are likely social pressures that I'm not aware of, pressures that charitable organizations are likely to exploit. The amount that I donate may be affected by the way options are listed on the donation form, by halo effects, by recent memories, indeed by a host of subtle influences that may affect me unwittingly. My final contribution cannot inform me about those factors that entered my decision-making process because, by definition, I knew about them before making my choice. But my final contribution can inform me about those other factors that influenced me that were not taken into account by my decision-making process because I may not have known about them. I might even be surprised by the amount of my own donation, in which case it's likely that one or more of these "hidden" causes are driving me. The choice is not diagnostic of the input into my decision-making process, but it is diagnostic of these other causes of my choice.

In sum, a rational analysis of intervention requires distinguishing causes of action that influence the decision-making process and causes of action that do not. A choice should be construed as an intervention only to the degree that the choice is governed by the process of decision making and not by other factors. Deliberate interventions entail treating choices as non-diagnostic of their causes; other interventions entail this kind of undoing only to the degree that choice is governed by decision making. But note that normative considerations demand that my beliefs about those factors that do govern the decision-making process are not influenced by choice; such beliefs are governed by undoing because the decision-maker already knows as much about them before choice as he or she does after choice.

The conclusions so far have focused exclusively on decision making. But the same distinctions arise when reasoning in the absence of a decision. To answer a question like "what would be the optimal intervention in such-and-such a situation?" I might use a causal model to think about the situation. To the degree that my answer is governed by my thinking, any intervention I consider should be construed as a deliberate intervention and not diagnostic of its normal causes. But to the degree that my answer is governed by factors that don't enter my causal model, like an unconscious desire to intervene in a specific way, the intervention is diagnostic of those factors. This situation is really not so different than decision making and might just be a special case of it.

However, if I am given an intervention and asked to reason about it, then I must always treat it as a deliberate intervention. To answer the question "what would happen if I intervened on X?" a reasoner

always has complete control over the variable intervened on because what we imagine to be true just is true in our imaginary world. In that sense, imagined interventions in reasoning should always obey undoing; they reveal nothing about causes other than the intervention because the intervention uniquely determines their value.

Whether or not people are obeying the logic of intervention when reasoning and decision making, their thinking always seems to be guided by causal beliefs. Even the signal value of actions derives from belief in a causal mechanism determining the action. Without the belief that heart type causes more or less tolerance for pain, participants would not have been motivated to hold their hands in water for more or less time. So even when people are violating the logic of intervention, a causal model guides their actions.

Note. This essay has benefited from the comments of York Hagmayer.

References

Bodner, R. & D. Prelec. (2002). Self-signaling and diagnostic utility in everyday decision making. In Collected Essays in Psychology and Economics. I. Brocas and J. Carillo (Eds.), Oxford University Press,
Campbell, R. & Sowden, L. (1985). (Eds.), Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem. Vancouver: University of British Columbia Press.
Hagmayer Y., & Sloman, S. A. (2005). Causal Models of Decision Making: Choice as Intervention. Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society, Stresa, Italy.
Hagmayer Y., & Sloman, S.A., Lagnado, D. A., & Waldmann, M. R., (in press). Causal reasoning through intervention.In Gopnik, A. & Schulz, L. (Eds.), Causal learning: Psychology, philosophy, and computation. Oxford: Oxford University Press.
Lagnado, D. A., & Sloman, S. A. (2004). The advantage of timely intervention. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 856-876.
Lagnado, D. A., Waldmann, M. R., Hagmayer Y., & Sloman, S.A., (in press). Beyond covariation: Cues to causal structure. In Gopnik, A. & Schulz, L. (Eds.), Causal learning: Psychology, philosophy, and computation. Oxford: OxfordUniversityPress.
Nozick, R. (1995). The nature of rationality. Princeton. Princeton University Press.
Pearl, J. (2000). Causality. Cambridge: Cambridge University Press.
Quattrone, G. & Tversky A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. Journal of Personality and Social Psychology, 46, 237-248.
Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. Cognitive Psychology, 24, 449-474.
Sloman, S. A., & Lagnado, D. (2005). Do we “do”? Cognitive Science, 29, pp. 5-39.
Spirtes, P., Glymour, C., & Scheines, R. (1993). Causation, prediction, and search. New York: Springer.
Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. Journal of Experimental Psychology: Learning, Motivation, and Cognition, 31, 216-227.
Woodward, J. (2003). Making things happen. A theory of causal explanation. Oxford: Oxford University Press.

Discussion

The examples fail in Real World applications
Robert Stonjek
Sep 5, 2005 14:15 UT

What is missing from Steven Sloman’s opening arguments is any mention of relativity, without which is examples make no sense.

The only way of changing the Earth’s gravitational pull is to change the mass of the Earth. We can introduce a Science fiction device for teleporting matter for the purposes of discussion. Once done the Earth’s gravitational force weakens and the Moon’s orbit changes, but the Earth is also influenced by the moon’s gravitational force, relatively more so if the Earth loses mass.

If the mass of the Earth and Moon were equal, then they would orbit each other as equal partners so that neither one is a satellite of the other.

So the moon's gravity affects the path the Earth takes through space, and the Earth's gravity affects the path that the moon describes through space. Reducing the Earth's gravity effects both moon and Earth.

The grass analogue is closer to home. The main determinant of the height of the grass is its genetic predisposition – the genes it has inherited. The effect of water can only be considered as a determinant when the relative height of grass is considered ie relative to the same species of grass with less or more water.

Intervention, then, is useful as a tool only if one has a control (the relative value with which one is to make subsequent comparisons) and can be sure that no uncatalogued interventions occur during the period of observation ie no contamination, noise or other unwanted 'interventions'.

On the rocketship question, the two components of the simplest possible rocket are the rocket structure (let this be B) and the thrust (let this be A). The thrust moves in the opposite direction to the rocket and propels it. Near the Earth, if the thrust stops then the rocket falls back to earth, eventually stopping, but in space the rocket continues to move, perhaps forever.

If the rocket is prevented from moving, say by being fixed to the launchpad, then the thrust does move but the rocket remains stationary.

Thus an actual example of A and B shows that i) if B (the rocket) is prevented from moving then A (the thrust) continues, (eg trying to climb out of a strong gravitational field): the answer is "yes". ii) If the thrust stops and the rocket is near Earth then the rocket will continue moving for a finite period, until it falls back to Earth, but if in space, which is where most rocket ships are to be found, then A continues. There is not enough information to determine an answer.

I am unaware of any examples of propulsive units that would fit the description of B and return the answers given in the paper.

If the school of logic is to contribute to science, they must at least get their basic science right so that prospective recipients of what might be useful thinking tools do not discard the paper in disgust before reading down to the crux of the issue.

I guess the real world isn't my bag
Steven Sloman
Sep 7, 2005 19:29 UT

I specifically chose the solar system example because it is not amenable to actual, physical intervention (not easily anyway). The point was to use an example in which intervention had to be done in imagination, on a model of the earth-moon system. In such a case, changing the Earth's gravitational pull would indeed change the moon's orbit whatever else also gets changed. Yes, actual intervention in the world indeed runs into all the problems that running experiments does, and there are many of those indeed. Fortunately the logic of intervention remains invariant whatever complexities arise in the real world, and fortunately imagined interventions don't suffer from all these complexities. By the way, if I were looking for a rocketship, I'd probably do most of my searching on Earth, not in outer space.

Signaling and intervening
Gloria Origgi
Sep 7, 2005 18:09 UT

Sloman says that self-deception when people consider their actions as a signal of a certain state of affairs is not a failure of causal reasoning but a failure of interventional logic, that is, people treat a certain correlation as it were indipendent of their own intervention whereas it is dependent on it. But it seems to me that signaling cases are a failure also of normal causal reasoning, that is, people invert the role of causes and effects by giving causal powers to actions (as for example keeping the arm in the cold water for longer) that do not have any causal power on a particular state of affairs. Their causal reasoning about the correlation between having a healthy heart and keeping an arm longer under cold water is incorrect. It seems to me the the failure of interventional logic in these signaling cases depends on a wrong way of framing the causal structure of the situation.

> **Where is the failure?**
> Steven Sloman
> Sep 7, 2005 19:18 UT
>
> I see the logic of intervention as a critical part of causal logic, so any failure of the logic of intervention is also a failure of causal reasoning. Origgi may be correct that in cases of self-deception people are inverting a causal relation, but there is another possibility. It may be that people manipulate the effect on the belief that the value of the effect SIGNALS the value of the cause by virtue of the effect's diagnosticity for the cause, without believing that the effect has causal efficacy. One or both of these beliefs may underlie self-deception. It's an empirical question which. It matters in that signal value can sometimes be justified; I don't think inverting a causal relation can ever be.

Intervention and development
Anne Reboul
Sep 8, 2005 10:19 UT

Steve's contribution introduces the important notion of intervention. Intervention can both correspond to an action performed by the subject or to an event observed by the subject. In the first case, regarding the development of the logic of intervention, a more or less Piagetian view might be adopted. However, with very young children, whose sensori-motor development is still very limited, one would expect the logic of intervention to be more or less inexistent. Yet, some of the folk physics experiments using the habituation/dishabituation paradigm with fairly young children could be interpreted as showing some abilities in the logic of intervention: for instance, in one of the classical experiments, infants are shown a red ball hitting a blue ball and causing it to move (intervention) or the red ball stopping before it hits the blue ball which nevertheless moves (independence). Children look longer at the second situation which has been interpreted as showing surprise. I have two questions: does it make sense to interpret those very early abilities as showing some inkling of the logic of intervention? If it does, what could the developmental story be?

> **Developmental work on intervention**
> Steven Sloman
> Sep 8, 2005 13:07 UT
>
> The example that Anne describes strikes me as more a demonstration that infants are sensitive to the specific spatio-temporal characteristics of physical contact, or of what Michotte called "launching." But she's absolutely right that young children are sensitive to the logic of intervention at least in the context of learning causal relations. Several developmental psychologists have been studying whether young children can use intervention to aid learning including Laura Shultz and Dave Sobel. To illustrate, Shultz presented 4-year-olds with two creatures A and B. The children were told that one of the two creatures was the boss, and that the boss made the other creature move but that the other creature didn't make the boss move. The children had to figure out whether the boss was A or B. No amount of observation of A and B's movement would tell them because A

and B were always moving together so there were no temporal cues to help. The kids were then shown a button that made one of the creatures (B) move. After an intervention – pressing the button – they discovered that only creature B moved. Observing this intervention was sufficient to convince the kids (and adults too) that creature A was the boss.

A diagnostic interpretation of the signaling exemple
Anne Reboul
Sep 8, 2005 10:43 UT

Steve gives the interesting example of students holding their arms in cold water for a longer or shorter time, depending on what they'd been told about the relation between having a healthy heart and being able to hold one's arms in cold water. This difference in behavior is interpreted as linked to signalling (in this case, possibly to oneself) that one is healthy. However, I think that another interpretation is possible, in terms, not of signalling but of diagnostic. The idea is that, after all, one usually doesn't know the state of one's own heart. So, being told of a relation between cardiac health and a physical ability and given an opportunity to put one's heart to the test, one might just go ahead and do it. This does not mean that no element of wishful thinking does intervene (for instance, one might expect the group that has been told that healthy hearts have less tolerance for the behavior to give up earlier), but that, though the subjects fail to appreciate the impact of what they've been told on their behavior, they do, at the same time, manifest a correct causal reasoning, using the causal connection between cardiac health and tolerance to cold water as diagnostic in their own case. In other words, taking the positive scenario, a subject might think that, given that a healthy heart is causally related to tolerance, holding his arms in cold water for as long as he can will (not only signal) but demonstrate either the soundness and the unsoundness of his cardio-vascular system. A final remark: regarding the possible paradox involved in performing an action for the sake of its signaling value, a near version of this can be found in what Johansson has called "antiperformatives", which are self-falsifying assertions (e.g., "I'm very humble"). This may be limited to a set of more or less ethic situations or attitudes (altruism, humility, etc.). If this is the case the paradox may be sufficiently limited not to be a general threat to either general signalling or rationality.

REFERENCES Johansson, I (2003), "Performatives and antiperformatives", Linguistics and philosophy 22/6, 661-702.

a diagnostic interpretation
Steven Sloman
Sep 8, 2005 19:04 UT

Quattrone and Tversky ran two groups: One was told that a good heart gave people the ability to hold their arms in cold water for longer, the other group was told the opposite, that a good heart caused less tolerance for pain. I interpret the fact that those in the first group held their arms in cold water for longer than those in the second group as reflecting either i. an attempt to change the quality of their heart or ii. an attempt to signal a strong heart. i. requires the belief that participants could change the quality of their heart which is on the face of it implausible. However, it's possible that such a belief was operative at an unconscious level. ii. is more likely. My interpretation was much like Anne's, that they were treating the causal link from heart quality to time in cold as supporting a diagnostic relation in the reverse direction. The longer they held their arms in cold water, the more evidence they acrued that they had good hearts. Of course, this inference is only justifiable if heart quality, and not desire to have a healthy heart, determined the length of time spent. The fact that the two groups behaved differently suggests that it was the desire to signal a healthy heart that was critical. The signal of course obtained its value from the diagnosticity of the time spent in the cold; but the difference between two groups of otherwise identical individuals can only be explained by their desire to appear to have good hearts, not by the actual quality of their hearts. I can't see how to rationalize this behavior as it implies that a non-diagnostic signal was being treated as diagnostic.

Signalling: – conscious or subliminal?
Robert Stonjek
Sep 8, 2005 12:38 UT

The section titled “Failures of interventional logic: Signalling” describing the cold water tolerance experiment which showed that participants are influenced by information regarding the health of the heart and tolerance to cold water, makes naïve (with respect to cognitive science and to paths well trodden by psychology) assumptions about what participants believed and what their conscious intensions were.

The assumption is that individuals were consciously aware that they were influenced by the knowledge of the health of the heart relative to cold water tolerance, that they consciously adjusted their demonstration of tolerance (by removing their arms earlier or later than they otherwise would have) and that they consciously tried to mislead experimenters by misrepresenting their deception vis “They acted as if they had not intervened when in fact they had.” And “On the assumption that participants were making every effort to be honest, and really did not believe that the hypothesis influenced them, signalling in this case was a form of self-deception. People deceived themselves into believing they had not intervened in order to use their action as a signal that they were healthy.”

These are large assumptions and don’t bear up to scrutiny. It is well known and easily demonstrable that subjects can be subliminally influenced in a variety of ways. In ‘Placebo: The Belief Effect’ by Dylan Evans (Harper-Collins, 2003) we read about numerous examples of non-active pills having a beneficial effect that sometimes equals the active drug, at least in the short term, when treating conditions such as anxiety, depression or pain.

If the placebo effect were one of conscious decision making where sufferers who benefit from placebos “acted as if they had not intervened when in fact they had.” then there would be no excuse for these conditions – sufferers could intervene at any time. The ‘Placebo Effect’ is not a form of ‘self deception’.

It is far more likely that participants in the above mentioned experiment (cold water tolerance) were subliminally effected by the erroneous information regarding the health of the heart and tolerance for cold water planted by experiments, that they genuinely believed that that information was inconsequential and that they had followed instructions and held there arms in the water for as long as they could.

The following section “The rationality of signalling” makes the opposite assumption – that people are not aware of the subliminal influences that guide some portion of their behaviour, particularly where personal preferences are concerned.

Why is it that we must infer from observing our choices what the underlying influences must be yet we are cognisant of a subliminal influence that biases our behaviour in a cold water tolerance experiment?

This seems inconsistent and presents problems where perfectly logical deduction is to be applied to real world conditions. It will be difficult for such plausible and possibly utilitarian logic to gain traction in the areas of its greatest utility if inconsistencies can’t be separated by at least a more than a few paragraphs.

> **Exactly wrong**
> Steven Sloman
> Sep 9, 2005 13:57 UT
>
> Maybe Quattrone & Tversky's assumptions are naive, but they were exactly the opposite of those stated by the respondent. As stated very clearly in the text, participants in their experiments claimed that they were NOT influenced by knowledge of the hypotheses and, taking those assertions at face value, Quattrone & Tversky (and I) assume that any influence on pain tolerance was unconscious. So at least on this point the respondent and I have no disagreement.

Interdisciplinary Semantics?
Robert Stonjek
Sep 11, 2005 13:14 UT

If the purpose of an interdisciplinary forum is to test the meshing and cooperation of a diverse range of disciplines in solving issues which may require such an interdisciplinary approach then accuracy of rhetoric, without necessarily becoming tedious, is essential. In particular, phraseology having specific meaning only in a particular discipline should be avoided or balanced with an explanation upon the first useage of such discipline-specific phrases (or words)

After reading the response to a previous note of mine (**"Signalling: – conscious or subliminal?"** and the response **"Exactly Wrong"**) I see that the following statement:

"In other words, they used their tolerance for cold water as a signal that they were healthy by not being aware of or denying the influence of the heart hypothesis on their action."

actually meant that 'they' were not using cold water tolerance as a signalling device but in fact were fooled by 'their' own subconscious into unwittingly signalling.

This may seem to be hair splitting when whole subjects are considered, but cognitive sciences, particularly those dealing with consciousness, must differentiate the entire individual's behaviour from consciously willed, initiated or monitored behaviour.

The above quoted sentence, typical of the paper with regard to conscious-subconscious intentionality, readily confounds the consciously monitored and subconscious only behaviour eg *"…or denying the influence of the heart hypothesis on their action."* reads very much like a consciously considered response to an experimenter's probing.

To be consistent across disciplines, the use of descriptors such as "they", "he", "she" can not be arbitrarily assigned to conscious, subconscious, or whole person behaviour without some portion of the readers of that paper confounding the author's intentions.

In consciousness studies, in particular, 'they' in combination with behaviour usually indicates a subject's consciously monitored behaviour such that upon later questioning a subject consciously recalls that behaviour (action, perception, response etc).

It is heartening that upon resolution of this conflict we are in some agreement, though from an earlier note I would point out that 'rocketships' are designed for space and spend most of their active life there, though non-moving (before and after a flight) rocketships can be found more readily here on Earth (as Steven pointed out).

Death as an Empirical Backdoor to the Representation of Mental Causality

Jesse M. Bering (Cognitive scientist, University of Arkansas)
(Date of publication: 3 October 2005)

Abstract: Investigating peoples' understanding of death may help to disentangle the complex relationship between the causal construals of self and other. From a simulationist perspective of other minds, because representations of postmortem mental states cannot be informed by firsthand experience with personal death, theoretical constructs dealing with the self and others' minds after death suffer from the logical impoverishment of hypothesis disconfirmation. As lamented by the Spanish philosopher de Unamuno, "the effort to comprehend it causes the most tormenting dizziness. We cannot conceive of ourselves as not existing" (1954, p. 38). Simulation constraints may lead to a number of telltale errors, namely Type I errors (inferring mental states when in fact there are none), regarding the psychological status of dead agents. But even if one does not embrace simulation theory, reasoning about the fate of minds after death in a materialist fashion is problematic.

Introduction

One of the hallmarks of intentional agents is the fact that natural selection has rubber-stamped their causal systems with an approximate expiration date. Miracles and medical caveats aside, brains become irrevocably defunct at biological death. Dying therefore makes intentional causality special because mental states, unlike other causal states, are absolutely absent in dead organisms. Consider the death of a human being. A carcass might tumble out of its coffin (thus the body retains physical causality); or enzymes from a corpse might seep into the soil of its grave (thus the body retains biological causality). But alas, the body has entirely lost its capacity for intentional causality. This is because the brain — that "thing" which in better days caused this body to stand up, to dance, to cry, to laugh — abruptly halts its production of mental states whenever it expires or exceeds its tolerance for assault.

All of this is bound to strike a scientific audience as painfully obvious. Indeed, one need not have extensive training in cognitive neuroscience to recognize death's impact on the mind; anybody who has thought about the topic with any empirical gusto is likely to find themselves adopting a materialist stance. But this "Duh" effect is precisely why the study of death should be of interest to cognitive scientists. If it is so obvious that mental states are obliterated by death, then why do people the world over believe and behave otherwise?

Let me not prematurely raise hopes: We don't really know yet. But as I discuss in this article, there's reason to be optimistic that cognitive science holds the key to answering this question. Of course, cognitive scientists are not the first to take an interest in this topic; we are in fact nipping at the heels of over a half century's worth of psychology that has done the emotional bricklaying. From Freud's more vacuous concept of "wish fulfillment" to Ernest Becker's inspired "Terror Management Theory" and its huge catalogue of elegant experiments on death anxiety (e.g., Dechesne et al., 2003), psychologists in this area historically have focused almost exclusively on fear of death. Central to these models is the notion that human beings have 'invented' the idea of an immortal soul to allay their concerns about nonexistence. (Below I review recent evidence that calls into question this popular model of afterlife beliefs.) The reason that cognitive science is now nipping at these heels is to help orient psychologists to the fact that any human emotions that play a role in reasoning about the mind's fate after death are dependent on the more basic capacity to think about minds.

A Lacuna of Comparative Investigations

One might therefore begin by asking whether other species reason about death in a fashion similar to human beings. Miguel de Unamuno, the Spanish philosopher who penned the notoriously brooding The Tragic Sense of Life, suggested that "The gorilla, the chimpanzee, the orangutan, and their kind, must look upon man as a feeble and infirm animal, whose strange custom it is to store up his dead. Wherefore?" (1926, p. 20; italics added). From an evolutionary perspective, this is an important question because it bears on the phylogeny of mental representational systems. If the soul is, in essence, the disembodied

mind, then the capacity to reason about this mythical entity is quintessentially rooted in evolved social cognition. Comparative studies that investigate other species' reasoning about death would therefore be fruitful if they helped to map out the evolutionary origins of cognitive traits such as theory of mind.

Unfortunately, to my knowledge no comparative psychologist has taken on this specific task, so we have only anecdotal reports to frame our current understanding. When reviewing these cases, we are wise to be wary of the human penchant for seeing our own minds mirrored in the behaviors of other species. E. O. Wilson (1971) offers a good cautionary tale about the hazards of extravagant theorizing in this area (read the many popular science accounts of elephants' fondling their grandmothers' bones and great apes mourning themselves sick). He once described how ants (Pogonomyrmex barbatus) routinely inspected with their antennae day old corpses of nest mates that had been decomposing in the open air. The first ant to find the corpse picked it up and carried it off to the refuse pile, where it was safely removed from healthy sister workers. Do ants therefore understand death as death? Probably not. As Wilsonobserved, it is just natural selection's way of keeping inclusive fitness levels at a healthy high. A closer look revealed that these predictable behavioral responses were motivated by biochemical cues; bits of paper daubed with acetone extracts of ant corpses were treated by the workers in an identical fashion — they too were carried off to the refuse pile!

Some anecdotes are relevant to the current discussion, however, because they detail natural behavior patterns that show a reliance on causal agency cues in distinguishing between a dead organism and a live one. Both de Waal (1996) and Goodall (1990), for instance, have described the behaviors of chimpanzees (Pan troglodytes) in response to the deaths of group members as involving (i) close inspection of the corpse; (ii) attenuated social arousal and (iii) momentary group cohesion (see also Teleki, 1973). Miller and Brigham (1998, p. 79) even describe an incidence of "ceremonial gathering" by black-billed magpies (Pica pica) that occurred in response to the mid-flight death of a solitary bird: "Within 5 minutes a flock of 13-14 magpies had gathered in a circle around the body. . .There was no overt sign of aggression nor did the birds make any attempt to scavenge from the carcass. . The remaining birds stayed around the carcass for at least 5 minutes before spontaneously flying off together." Barrett and Behne (2005) have recently identified agency detection mechanisms as central to human reasoning about death as well. (To illustrate, simply imagine a runny-nosed second grader using a stick to poke at a dead raccoon on the road.) Corpses are one of the great ambiguities of nature. They place unusual strain on inference systems devoted to reasoning about hidden causes because a body's absence of action does not necessarily entail an absence of intentional states.

In one of the few serious theoretical papers in this area, Allen and Hauser (1991, p. 231) suggest that our species may be qualitatively unique because we can mentally represent the concept of death. They argue that: "Humans are capable of recognizing something as dead because they have an internal representation of death that is distinct from the perceptual information that is used for evidence of death. It is this separate representation that is capable of explaining the human ability to reason about death rather than merely respond to death in the environment." Following this logic, the authors proceed to describe several possible experiments that would test the hypothesis that our species is alone in harboring this internal representation. For example, they suggest that researchers might observe the behaviors of female vervet monkeys who are listening to vocalization playbacks of distress calls from their recently dead infants. Presumably, if the mother ignored her dead infant's distress call, this would be evidence that she has "turned off" her localization and search response because her internal representation allows her to appreciate the finality of death. There may be better — and less cruel! — ways to test such hypotheses, but it is indeed critical to put the many comparative anecdotes on death through the sieve of the scientific method, just as these authors advise.

Autistic Souls?

Because their impoverishments are localized to social cognitive dysfunctions, investigating autistic people's death concepts may also prove informative. Do — or, more properly, can — autists believe in souls? Unfortunately, as is the case for comparative research on intuitive reasoning about death, no controlled studies have yet been conducted with autistic people. There are a handful of anecdotal reports to pique our curiosity, however. In her autobiography Thinking in Pictures: And Other Reports from My Life with Autism, TempleGrandinsoliloquizes on immortality after describing her invention of a humane

slaughtering apparatus for cattle. Indeed, she devotes her final chapter to the subjects of Heaven and God. Of particular interest is a diary entry which blandly reads: "I believe that a person goes on to somewhere else after they die. I do not know where" (1995, p. 197).

Evolutionary Hypotheses

In light of the question I asked at the outset of this paper ("If it is so obvious that mental states are obliterated by death, then why do people the world over believe and behave otherwise?"), it seems reasonable to postulate possible genetic advantages associated with this ubiquitous 'irrationality.' In both hunter-gatherer and modern societies, the fear of ghosts abounds (e.g., see Reynolds & Tanner, 1995). In children, this fear rivals such evolutionarily plausible fears as those of snakes and spiders, and it is apparently even more resistant to treatment than fear of strangers (Gullone et al., 2000). Recently, Katrina McCleod, Todd Shackelford and I hypothesized that the fear of ghosts may facilitate the inhibition of selfish behaviors, thus preserving reputation in situations where individuals underestimated the risk of detection by living group members (Bering, McLeod, & Shackelford, in press; see also Boyer, 2001). Sometimes, of course, it pays to cheat, but in general the costs of underestimating the risk of social detection would have been disproportionately greater than the costs of prosocial decisions that were contextually maladaptive.

Our study was partially inspired by another recent study by Burnham and Hare (in press), who reported that, in anonymous and final interactions, participants contributed significantly more to a public good when 'watched' by a robot with large, human-like eyes (see also Haley & Fessler, 2005). Although their hypothesis was that human eyes would trigger non-conscious mechanisms that gauge privacy, and thus serve to elicit prosocial behaviors, we suspected that the presumed presence of a ghost in the room may similarly prime cooperative effort.

Our experimental design presented undergraduate students with several opportunities to cheat at a challenging spatial intelligence test, which promised the highest scorer a generous monetary prize. On a random sample of mental rotation items, the correct answer was "accidentally" revealed on the screen prior to the question. In both the written and verbal instructions, participants had been informed earlier that, due to a glitch in the computer program, this would periodically happen. If it did, they were told, they should immediately press the space bar so that they could "answer the question honestly." The catch was that a third of these people had also been told earlier that the ghost of a dead graduate student had recently been spotted in the laboratory. Another group simply read an In Memoriam announcement in honor of this dead student, while the remainder heard nothing about the fictitious decedent. As predicted, there was a significant effect of condition on latency of response (i.e., how quickly the participants pressed the space bar to clear the screen of the correct answer) on glitch items. Participants who were randomly assigned to the ghost condition had faster response latencies than those in the control group and those in the In Memoriam group.

'Signs'

This inhibitory effect hypothesis of ghost primes has additional adaptationist veracity when considering that many people are under the impression that the dead are arbiters of fortune; they are often seen as meting out punishment for social transgressions. In a recent review, Dominic Johnson and I poured through the Human Relations Area Files (a vast archive of cultural ethnographies that includes many societies which have since gone extinct) for evidence of the cross-cultural tendency to ascribe to the dead responsibility for natural events (Bering & Johnson, 2005). To say that this is a culturally recurrent phenomenon would be an understatement. Thus, not only do many people believe in ambient spirits, they also believe that these spirits are causal agents in the environment, intentional agents that use natural events as signs to communicate messages (often symbolizing their discontent).

Given the right emotional construal, I'd wager that everybody — materialist and dualist alike — is susceptible to knee-jerk attributions of intentions to disembodied minds, whether they profess to 'believe' in such things or not. For instance, I am an unswerving atheist and I fully 'believe' that my mind will burn out in harmony with the death rattle of my brain cells. Furthermore, I confess that on more than one occasion I have shivered with shock and repugnance at the dogged naiveté of the faithful. And yet — and

yet — on the eve of my mother's death, when the wind chimes outside her window began to sound with no noticeable breeze in the air, I intuitively joined my more religious siblings in their chorus of supernatural affirmation. Under similar circumstances, who among us wouldn't fleetingly feel that this was their dead mother's signal from the great beyond? (Rest assured that my consciousness immediately smothered this intuition with sound scientific reasoning.)

There were at least three possible ways that I might have reasoned about the cause of the wind chimes' movement that evening. I might have: (i) intuitively thought that the movement was brought about by some imperceptible brush of wind, or perhaps a falling branch tickling its strings on its way to the ground (natural cause); (ii) intuitively thought that my mother's spirit intentionally caused the wind chimes to move, for no particular reason other than that she was traipsing about outside and wanted to make them move (intentional cause), or (iii) intuitively thought that she caused the wind chimes to move, but that she did so in order to communicate with us: to 'tell' us that her soul had transitioned safely ("It's okay, kids; all is swell on the other side of life!") (declarative cause). Recall that, in the emotional construal of the moment, my intuitive causal ascription fell along the lines of the third variety.

This personal experience with my dead mother's ghost was fodder for the experimentalist in me. And so, naturally, I decided to properly investigate this phenomenon in my laboratory. In a recent study, Becky Parker and I had 3- to 9-year-old children play a simple guessing game in which they were to find the location of a hidden ball inside one of two boxes (Bering & Parker, in press). Sticker prizes served as rewards for all correct guesses. To choose a box, children had only to place their hand on top of that box, and in fact they were given 15-s per trial in case they changed their mind and wished to move their hand. The methodological rub was that half of the children were informed that a friendly invisible princess ("Princess Alice") in the room would help them find the ball by telling them, somehow, when they chose the wrong box. Then, on a counterbalanced 2 of the 4 trials, experimenters in an adjacent room triggered an unexpected event in the laboratory —a table lamp flickering and a picture crashing to the floor —as soon as the child's hand first made contact with one of the two boxes. If children regarded these events as declarative messages from Princess Alice, then compared to the nonevent trials they would be more likely to move their hands and their verbal judgments would reflect this type of causal attribution (e.g., "it happened because I picked the wrong box"). In contrast, if the events were seen as only intentionally caused by Princess Alice, but not as communicative signs, then children shouldn't move their hands in response to them nor explain the events as being about their choice of box. Results showed a significant effect of condition by age group. Only the oldest children (M = 7 years) who were assigned to the Princess Alice condition moved their hands and made verbal judgments reflecting declarative causal reasoning. Younger children (M = 5 years) failed to move their hands in response to the events; instead they saw Princess Alice as a trickster who caused the events only because she wanted to (e.g., "because she likes the picture better on the ground"). Finally, the youngest children (M = 4 years) either shrugged their shoulders or gave good scientific answers (e.g., "because the light's broken").

These age differences are puzzling given that even 2.5-year-olds see deictic gestures such as eye gaze and indexical pointing as referential and declarative. Only follow-up studies will help to disentangle alternative interpretations of these data. At the moment, however, I suspect that second-order theory of mind plays a role in this domain of causal reasoning (e.g., "As she can see from my behavior, Princess Alice knows [I don't know] where the ball is actually hidden; thus, that event is her informing me that I have a false belief").

On 'Being' Dead

But all this still begs the question: "If it is so obvious that mental states are obliterated by death, then why do people the world over believe and behave otherwise?" In much of my own work in this area, I have argued for something called the "simulation constraint hypothesis" to account for people's intuitive reasoning about dead agents' minds (which is of course an oxymoron!). Ever wonder why you can never actually die in a dream sequence? According to the simulation constraint hypothesis, it is because it is impossible to ever know what it is like to be dead; our phenomenological systems are literally forced to construe theoretical models of a subjective existence beyond death. For my doctoral dissertation, I had people reason about the psychological functioning of a protagonist who had just died in a car accident (Bering, 2002). For example, could he taste the breath mint he ate right before he died? Could he

experience lust? Did he know that he was dead? Here is how one young 'materialist' answered this last question: "Yeah, he'd know, because I don't believe in the afterlife. It is non-existent; he sees that now." In an age-modified task in which a puppet mouse was killed by a puppet alligator, children similarly found it hard to disavow themselves of the possibility that the mouse continued to experience psychological states after its death (particularly emotional, desire, and epistemic states) (Bering & Bjorklund, 2004; Bering, Hernández-Blasi, & Bjorklund, in press).

In fact, the younger the child, the more likely he or she was to say that the dead mouse retained various aspects of its consciousness, which is precisely the opposite pattern that one would expect to find if the origins of such beliefs could be traced exclusively to cultural indoctrination. In fact, religious-type answers (e.g. Heaven, God, spirits, etc.) among the youngest children were extraordinarily rare. Recent findings by Paul Harris and Rita Astuti, however, show that the social context can also be highly influential when it comes to children's reasoning about the afterlife. For instance, children are more likely to say that consciousness survives death when the story involves a religious figure (e.g., a priest or medicine man) rather than a doctor (Astuti, 2005; Harris & Giménez, 2005).

Conclusion

In conclusion, investigating folk beliefs about dead agents' minds may provide important information concerning the evolution of human cognition. Since the idea of a soul hinges on the cognitive capacity to represent mental states, studies of peoples' underlying beliefs about the fate of consciousness after death may open an empirical backdoor to this representational system. Reactions to dead bodies, the ability to reason about dead agents' minds, or beliefs concerning the intelligent design of souls are just some of the empirical topics waiting to be further explored.

References

Allen, C., & Hauser, M. D. (1991). Concept attribution in nonhuman animals: Theoretical and methodological problems in ascribing complex mental processes. Philosophy of Science, 58, 221-240.

Astuti, R. (2005). Turning belief into ethnography. Unpublished manuscript.

Bering, J. M. (2002). Intuitive conceptions of dead agents' minds: The natural foundations of afterlife beliefs as phenomenological boundary. Journal of Cognition and Culture, 2, 263-308.

Bering, J. M., & Bjorklund, D. F. (2004). The natural emergence of reasoning about the afterlife as a developmental regularity. Developmental Psychology, 40, 217-233.

Bering, J. M., Hernández-Blasi, C., Bjorklund, D. F. (in press). The development of 'afterlife' beliefs in secularly and religiously schooled children. British Journal of Developmental Psychology.

Bering, J. M., & Johnson, D.D.P. (2005). "O Lord . . . you perceive my thoughts from afar": Recursiveness and the evolution of supernatural agency. Journal of Cognition and Culture, 5, 118-142.

Bering, J. M., McLeod, K. A., & Shackelford, T. K. (in press). Reasoning about dead agents reveals possible adaptive trends. Human Nature.

Bering, J. M., & Parker, B. D. (in press). Children's attributions of intentions to an invisible agent. Developmental Psychologist.

Boyer, P. (2001). Religion explained: The evolutionary origins of religious thought. New York, NY: Basic Books.

Burnham, T., & Hare, B. (2005). Engineering human cooperation: Does involuntary neural activation increase public goods contributions in adult humans? Human Nature.

Dechesne, M., Pyszczynski, T., Arndt, J., Ransom, S., Sheldon, K. M., van Knippenberg, A., & Janssen, J. (2003). Literal and symbolic immortality: The effect of evidence of literal immortality on self-esteem striving in response to mortality salience. Journal of Personality and Social Psychology, 84, 722-737.

Goodall, J. (1990). Through a window. Boston, MA: Houghton Mifflin.

Gullone, E., King, N. J., Tonge, B., Heyne, D., & Ollendick, T. H. (2000). The fear survey schedule in children – II (FSSC-II): Validity data as a treatment outcome measure. Australian Psychologist, 35, 238-243.

Haley, K., & Fessler, D. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior 26:245-256.

Harris, P. L., & Giménez, M. (2005). Children's acceptance of conflicting testimony: The case of death. Journal of Cognition and Culture 5:143-162.

Miller, W. R. & Brigham, R. M. (1990). “Ceremonial” gathering of black-billed magpies (Pica pica) after the sudden death of a conspecific. Murrelet, 69, 78-79.
Reynolds, V., & Tanner, R. (1995). The social ecology of religion. New York: OxfordUniversity Press.
Wilson, E. O. (1971). The insect societies. Cambridge, MA: Belknap Press.
de Unamuno, M. (1926). The tragic sense of life. London: MacMillan and Co., Ltd.
de Waal, F. B. M. (1996). Good natured: The origins of right and wrong in humans and other animals. Cambridge, MA: HarvardUniversityPress.

Discussion

General observations
Robert Stonjek
Oct 3, 2005 12:27 UT

Humans do not, as is often imagined, make up an after death theory to “allay their concerns about non-existence”. In fact, very few people can imagine a state of non-existence after their death. This leads inevitably toward a continuation after death.

What people do fear is not whether or not there is a continuation, but in what form that continuation takes. Stories contrived in this regard can alley fears to such a degree that some may seek death. Early Christians were known to be very ‘accident prone’, no doubt hoping for heaven, such that they were often getting lost in the desert or eaten by hungry beasts.

It is this very point that interested and entertained the Romans who wanted to see for themselves the Christians in this act – well, there was no TV back then so topical events were brought to the arena. The lions were well fed, so the antics of Christians who would try to annoy their lions or place the heads in the lions mouths was near enough to Roman comedy. The story of Jesus being able to ‘survive’ was no doubt true because lions only feed every few days anyway – in between time the lamb can safely sleep with the lion.

In modern times, the lure of 70 virgins has enticed many a suicide bomber. The opposite is also true – the threat of hell is a powerful one.

In **‘signs’**, we see the progression of the ability to associated otherwise unrelated events into a single composite image or model. That this ability is prone to error, such as the ascription of ghosts and spirits to causes of what are probably coincidences, is a tolerable imperfection when the benefits of global modelling are considered.

It is notable that whenever the power fails in the average sized city, numerous people will, at the exact moment the power fails, will have just switched on an electrical appliance, and following that many will feel responsible for the entire blackout.

As for death in dream sequences, this is not uncommon in clinical depression and anxiety. It is a feature in ‘night terrors’. The one caveat that should be noted is that such dreams end both with death and sudden awakeness, not uncommonly in a cold sweat and a reluctance to return to sleep for fear of a recurrence of the nightmare.

I wouldn’t place too much store in studies done on young children and puppet characters. Children have saturation exposure to cartoon characters which are killed in all sorts of ways, only to spring back to life moments later.

One must make a clearer differentiation between the beliefs that a subject has or has developed from their own anticipation of death and possible consequences, and the beliefs etc that one has about an agent's afterlife prospects. Jesse does not make clear to which she refers to in, say, the consideration of stories contrived to relieve death anxiety – is this anxiety about one's own death, or the death of loved ones?

Some clarifying points
Jesse Bering
Oct 4, 2005 11:54 UT

Many thanks to Robert for his very thoughtful response to my paper. I've just a few quick 'rejoinders' to his insightful comments.

(1) Perhaps it was not made entirely clear, but I am in fact in total agreement with you Robert that afterlife beliefs are principally motivated by an inability to imagine a state of non-existence after death. (Although I have one quibble over your line that "...very few people can imagine a state of non-existence after their death" Can anybody?!) In any event, I was simply noting that this commonsense view that the afterlife is an "invention" that serves to "allay death anxiety" is short-sighted (and probably wrong). I certainly wasn't endorsing it! I've actually taken great pains to argue this alternative model of simulation constraints elsewhere. (2) I'm fascinated by the term, but I'm not sure what you mean by "global modelling" in relation to 'signs'? (3) The pretense characters from our afterlife puppet shows are admittedly a potential limitation of these developmental studies. As you might imagine, as methodologists in this area we're constrained by an entire parapet of IRB red tape. However, I'm very sensitive to this issue of the puppets too, and so one of my current PhD students is currently working on replicating these data in a study that uses the death of a real organism (a big fat carpenter ant being squashed). (4) Finally, the confusion is more than understandable given the androgynous name, but (for some reason) I feel it necessary to point out that Jesse is a "he," not a "she".

RE: Clarifying Points
Robert Stonjek
Oct 5, 2005 12:38 UT

Thanks Jesse, On point **One**: we agree that the simple statement that *afterlife stories help people to deal with the nothingness of after death non-existence* is a fiction, but I point out that although we readily accept a continuance, the nature of that continuance is a source of anxiety, particularly if negative outcomes are suggested. A readiness to accept stories and relate revelations recalled from dreams follows. I would still count this as death anxiety, not of the fact of death or the prospect of nothingness but on the unknown nature or mystery of it.

On point **Two**: I'm using the term *Global Modelling* to identify the elements or precursory or contributory concepts that are required to form a *World View*. To make observations about the maturation of cubs into lions, to note their feeding habits, their social life, the way they hunt etc is one thing, but to find a generalizable form is quite another eg that all animals are born, mature, mate, and eventually die.

Whilst I've given an example of Global Modelling that may lead to a world view not dissimilar to the science of biology (esp. ethology), it is likely that early global modelling strongly anthropomorphised observations of nature, say to a story of the God of life and death (for the above example) and this story, along with other Global Modelling in the same vein, will lead to a spiritual world view.

I note that the capacity for Global Modelling is intrinsic in Homo sapiens and when overstimulated through drugs, emotional insult or psychiatric conditions such as schizophrenia, will readily lead to unwanted Global Modelling such as are found in conspiracy theories.

On point **Four**: sorry about the unsolicited sex change – my inner dialogue has been updated to a male Arkansas gentleman modelled on Bill Clinton, who makes even the drollest dialogue in your paper sound riveting.

Kind Regards, Robert Karl Stonjek

'Theory of Mind', dead agents and supernatural abilities
Teresa Bejarano
Oct 5, 2005 11:37 UT

Any human emotions that play a role in reasoning about the mind's fate after death are dependent on a peculiarly human cognitive mechanism -the capacity to think about minds. This dependence is unfolded by Bering in two ways. 1) The fear of ghosts (or dead agents) facilitates the inhibition of selfish behaviours. Bering suspects that second-order ToM plays a role in that domain (I agree. Let's remember that shame and guilt are generally regarded as second-order emotions). This adaptively advantageous fear of ghosts would explain why humans prefer 'declarative causal explanations', why they see 'signs'. 2) Simulation constraint. I want to suggest a third way of dependence on ToM. Dead agents' minds are nearly always envisioned as possessing supernatural abilities. In the afterlife, souls can see future events, and they can have simultaneous perceptual access to very distant places. In addition, after death, alien and own mental states are always equally perceptible. Certainly, we could try to explain these beliefs without suggesting their dependence on ToM. It might be argued that there is an enhanced memory, and an enhanced cultural survival, for concepts that violate core ontological assumptions (cf. Boyer or Barrett). Or we might think of a 'physical explanation'. The disembodied, Platonic, soul would fly and would travel very quickly. I do not refuse these explanations. However, I would suggest a link between ToM and those beliefs. Thinking of supernatural abilities would be a direct consequence of having a ToM. Why? 'Travel time' ('a metamental capacity': Suddendorf) allows humans to think of non-current things: Great!. But -hélas!- in human beings, non-immediate future events are weak representations (cf. neurological studies about immediate and delayed rewards: McClure et al.). This same pattern (ability-cum-limitation) is observed in the perception of alien mental states. Certainly, humans can take the perspective of the other. However, this role-taking is a very brief and weak state. Self and other are very different things. In short, humans are pushed into two opposite views. ToM allows them to represent current and non-current time (and likewise, self and other) as similar to each other. But this presumed similarity cannot be really perceived or exercised. Thus, ToM would allow human beings to see human limitations and barriers, and, consequently, to think of supernatural abilities.

References. Boyer, P. (2000). Mind and culture: religious concepts as a limiting-case. In Carruthers & Chamberlain (eds.) Evolution and the human mind, Cambridge U.P, 93-112. Suddendorf, T. & Busby, J. (2003). Mental time travel in animals? Trends in Cognitive Sciences, 391-396. McClure, S. M. et al. (2004) Separate neural systems value immediate and delayed monetary rewards, Science, 503-507.

Theory of (the deceased) Mind
Robert Stonjek
Oct 6, 2005 11:06 UT

Possible sharing many of the same neural resources at those used for Theory of Mind is our ability to use placeholders for people who are not in our immediate vicinity.

This ability allows one to anticipate another's response to a range of scenarios such as their response to plans we are currently formulating.

Apart from modelling future possible responses to our plans, typical behaviour may also be modelled. How many times do we here the bereaved saying that they almost expect their dead spouse to enter the room, phone, or otherwise enter some form of proximal reality?

It is common for placeholders (taking the place of the real person when they are not physically present) to include aspects of personality, appearance, and some property that makes them seem real rather than mere mannequins of the imagination.

We know that these placeholders persist long after the real person has left us or died. There is no mechanism to 'kill' the placeholder, so in moments of absent mindedness we fail to correct what the imagination so readily animates and we might feel that a phone call is imminent, though the caller has passed away.

Theory of Mind requires just this kind of modelling – we must create an imagined substitute for the real person, who may be in our presence, that is emotionally transparent such that we can see and even experience their pain, joy, surprise or intellectual independence.

That the agent of ToM can persist after the real person is no longer present and therefore can no longer stimulate the ToM agent is a natural extension of that mental ability. The further extension sees its persistence even after the real person has passed away.

This observation makes no judgement about whether or not such after-death persistence *actually occurs*. It is clear, however, that the mechanism by which such after-death modelling occurs is closely related to, and no doubt evolved in step with, Theory of Mind.

Kind Regards, Robert Karl Stonjek

offline social processing
Jesse Bering
Oct 8, 2005 10:01 UT

Robert, we're definitely on the same page here. Here is an excerpt from a forthcoming article of mine:

"...we are wise to remember that social reasoning is not limited to making empirical observations of behaviors, or for that matter even bodies. In fact, human relationships are characterized by mostly offline social events; those with whom we have relationships are only periodically directly observable. Otherwise they continue to exist outside of our perceptual awareness without compromising our strategic ability to represent and monitor, through indirect sources, their ongoing presence in the social environment (e.g. Dunbar 1993 2004). When it comes to death, human cognition apparently is not well equipped to update the list of players in our complex social rosters by accommodating the recent nonexistence of any one of them. This is especially the case, of course, for individuals who have played primary roles in our social lives, who did so for a long time, and who were never presumed to be continuously stationary when they were out of our sight. Because human cognition is designed for offline as well as online social processing, it "expects" the periodic physical absence of social partners. Casual observation reveals that individuals will often, for instance, pick up the phone with the intention of calling the decedent or fleetingly imagine how the decedent will react when told about some good news, only to remember that the person is dead. Although these automatic cognitions are probably the residue of habitual social behaviors, they also reveal something about the challenges faced by the human cognitive system when it attempts to process information concerning the truth about dead agents' physical whereabouts. In his famous novel The Magus, John Fowles writes of his protagonist's difficulty in this specific area when mulling over the recent death of his lover: "I forced myself to stop thinking of her as someone still somewhere, if only in memory, still obscurely alive, breathing, doing, moving, but as a shovelful of ashes; as a broken link, a biological dead end, an eternal withdrawal from reality." In summary, a person who has recently died and whose body has already been disposed of may continue to be processed by an offline social system for an undetermined period of time, especially if compounded by nonnegotiable simulation constraints."

Offline Social Processing
Robert Stonjek
Oct 8, 2005 10:58 UT

It is particularly pertinent to consider what the 'offline' representation is composed of. Apart from imagery, which can fade or change with time, the most tangible substance of, for instance, one's internal representation of their mother, is a set of emotions, attitudes, generic and possibly genetically based predispositions and feelings that typically accompany mirrored states such as the feeling one has upon seeing a mother nurturing a young baby.

'Mother', in my example, *must* have some innate component specific to each species, from the duck's 'first moving object I see after birth' to our general predisposition to seek out 'mother' from the proximal environment after our birth and the more sophistication elaborations that develop later.

From a general stock of emotion, an association between particular emotional states is woven together with the image of an individual to form the internal representation. When that person dies, 'love' does not also die. Nor does any other component of that representation cease to exist.

The combination of components that make up that internal representation is the unique component of each individual so represented, and there is no reason for that to die or otherwise cease to exist along with the target of that representation.

Indeed – what part of the internal representation *should* change and how? One speculates that a particular representation should become benign or no longer active, but this may actually be a complicated process involving the selective removal of particular components including expectations and so on. An ability to globally reset the intentional states of an internal representations may be a yet-to-evolve elaboration available only to some much later human form.

The simulation constraint hypothesis
Anne Reboul
Oct 7, 2005 15:18 UT

I like the simulation constraint hypothesis. I just would like to point out that it is in fact strongly linked to the fact that being dead is like being non-existent. The idea is that we can't know what it is like to be dead because we can't know what it is like to be non-existent. Or, in other words, we can't know what it is like to be dead, because there is nothing it is like to be non-existent and thus we can't experience non-existence and, hence, death being one kind of non-existence, we can't experience being dead. If, indeed, this is the explanation, then one would expect non-existence before conception to be just as much a problem as is death. Indeed, this can be seen in young children of around three to four years of age. When you look at photographs of them as infants, they typically ask where they were before they were born. If you answer that they were in their mother's belly, given the relative sophistication of children of our time regarding gestation and birth, they ask where they were before they were in their mother's belly. When you answer that they were nowhere, before they did not exist, the answer does not make any kind of sense to them because they can't make sense either of coming to exist or of ceasing to exist. One interesting question is why this preoccupation with non-existence before conception disappears in the next few years (it's a safe bet that though 6 years old worry about death, they don't worry about non-existence before conception)

pre-existence studies
Jesse Bering
Oct 8, 2005 10:15 UT

Anne, absolutely. If the simulation constraint hypothesis has any veracity, then intuitive representations of "life before birth" should be similar to those of "life after death." I also suspect that this is where the intelligent design stance may articulate with an innate dualism. Getting at this experimentally with children is tough, though. Last year I ran a pilot study where children were asked to reason about the biological and psychological functioning of a single human being at three

distinct temporal periods: prior to conception (e.g., "before he was in his mother's belly"), during prenatal development (e.g., "while he was in his mother's belly"), and after birth. Many of the younger children had trouble even grasping the "pre-existence" category (which sounds revealing, but in fact I suspect it had more to do with language limitations than simulation constraints). I'm working on it though!

The future and the past
Teresa Bejarano
Oct 13, 2005 11:09 UT

Why the preoccupation with non-existence before conception disappears in 6-year-olds? Why is the future- the expectation of the future- more important than the past? All animal goal-driven behaviour can be explained in terms of expectations –either innate or learned -. Certainly, human beings have autonoetic episodic memory. However, we do not allocate affective force in past contents except when we want -or fear- some elements of these contents to return. (Perhaps the adaptively advantageous function of human tears is related to this. Tears would consume the affective force that focuses on an impossible desire, and that otherwise prevented us from continuing our activity). Thus, it makes sense the observed contrast -the absence of preoccupation with non-existence before conception, and the often anxious reasoning about the mind’s fate after death.

Causal Inferences. Evolutionary Domains and Neural Systems

Clark Barrett (Assistant Professor, UCLA Department of Anthropology)
Pascal Boyer
(Date of publication: 15 October 2005)

Abstract: Causal perception is a domain in which cognitive neuroscience studies and developmental psychology findings should be confronted. We still have no good description of the neural correlates of such simple causal thinking as connecting two events because of temporal contiguity. In particular, this causal binding of events could be operated either by specialised, modality specific low-level systems or higher-level contingency detection, or both. It would be useful to examine how developmental results illuminate this question.

Here we consider two apparently distinct questions: [1] What are the neural correlates of causal inference? And [2] How do we distinguish between different domains of causal inferences? To understand the varieties of causal thinking in human minds, we need to bring together behavioral and developmental data on the one hand and information from both neuro-psychology and neuro-imaging on the other. Once this evidence is replaced in an evolutionary framework, it becomes easier to understand the functional divisions between neural systems. We discuss these questions in the context, first, of high-level conceptual differences between living-things and artifacts, and then of low-level causal perception.

Computation, neural systems and evolution

In some domains, the mapping between computational function and neural structure seems rather straightforward. Consider face-recognition. There is substantial behavioral evidence to support the notion of a specialized computational device (Young et al., 1987; Tanaka & Sengco, 1997). Some functional features (e.g. configural processing, sensitivity to inversion) are found only in the treatment of face-like stimuli (see (Kanwisher, 2000)for a detailed discussion). There is also developmental evidence in infancy (Pascalis et al., 1995; Slater & Quinn, 2001). Neuro-psychology has documented many cases of prosopagnosia (Farah, 1994)in which the structural processing of objects, object-recognition and even imagination for faces can be preserved while face-recognition is impaired (Duchaine, 2000; Michelon & Biederman, 2003).

In this case, functional specialization happens to map onto to neural structure. Neuro-imaging studies have reliably shown a specific modulation of activity of the fusiform gyrus of the temporal lobe during identification or passive viewing of faces (Kanwisher et al., 1997). Specialized systems handle the invariant properties of faces (that allow recognition) while other networks handle changing aspects such as gaze, smile and emotional expression (Haxby et al., 2002).

The example of face-recognition is exceptional, in the sense that there is a clear mapping here between a unique and clearly defined function on the one hand and a unique structure on the other. But the example is relevant here in that it shows how evolutionary considerations are the necessary background to functional distinctions. Let us consider this in more detail.

Despite the impressive behavioral and neural evidence, some psychologists have argued that the specificity of face-perception was an illusion, and that human beings simply became expert recognizers of faces by using unspecialized visual capacities. One central argument is that one can observe inversion effects (Diamond & Carey, 1986; Gauthier et al., 1998)and fusiform gyrus activation (Gauthier et al., 1999; Tarr & Gauthier, 2000)when trained experts examine and identify automobiles, birds, dogs or even abstract geometrical shapes (the effects occur only in their particular domain of expertise). So, the argument goes (we simplify a bit), the system is not actually dedicated to faces, but to a broader domain of (say) 'visual stimuli of personal importance with similar overall structure and fine-grained distinctive features'.

There are three possible objections to this alternative interpretation, two of which are tempting but

misguided:

[1] One may object that 'face' is a simpler concept that 'visual stimuli of … etc.' so we should prefer the first predicate, even if the cognitive system considered is often rather lax in its interpretation of what a 'face' is. However, this is misguided because there really is nothing intrinsically more parsimonious in one concept than the other. The complexity of predicates depends on a matrix of other predicates they are related to, so that 'visual stimuli of … etc.' may be the simpler predicate in some alternative ontology (Goodman, 1955).

[2] One may think that 'face' is a more natural predicate, because there are such things as faces in the world (people have faces, animals have faces), whereas the domain of 'visual stimuli of… etc.' does not to correspond to a proper natural kind. This too is misguided, because as it is difficult to argue that faces really are a natural kind of objects. Perhaps faces as such do not enter into any causal laws except in the computational states of face-recognizers… which leads us to the third argument:

[3] Being good at (or slightly better than others at) distinguishing conspecifics can have important fitness effects. Indeed, various human behaviors of great evolutionary significance (social exchange, friendship, cooperation, warfare, coalitional affiliation) depend on the precise identification of conspecifics. To the extent that the top front part of conspecifics (what we usually call "the face") is used as a source of information for that purpose, 'faces' are part of the cognitive environment of human beings in a way that 'visual stimuli of … etc.' are not. Some species of primates evolved to become better recognizers of 'faces', not better recognizers of 'visual stimuli of … etc.' because the only domain where their performance mattered to fitness was the distinction between conspecifics.

This may seem rather obvious – because the functional specialization of face-recognition is not in fact really controversial. We emphasize the point, because such evolutionary considerations become even more important as we turn to domains where functional distinctions are less obvious.

Domains: Living things / artifacts-materials

Let us turn to the high-level distinction between the domains of 'living things' and 'artifacts', generally interpreted as a functional distinction based on specific causal principles. Animal species are intuitively construed in terms of species-specific "causal essences" (Atran, 1998). By contrast, man-made objects are principally construed in terms of their functions. (Kemler Nelson, 1995)(Richards et al., 1989). Artifacts seem to be construed by adults in terms of their designers' intentions as well as actual use (Bloom, 1996)and pre-schoolers too consider intentions as relevant to an artifact's 'genuine' function (Gelman & Bloom, 2000).

These are differences of inferential principles. The fact that an object is identified as either living or man-made leads to [a] paying attention to different aspects of the object; [b] producing different inferences from similar input; [c] producing categories with different internal structures (observable features index possession of an essence [animals] or presence of a human intention [artifacts]; [d] assembling the categories themselves in different ways (there is no hierarchical, nested taxonomy for artifacts, only juxtaposed kind-concepts).

The neuro-psychological evidence seems to supports this notion of distinct structures. Some types of brain damage result in impaired content or retrieval of linguistic and conceptual information in either one of the two domains. The first cases to appear in the clinical literature showed selective impairment of the living thing domain, in particular knowledge for the names, shapes orassociative features of animals (Warrington & McCarthy, 1983; Sartori et al., 1993; Sheridan & Humphreys, 1993; Sartori et al., 1994; Moss & Tyler, 2000). But there is also evidence for double dissociation, for the symmetrical impairment in the artifact domain with preserved knowledge of living things (Warrington & McCarthy, 1987; Sacchett & Humphreys, 1992).
However, it is very difficult in this case to map the functional distinctions and their behavioral manifestations onto neural structures or systems. In particular, the neuro-imaging evidence is less than altogether compelling. A host of neuro-imaging studies, using both PET and fMRI scans, with either word- or image-recognition or generation, have shown significantly different cortical activations for living things

and artifacts (Martin et al., 1994; Perani et al., 1995; Spitzer et al., 1995; Martin et al., 1996; Spitzer et al., 1998; Moore & Price, 1999; Gerlach et al., 2000). However, few of these findings are clearly replicated. Also, in many studies the variety of activation peaks reported for either type of stimuli could not plausibly be described as constituting a functional network. That is, there is no clear indication that joint involvement of such areas is required for the processing of such stimuli. The gross anatomy does not suggest particular and exclusive connectivity between those regions either. Finally, some of the findings may turn out to be false positives (Devlin et al., 2002).

Reinterpretation from an evolutionary viewpoint

The problem in this case is that the domains themselves are not construed in a principled way. In most studies of domain-specificity, the precise understanding of what are 'artifacts' (often oddly called 'objects') or 'animals' or 'living things' is left to… the experimenter's commonsense, as if that was a privileged road to cognitive structure. In some other cases experimenters stick to scientific distinctions between the 'living' and 'non-living', somehow implying that the organization of mental faculties corresponds to the way the world really is.

From an evolutionary viewpoint, we should expect cognitive domains to correspond to recurrent fitness-related situations or problems (e.g. 'predators', 'competitors', 'tools', 'foraging techniques', 'mate selection', 'social exchange', 'interactions with kin', etc.). This suggests that we should find systems specialized, not so much in different kinds of objects, but in different kinds of interaction with objects, likely to impinge on the organism's fitness.

From that standpoint, humans certainly do not interact with "living things" in general. Living things comprise plants, bacteria, and middle-sized animals including human beings. We interact in very different ways with predators, prey, potential foodstuffs, competitors, and parasites. Nor do humans handle "artifacts" in general. Man-made objects include foodstuffs, tools and weapons, buildings and shelters, visual representations, as well as paths, dams and other modifications of the natural environment. We should expect the input format and activation cues of domain-specific inference systems to reflect this fine-grained specificity.

This hypothesis of a set of finer-grained, fitness-relevant systems receives some support from the neuro-imaging evidence. For instance, the naming of artifacts, or even simple viewing of pictures of artifacts, seems to result in pre-motor activation. Viewing an artifact-like object automatically triggers the search for (and simulation of) motor plans that involve the object in question. Indeed, the areas activated (pre-motor cortex, anterior cingulate, orbito-frontal) are all consistent with this interpretation of a motor plan that is both activated and inhibited. This suggests that "man-made object" is probably not the right criterion here. Houses are man-made but do not afford motor plans that include handling. If motor plans are triggered, they are about tools rather than man-made objects in general (Moore & Price, 1999). A direct confirmation can be found in a study of manipulable versus non-manipulable artifacts, which finds the classical left ventral frontal (pre-motor) activation only for the former kind of stimuli (Mecklinger et al., 2002). By contrast, some infero-temporal areas (BA20) are found to be exclusively activated by animal pictures (Perani et al., 1995), as are some occipital areas (left medial occipital) (Martin et al., 1996). The latter activation would only suggest higher modulation of early visual processing for animals. This is consistent with the notion that identification of different animal species requires finer-grained distinctions than that of artifacts: animals of different species (cat, dog) often share a basic Bauplan (trunk, legs, head, fur) and differ in details (shape of head, limbs, etc.), while tools (e.g. screwdriver, hammer) differ in overall structure. Animal-specific activations of the posterior temporal lobe seem to vanish when the stimuli are easier to identify (Moore & Price, 1999) hich would confirm this interpretation as an effect of fine-grained, relatively effortful processing[i].

Neuro-imaging findings and developmental evidence converge in supporting the evolutionarily plausible view, that inference systems are not about ontological categories like "man-made object" or "living thing" but about types of situations, such as "fast identification of potential predator-prey" or "detection of possible use of tool or weapon".

Neural evidence for causal perception

Let us now turn to causal perception, the fast and automatic interpretation of distinct visual events as causally related. The cognitive approach to causal perception stems from Hume – and from Michotte's study of causal inferences for the trajectories of billiard-ball-like shapes (Michotte, 1946). The interpretation of such visual events as causal or non-causal depends on a subtle psycho-physics that combines time- or space-contiguity between events with relative velocities (Schlottman & Shanks, 1992)from infancy (Leslie, 1984; Schlottmann & Surian, 1999). To the extent that there is principled interpretation of particular stimuli, what specific neural structures are involved?

Surprisingly, this was not the object of much research until recently. One major reason may be that there seem to be no cases of selective impairment of causal reasoning. So there is no neuro-psychological evidence and no tentative connections to localized infarction or other damage. This may suggest that no specific structure is involved in causal perception, that the process is distributed among many different systems. Or it may suggest that the process in question is so fundamental to other (higher-level) domains of cognition that an impairment in this capacity would result in general confusion and therefore not to a diagnosis of selective impairment.

Be is as it may, there is very little neuro-imaging investigation of the networks modulated by causal perception. One exception is an event-related fMRI study by Blakemore and colleagues, with stimuli including various psycho-physical variants of causal and non-causal collisions (Blakemore et al., 2001). Results of the causal – noncausal subtraction show higher activation MT/V5, STS (superior temporal sulcus) and the left IPS (intra-parietal sulcus). Finding activation in MT/V5 is not in itself surprising (the region is sensitive to relative motion) but the difference in activation may suggest that a causal event is more complex or triggers more processing than a non-causal event. The other two regions are generally involved in the higher-level interpretation of visual events. Taken together, these activations suggest that causal events are treated as special from the lowest stages of visual processing. But note that the system involved is not a "Hume module" for causal inference in general but rather a detector of specific visual psycho-physics.

Pure causal processing or joint activation?

The search for a "Hume module" may be a wild goose chase because the system's putative domain of operation is unconstrained. Do organisms really need to process causal events as such and distinguish them as a general class from non-causal events? This would be self-evident if brains had been designed to be philosophically correct, as it were. But they were designed to enhance fitness.

From this standpoint, the mere detection of contingencies in one modality may be of great interest to the extent that it can feed into specialized downstream processing. That is to say, the quick perception of two events as contiguous (which may or may not involve cortical structures) would be quickly followed by activation of possibly relevant systems specialized in fitness-relevant situations. For instance, it seems plausible that the detection of potential predators occurs as a fast response to specific kinds of unexpected stimuli (e.g. sudden noise without visual correlate). Another kind of detection is involved in the way organisms monitor the effect of their own actions on objects. Specific parameters of the contingency relations between events may trigger activation of such a detection system.

All this is necessarily rather speculative, but would at least suggest that it is not the contingency detectors themselves that produce causal inferences, but the joint activation of contingency-detection and situation-relevant inference engines. This would have interesting consequences. For instance, it would suggest that causal inferencing in the brain rarely is a matter of single-modality information-processing, but usually involves comparison of several modalities. Although there is evidence that cross-modality contingency detection is a fast and automatic system (Driver & Spence, 1998; Schmitt et al., 2000), there is no study of such effects in specific, biologically relevant situations.

Let us mention a last example that supports this notion of a joint activation of contingency detection and higher-level templates for causal inference. Blakemore and colleagues compared the neural activities triggered by both mechanical and "intentional" causal events between shapes on a screen (Blakemore et

al., 2003). In the intentional condition one shape oriented in such a way as to “attend” to another shape, a much simpler intentional event than the familiar Heider films of triangles and squares chasing and helping each other (Heider & Simmel, 1944). Subjects either attended to the direction of motion or to the contingency relations. Results showed some interesting differences between mechanical and intentional events. But more interesting for our discussion here was the difference within the intentional condition, depending on the subject’s focus of attention. When subjects focused their attention on the contingent relationships between the objects in the displays, as opposed to physical aspects of the objects’ movement, there was significant activation of the right middle frontal cortex and the left STS, a subset of the regions that are consistently activated by theory of mind tasks (Fletcher et al., 1995; Brunet et al., 2000; Castelli et al., 2000; Gallagher et al., 2000; Vogeley et al., 2001). Although this activation had been previously described as an automatic result of “intentional-looking” contingencies (Happé et al., 1999), it seems more accurate to describe it as the effect of contingencies in the context of intentional expectations (Blakemore et al., 2003).In other words, given a biologically relevant situation (looking for possible animate agents in an unknown environment), the detection of contingencies involves both lower-level and specific higher-level neural systems. Rather than a Hume system, this is a “look-for-animate-agents” system.

Conclusions

There are obvious limits to this picture, notably because neuro-psychological cases are ambiguous and neuro-imaging techniques are limited. But the evidence reviewed here should point to a general lesson.

Evolutionary considerations are not a complement or footnote to the study of cognitive function, a sort of “by the way, this is how the system got to be this way…” commentary on prior results, independently produced by neuro-psychology or neuro-science or experimental psychology. A description of the “proper domains” of function is intrinsically an evolutionary description (Sperber, 1996). In the near future, we may expect to understand much better how the brain handles information about causal relations. The price for this better picture may be to jettison some philosophical baggage, such as the notion that organisms detect “causation” in general. Contrary to this “philosophical” way of proceeding, we should start from the computational requirements of specific fitness-relevant situations, predict which kinds of contingencies are pertinent to each specific type, and test for the existence and autonomy of corresponding neural systems. The examples reviewed here suggest that this may be a more successful strategy in the description of causal thinking.

References

Atran, S. A. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. Behavioral & Brain Sciences, 21(4), 547-609.
Blakemore, S.-J., Boyer, P., Pachot-Clouard, M., Meltzoff, A. N., & Decety, J. (2003). Detection of contingency and animacy in the human brain. Cerebral Cortex, 13, 837-844.
Blakemore, S.-J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., et al. (2001). How the brain perceives causality: An event-related fmri study. Neuroreport: For Rapid Communication of Neuroscience Research, 12(17), 3741-3746.
Bloom, P. (1996). Intention, history and artifact concepts. Cognition, 60, 1-29.
Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., & Moss. (2002). Is there an anatomical basis for category- specificity? Semantic memory studies in pet and fmri. Neuropsychologia, 40(1), 54-75.
Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. Journal of experimental psychology General, 115(2), 107-117.
Driver, J., & Spence, C. (1998). Crossmodal attention. Current Opinion in Neurobiology, 8(2), 245-253.
Duchaine, B. C. (2000). Developmental prosopagnosia with normal configural processing. Neuroreport: For Rapid Communication of Neuroscience Research, 11(1), 79-83.
Farah, M. (1994). Specialization within visual object recognition: Clues from prosopagnosia and alexia. In G. R. Martha J. Farah (Ed.), The neuropsychology of high-level vision: Collected tutorial essays. Carnegie mellon symposia on cognition. (pp. 133-146): Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.
Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. Nature Neuroscience, 2(6),

568-573.
Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. Vision Research Special Issue: Models of recognition, 38(15-16), 2401-2428.
Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. Cognition, 76(2), 91-103.
Gerlach, C., Law, I., Gade, A., & Paulson, O. B. (2000). Categorization and category effects in normal object recognition: A pet study. Neuropsychologia, 38(13), 1693-1703.
Goodman, N. (1955). Fact, fiction and forecast. Cambridge: Harvard University Press.
Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. Biological psychiatry, 51(1), 59-67.
Heider, F., & Simmel, M. (1944). An experimental study of apparent behaviour. The American Journal of Psychology, 57, 243-259.
Kanwisher, N. (2000). Domain specificity in face perception. Nature Neuroscience, 3(8), 759-763.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17(11), 4302-4311.
Kemler Nelson, D. G. (1995). Principle-based inferences in young children's categorization. Revisiting the impact of function on the naming of artefacts. Cognitive Development, 10, 347-380.
Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. Perception, 13(3), 287-305.
Martin, A., Haxby, J. V., Lalonde, F. J., Wiggs, C. L., Parasuraman, R., & Ungerleider, L. G. (1994). A distributed cortical network for object knowledge. Society for Neuroscience Abstracts, 20(1-2), 5.
Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. Nature (London), 379(6566), 649-652.
Mcklinger, A., Gruenewald, C., Besson, M., Magnie, M.-N., & Von Cramon, Y. (2002). Separable neuronal circuitries for manipulable and non-manipulable objects in working memory. Cerebral cortex, 12, 1115-1123.
Michelon, P., & Biederman, I. (2003). Less impairment in face imagery than face perception in early prosopagnosia. Neuropsychologia, 41(4), 421-441.
Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. Brain, 122(5), 943-962.
Moss, H. E., & Tyler, L. K. (2000). A progressive category-specific semantic deficit for non-living things. Neuropsychologia, 38(1), 60-82.
Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & et al. (1995). Mother's face recognition by neonates: A replication and an extension. Infant Behavior & Development, 18(1), 79-85.
Perani, D., Cappa, S. F., Bettinardi, V., Bressi, S., Gorno-Tempini, M., Matarrese, M., et al. (1995). Different neural systems for the recognition of animals and man-made tools. Society for Neuroscience Abstracts, 21(1-3), 1498.
Richards, D. D., Goldfarb, J., Richards, A. L., & Hassen, P. (1989). The role of the functionality rule in the categorization of well-defined concepts. Journal of Experimental Child Psychology, 47, 97-115.
Sacchett, C., & Humphreys, G. W. (1992). Calling a squirrel a squirrel but a canoe a wigwam: A category specific deficit for artefactual objects and body parts. Cognitive Neuropsychology, 9, 73-86.
Sartori, G., Coltheart, M., Miozzo, M., & Job, R. (1994). Category specificity and informational specificity in neuropsychological impairment of semantic memory. In C. Umilta & M. Moscovitch (Eds.), Attention and performance 15: Conscious and nonconscious information processing. (pp. 537-550). Cambridge, MA, US: The MIT Press.
Sartori, G., Job, R., Miozzo, M., Zago, S., & et al. (1993). Category-specific form-knowledge deficit in a patient with herpes simplex virus encephalitis. Journal of Clinical & Experimental Neuropsychology, 15(2), 280-299.
Schlottman, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 44(2), 321-342.
Schlottmann, A., & Surian, L. (1999). Do 9-month-olds perceive causation-at-a-distance? Perception, 28(9), 1105-1113.
Schmitt, M., Postma, A., & De Haan, E. (2000). Interactions between exogenous auditory and visual spatial attention. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53A(1), 105-130.

Sheridan, J., & Humphreys, G. W. (1993). A verbal-semantic category-specific recognition impairment. Cognitive Neuropsychology, 10(2), 143-184.
Slater, A., & Quinn, P. C. (2001). Face recognition in the newborn infant. Infant & Child Development Special Issue: Face Processing in Infancy and Early Childhood, 10(1-2), 21-24.
Sperber, D. (1996). Explaining culture: A naturalistic approach. Oxford: Blackwell.
Spitzer, M., Kischka, U., Gueckel, F., Bellemann, M. E., Kammer, T., Seyyedi, S., et al. (1998). Functional magnetic resonance imaging of category-specific cortical activation: Evidence for semantic maps. Cognitive Brain Research, 6(4), 309-319.
Spitzer, M., Kwong, K. K., Kennedy, W., & Rosen, B. R. (1995). Category-specific brain activation in fmri during picture naming. Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience, 6(16), 2109-2112.
Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. Memory & Cognition, 25(5), 583-592.
Tarr, M. J., & Gauthier, I. (2000). Ffa: A flexible fusiform area for subordinate-level visual processing automatized by expertise. Nature Neuroscience, 3(8), 764-769.
Warrington, E. K., & McCarthy, R. (1983). Category-specific access dysphasia. Brain, 106, 859-878.
Warrington, E. K., & McCarthy, R. (1987). Categories of knowledge: Further fractionations and an attempted integration. Brain, 110, 1273-1296.
Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. Perception, 16(6), 747-759.

Discussion

Evolutionary Steps toward Logical Deduction of Causative Agents
Robert Stonjek
Oct 16, 2005 23:28 UT

In its most basic form the evolution question asks “what is the selective advantage to a species of discovering the causal agent of particular phenomena?”

Survival is also about the ability of an animal to anticipate and avoid danger. Successful response strategies may evolve without the need for each animal to discover the best strategy on its own, such as the three distinct calls of the Vervet monkey which is a partially genetic and partially culturally derived strategy.

A further development is the recollection of previous successful responses to danger eg how a predator was evaded.

Requiring considerably more brainpower is the calculation of causation. If causation can be deduced from the evidence gathered at the scene of an event, say a conspecific has died from an attack of a predator then experience can be gained vicariously for a very wide range of phenomena – far more than genetic evolution could ever provide.

Thus we would expect to see in any species capable of deducing causation a hierarchy of approaches to avoiding danger – the preadapted reflexive response, the learned (from experience or cultural transmission) response, and finally the deduction from available evidence. A further elaboration, possibly only available to humans, is the generalisation from deduced causation.

Chimps, approaching a banana suspended overhead and boxes randomly strewn about will typically try a direct approach until losing their temper and throwing a tantrum. Once done, they more carefully consider the problem before stacking the boxes one atop of the other so the boxes can be scaled and the banana acquired. But on successive occasions the chimp will go through the entire routine again, indicating that

the chimp is capable of deductive reasoning but not of generalising that reasoning, even to all similar situations.

Humans *are* capable of such generalisation, and erroneous generalisations go a long way to explaining the development of religious doctrine. But this is far from a general detection of causation as a matter of course. Humans still default to the reflexive and experiential forms before moving on to attempt deduction of causative agents. Isn't this why the Sherlock Holmes novels proved so enticing – that an individual could side step his own 'intuitive' reasoning, expose others false assumptions (largely based on their reflexive or intuitive reasoning and their experience with vaguely similar events or phenomena) and move directly on to deductive reasoning in the search for the causative agent?

Yes, humans can do it – they can parse sensory data with some objectivity to logically deduce the causative agent, but this is not the default approach. Evolutionary models can help to illustrate the underlying or earlier steps required to respond to danger, find food and to mate. We should be aware that deduction of causation was appended to these when neural resources were sufficient to accommodate it, and that early evolutionary steps are still the steps taken by default before logical deduction of causation can proceed.

Robert Karl Stonjek

Contingency detection plus domain-specific inferences
Gloria Origgi
Oct 17, 2005 17:10 UT

This is a naïve question of clarification. The authors argue that rather than looking for a module of causal detection, detection of contingencies should be investigated in order to understand its role in constraining downstream inferences. Roughly, the system detects a contingency relation in a specific environment and matches it with a set of inferences that lead to relevant conclusions in that environment. For example, the detection of contingency relations between two moving geometric shapes, like two triangles, may trigger an intentional reading of the movement, that is, an activation of a "look-for-animate-agents" system. So, the idea is that we detect a contingency relation in a specific context and this triggers specific inferences. But don't we need those specific inferential patterns to detect the contingency relations? In the case of the moving shapes, what is the difference between detecting "intentional-looking" contingencies, as Happé has defined them, and detecting contingencies in the context of intentional expectations? What would be the advantage of a two-step explanation of a contingency-detection that triggers a domain-specific inference? And also, what is precisely 'causal' in the detection of the co-presence of two or more stimuli in the environment, if their only effect is that they trigger an set of domain-specific inferences? How does this explain our expectations about causality, that is, our expectations about the *connection* of events and not simply their *correlation*?

> **Reply to Origgi**
> Clark Barrett
> Oct 27, 2005 17:27 UT
>
> There are at least two questions here. The first is whether there is a single module responsible for all causal perception and causal inferences, or a "Hume module." There are reasons to suspect this might be unlikely, if only because humans and other animals discriminate between different types of cause (e.g., mechanical, intentional), and therefore, one might expect multiple systems, each with distinct activation or discrimination criteria. A qecond question is whether causal perception and inference are a single process, or whether there are multiple processes or mechanisms involved. We suggest that there may in fact be at least two steps in any given case of causal inferece: joint activation of contingency detection and higher-level templates for causal inference. Origgi asks "don't we need those specific inferential patterns to detect the contingency relations?" It's possible; as we stressed, this is all a matter of speculation, partly because of lack of relevant research, and partly because it might be quite difficult in practice to tease apart distinct processes if, indeed, multiple processes are involved, especially if they are tightly coupled as we speculate. Moreover, as

perception researchers like to point out, perception is a kind of inference. Nevertheless, we speculate that perceptual mechanisms might exist that require only specific stimulus arrangements to trigger them, but that themselves are just akin to “categorization” mechanisms. These would not necessarily require principles of causal inference in order to carry out their job. For example, some pretty basic spatiotemporal cues might distinguish between biological and purely mechanical motion (e.g., contingency at a distance in the former case, and upon contact in the latter case). Inferences, such as that event A “caused” event B, might not be generated by the mechanism that identifies such events themselves, but by inference systems that take the output of these perceptual “detectors” as inputs. One might argue that it would just be splitting hairs to argue that two systems were involved, as well as difficult to test. We pointed to some research (e.g., Blakemore) that suggests that the perceptual and inferential elements in the process might to some degree be separable. The only advantage to such an explanation, would be that it would separate the computational steps, if indeed there were evidence for multiple steps. Finally, Origgi asks: “what is precisely ‘causal’ in the detection of the co-presence of two or more stimuli in the environment, if their only effect is that they trigger an set of domain-specific inferences? How does this explain our expectations about causality, that is, our expectations about the connection of events and not simply their correlation?” Here, we would agree, and this is precisely one of the point we are making in suggesting that causal perception and inference may be, to some degree, distinct: there is, as Origgi says, nothing “causal” in mere detection of contingencies. It is in the inference that they license that one observes true causal cognition, rather than just detection of patterns or correlations.

If we have an evolutionary
Giyoo Hatano & Kayoko Inagaki
Oct 24, 2005 14:06 UT

Although we fully agree to Barrett and Boyer that both bringing together behavioral data and information from cognitive neuroscience and considering evolutionary perspectives seriously are needed for the advancement of the understanding of human causal cognition, we, as researchers on children's naive biology, must give a few skeptical comments on their characterizations of the living-nonliving distinction.

First let us discuss naive biology including the living-nonliving distinction at the information-processing level. As presented much more in detail elsewhere (Inagaki & Hatano, 2002) we adopt Hirschfeld and Gelman's (1994) definition of domains, that is, "a body of knowledge that identifies and interprets a class of phenomena assumed to share certain properties and to be of a distinct and general type" (p. 21) and regard naive biology as a representative of them. This means that naive-biological knowledge (1) is constructed based on individual experiences; (2) using powerful learning mechanisms of conceptual nature; (3) helped by innate constraints; and (4) by interactive sociocultural constraints. Therefore, we believe, taking seriously not only the evolutionary perspective but also how innate constraints are instantiated through individual experiences in sociocultural contexts is indispensable to understand the development of naive biology. The situation will not change even if we divide the domain of naive biology into fitness-related sub-domains or modules. Innate constraints must be skeletal because, since the process of evolution is very slow, possessing specific pieces of innate knowledge may be detrimental when ecological environments change. For example, what could be "recurrent fitness-related situations" of predator identification for human ancestors? These situations had to be defined in highly abstract ways in order to be useful for millions of years in a great variety of ecological niches.

How about the two authors' claims at the neural level? We believe the same arguments as the above can be applied. We pointed out (p. 136) that neuroimaging experiments with normal participants for animals versus artifacts "show segregation between categories, but the specific areas involved differ across studies" (Caramazza & Shelton, 1998, p. 23), probably because different members of animal and tool categories were presented in different forms for different tasks. In addition, such terms as living things and artifacts are too inclusive and misleading. However, even when we find dedicated neural mechanisms for animals in general or specifically for potential predators, they are likely to reflect individual experiences as well as genetic endowments, considering the plasticity of the human brain. Which regions are activated by a given cognitive task may considerably vary depending upon participants' experiences.

If the authors succeed in finding more straightforward patterns of mapping between behavioral manifestations and neural structures by focusing on a set of finer-grained, fitness-relevant knowledge systems, that would be wonderful. However, this business is not enough for understanding how such systems emerge and develop.

Reference Inagaki, K. and Hatano, G. (2002) Young children's naive thinking about the biological world. Philadelphia, PA: Psychology Press

Reply to Hatano and Inagaki part 1
Clark Barrett
Oct 27, 2005 8:30 UT
Reply to Hatano and Inagaki

We agree with the general point that the developmental environment, including both social factors and other aspects of individual experience, can be quite important in the development of the phenotype. However, we wish to stress that much more can be said about the development of domain-specific cognition than just that the environment impacts development. We would go beyond this to suggest that the developmental systems that build domain-specific cognitive architectures use environmental and social information in specific, principled ways, that can vary depending on the domain. A good example would be Mineka et al.'s work (Mineka et al., 1984) on social acquisition of snake fear in rhesus monkeys, which shows that the system in question has quite specific design features for the integration of social cues with category-specific information about the objects in question (in this case, the category of the object, as well as the relevant social cues, matter in the acquisition of fear). This example suggests that "innate principles" and "sociocultural contexts" cannot be cleanly separated, as the innate principles themselves are designed to use contextual cues. Moreover, the influence of our evolutionary past is felt in the kinds of socio-cultural environments we build, which are highly species-specific. Hatano and Inagaki suggest that evolved dispositions should be rather abstract because of the variety of environments in which selective pressures have been manifest, and they give the example of predator-prey interaction. We would note two things here: first, it is ultimately an empirical question just how "skeletal" domain-specific dispositions are, and second, it is highly likely that this is something that varies from domain to domain. It is not a general feature of domain-specific systems that they are "skeletal" or "abstract." Within the domain of living things we may find, for example that there are quite specific detectors or templates for certain classes of dangerous animals, such as snakes, because snakes were present in a wide range of human environments, and there is a stable set of cues (e.g., shape, motion) that can be used to identify them. Richard Coss and colleagues (Coss et al., 1993) have shown that some squirrels, for example, possess a highly specified snake template that is not dependent on experience in order to function: squirrels raised with no experience of snakes still exhibit evasive behaviors towards snakes, but not other objects. Interestingly, Coss and colleagues have been able to document about how long it takes for this "snake template" to disappear in populations that have existed without snakes: something on the order of 10,000 years. Dan Blumstein and colleagues (Blumstein et al., 2000) have showed a similar phenomenon in Tammar Wallabies living on islands off the coast of Australia, where there have been no predators for thousands of years. Members of this species who have never seen a predator before (and whose ancestors have not experienced predators for millenia) still show fear and anti-predator behaviors when presented with stuffed models. Thus, it is quite possible for highly domain-specific mechanisms to exist that are not skeletal at all, and that do not require fleshing out by experience or social context. In the case of true predators on humans

Reply to Hatano and Inagaki part 2
Clark Barrett
Oct 27, 2005 8:33 UT

In the case of true predators on humans (snakes are dangerous to humans but do not usually prey on them, except in the case of some constrictors), Hatano and Inagaki point out, correctly, that the class of such predators was quite diverse. While the evolutionary persistence and stability of snakes across human environments suggests that a fairly well-specified snake template is quite plausible, it

is perhaps less likely (though not impossible) that there would exist specific templates for crocodiles, bears, lions, and so on (though felids were present across such a wide range of ancestral environments that a cat template is not that implausible; Barrett, 2005). Nevertheless, there are cues that can be used to reliably discriminate predators from other kinds of animals, at least on a statistical basis. Several of these have been discussed in the literature, from sharp teeth, to forward facing eyes, to large body size, to aspects of biological motion, including pursuit / evasion motion (see Barrett, 2005, for a review). To take motion as an example, the ability to discriminate pursuit and evasion from other kinds of motion on the basis of low-level cues has been shown across cultures and in childhood (Barrett et al., 2005), and infants have been shown to have specific expectations about the behavior of objects that exhibit pursuit / evasion cues (Csibra et al., 2003). Thus, it is not the case that there exist only skeletal principles general to the entire domain of animate things. Rather, these findings suggest specific, early-developing principles specific to the much more specific domain of pursuit and evasion, or predation. These considerations suggest that some aspects of the stalking or other situations may be extremely precise, so it is not at all implausible – especially given the huge importance to fitness of being able to avoid predators – that there may be correspondingly precise evolved architecture for identifying such situations, and the same may be true of other domains (social exchange, kin interactions, and so on). This is not to say that learning, and the use of social information, would not be important. Indeed, as Cheney and Seyfarth (1990) have shown, even the highly stereotyped, domain-specific antipredator alarm call behavior of vervets is tuned during development, such that the identification of relevant predators becomes more precise. Although there is little research on it, it seems certain that human children use social cues from adults and other humans in developing abilities to recognize and categorize dangerous and non-dangerous animals, but do so in principled ways. One would expect social learning as a design feature of danger avoidance systems, since knowledge about dangers can be transmitted socially, and learning about danger through direct experience can be costly (Barrett, 2005). This makes sense of results such as those of Mineka et al. (1984), showing principled sensitivity to social cues in danger learning. In summary, we find it quite plausible that evolution would result in highly precise dispositions. Others, such as Hatano and Inagaki, may find this less plausible, but the evidence from other species shows that there can be quite precise anti-predator adaptations, and we see no reason why this could not be the case in humans, and in domains other than predation. Again, it is ultimately an empirical matter, and little research has been done. However, in the studies that have looked at, for example, fear of and reactions towards dangerous animals such as spiders and snakes, the mechanisms involved appear to be quite specific rather than general (see Öhman and Mineka, 2001, for a review).

Reply to Hatano and Inagaki part 3 and bibliography
Clark Barrett
Oct 27, 2005 8:35 UT

We would also stress, again, that there is no general answer to the question of how skeletal domain-specific architecture will be; it will probably vary from domain to domain. There may be highly specialised structure in some domains and not others. We have the same remark about neural systems. Why should neural organisation reflect abstract organisation (e.g. ‘living things’, ‘artefacts’) rather than more specific domains (‘predators’, ‘contaminants’, ‘tools’, ‘representations’ etc.)? If there are micro-domains that have been particularly important to fitness, then highly domain-specific structures can be built.

Barrett, H. C. (2005). Adaptations to predators and prey. In D. M. Buss (Ed.). The handbook of evolutionary psychology. (pp. 200-223). New York: Wiley. Barrett, H.C., Todd, P.M., Miller, G.F., and Blythe, P. (2005). Accurate judgments of intention from motion alone: A cross-cultural study. Evolution and Human Behavior, 26, 313-331. Blumstein, D. T., Daniel, J. C., Griffin, A. S. & Evans, C. S. 2000. Insular tammar wallabies (Macropus eugenii) respond to visual but not acoustic cues from predators. Behavioral Ecology, 11, 528-535. Cheney, D., & Seyfarth, R. (1990). How monkeys see the world: Inside the mind of another species. Chicago: U. of Chicago Press. Coss, R.G., K.L. Guse, N.S. Poran, and D.G. Smith. 1993. Development of antisnake defenses in California ground squirrels (Spermophilus beecheyi): II. Microevolutionary effects of relaxed selection from rattlesnakes. Behaviour 124:137-164. Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-

year-old infants use teleological representations of actions productively. Cognitive Psychology, 27, 111-133. Mineka, S., Davidson, M., Cook, M., & Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. Journal of Abnormal Psychology, 93, 355-372. Öhman, A., & Mineka, S. (2001). Fear, phobias and preparedness: Toward an evolved module of fear and fear learning. Psychological Review, 108, 483-522.

Associative Learning in Animals and Humans

Leyre Castro (Postdoctoral Researcher; Department of Psychology, University of Iowa) and
Edward A. Wasserman (Professor of Psychology, University of Iowa)
(Date of publication: 14 November 2005)

Abstract: No one would dispute that humans are able to learn causal relationships. But do animals also possess this capacity? Both humans and animals have been subjected to similar evolutionary histories and they currently experience common survival demands: to predict and control the environment. Close correspondence between animal and human cognition might therefore not be at all farfetched. We propose that associative learning theories provide an applicable and promising viewpoint from which to understand many of the phenomena observed in animal conditioning and human causality learning.

Pavlovian Conditioning and Human Causal Learning

A strong undercurrent in thinking since Hume is that humans do not directly apprehend causality. Instead, we make causal inferences based on a restricted set of experiences. When (1) two events occur together in time and space, (2) one of the events precedes the other, and (3) the two events appear consistently together (that is, they do not occur alone), we normally infer the existence of a causal relationship between them (Hume, 1739/1964).

Human causal learning is affected by these primary Humean rules, which are the same factors that affect classical conditioning in animals: contiguity, priority, and contingency (e.g., Fales & Wasserman, 1992; Shanks & Dickinson, 1987). Moreover, both humans and animals exhibit behavioral phenomena such as "discounting" and "augmentation" (Kelley, 1973), which appear to implicate a sophisticated causal reasoning process; organisms not only take into account how a potential cause covaries with the effect, but also how this cause competes with rival explanations of the effect. Interestingly, Hume himself amplified his three primary rules with three others that better pinpoint causality: (4) the same cause always produces the same effect, and the same effect never arises but from the same cause, (5) where several different objects produce the same effect, it must be by means of some common feature, and (6) any difference in the effects of two resembling objects must proceed from that particular in which they differ.

One of the best-known cases of discounting is the cue validity effect, first reported by Wagner, Logan, Haberlandt, and Price (1968). In their experiments, target Cue X was equally often paired with the outcome in all experimental conditions; Cue X was paired half of the time with Cue A and half of the time with Cue B. In one condition, each AX trial was paired with the occurrence of the outcome and each BX trial was paired with the absence of the outcome; in the other condition, both AX and BX were assigned the same probability of occurrence of the outcome (0.50). Rats', rabbits', and pigeons' conditioned responding, and humans' causal judgments to Cue X systematically decrease as the differential predictiveness of AX and BX increases—discounting (Wasserman, 1990).

Associative learning theories such as that of Rescorla and Wagner (1972) can readily explain these results. Briefly, the Rescorla-Wagner model states that a reinforcer can sustain only a limited amount of associative strength; so, simultaneously presented cues must compete with one another as the best predictor—or cause—of the outcome. When AX is consistently followed by the outcome and BX is not, Cue A becomes a strong predictor to the detriment of Cue X. When both AX and BX are followed by the outcome half of the time and are not followed by the outcome half of the time, none of the cues can become a strong predictor, so Cue X can acquire moderate positive strength.

As promising as this associative account approach may be, problems have arisen when this model was applied to other types of discounting and augmentation effects, such as those involving absent cues: so-called "retrospective revaluation" phenomena (also addressed by Martin Giurfa in his contribution to this conference), which have been observed in both Pavlovian conditioning (e.g., Kaufman & Bolles, 1981; Miller & Matute, 1996) and human causal learning studies (e.g., Dickinson & Burke, 1996; Wasserman & Berglan, 1998). These phenomena involve the presentation of a compound of two cues, AB, that is

followed by the outcome, so that each of these individual cues will become a moderate predictor or cause of the outcome. After this training, Cue A is presented alone, either followed by the outcome or not followed by the outcome, with no further training of Cue B. Even though Cue B is not given, its associative value changes. When Cue A alone is paired with the outcome, subjects decrease their judgments of the causal strength of Cue B—backward blocking. On the other hand, when Cue A alone is not paired with the outcome, subjects now increase their judgments of the causal strength of Cue B—recovery from overshadowing.

One way to deal with these challenges for associative learning theory is to dismiss the theory outright as either incorrect or inappropriate to causal understanding. Another tactic is to reconsider some of the premises of associative learning theory and to modify it in light of the evidence. For instance, it seems reasonable to believe that, if the presented cue were to gain strength in light of evidence, then nonpresented cues might immediately and correspondingly lose strength; conversely, if the presented cue were to lose strength in light of evidence, then nonpresented cues might immediately and correspondingly gain strength. Van Hamme and Wasserman (1994) suggested that the Rescorla-Wagner (1972) model could be modified in such a way that different learning rate parameters are assigned to presented and nonpresented cues: positive and negative learning parameters, respectively. This theoretical maneuver allows the Rescorla-Wagner model to embrace results that at first had appeared to be so discomforting.

Associative learning theory has accordingly been modified and enriched by the similarities between human causal learning and animal conditioning. We propose that the existence of such parallels speaks to a common underlying process. If one assumes that, during a conditioning procedure, animals acquire information about the causal texture of their environment, then the correspondence between animal conditioning and human causal learning can be readily accepted. However, some deem these parallels to be inadequate to prove causal understanding in animals, because these studies concern merely "making predictions about the temporal and spatial relations between observable events" (see, for example, discussion on Jennifer Vonk's paper in this conference). We disagree on this point, precisely because—as the studies mentioned above show—causal understanding even in humans seems to be based on the observation of temporal and spatial regularities in the environment.

As well, we are concerned with drawing a strong theoretical distinction between making predictions and making causal inferences. We would suggest that the best predictor of an event is also the cause of that event. It is unlikely that environmental contingencies are organized in such a way that a non-cause would be a reliable predictor of an event. It would be peculiar if evolution were to have endowed organisms with the ability, not to detect causal relationships, but to detect predictive relationships, when the former ought to be more directly relevant to survival than the later.

What evolution may have done is to prepare organisms to preferentially forge some associative connections, thereby increasing the speed with which certain experiential contingencies promote learning. Garcia and Koelling's (1966) classical experiments showed that learning depends on the relevance of the potential cause to the potential effect. Rats readily associated a taste with later illness, but with much greater difficulty they associated audiovisual cues with the same illness. In the opposite fashion, rats readily associated audiovisual cues with shock, but they found it difficult to associate taste with shock. Hence, learning best proceeded when the potential cause was combined with a relevant effect. Similar results have been found in other species, such as pigeons (Shapiro, Jacobs, & LoLordo, 1980). It might be that, when a plausible causal link can be inferred—even when the underlying connection is not truly causal—organisms are better prepared to associate stimuli that are presented together.

Instrumental Learning and Causation

Causal knowledge allows us not only to predict, but also to control our environment. We are able to predict an effect on the basis of observed cues, but we are also able to predict the effects that our own actions will have on the environment. If animals understand that there is a causal relationship between events, then one might argue that, when the effect is highly valuable, the animal should work to make the cause occur. Instrumental conditioning relies on the ability of organisms to learn that their own actions can produce certain outcomes. Humans' and animals' manipulation and control of their environment may be based on the inference of a causal relationship between their own behavior and the consequences of

this behavior.

Man's first experience with causes probably came from his own behavior: things moved because he moved them (Skinner, 1971, p.7).

It is not difficult to imagine that all mobile organisms go through this very basic experience. Hence, it is reasonable to ask: Are nonhuman animals also able to distinguish between events that are caused by their own behavior from those that are not? Killeen (1981) "asked" his pigeons whether or not they were responsible for key light offset. The pigeons were able to discriminate whether it was their own behavior or "something else" that caused the change in the light. The rudiments of causal understanding can easily be note here.

Arguably more compelling evidence of causal understanding in instrumental conditioning comes from studies of outcome devalution. Adams and Dickinson (1981) trained rats to press a lever to get food pellets. Later, an aversion to the food was induced by injecting the animals with a mild toxin that produced gastric illness. During this aversive conditioning, the lever was not present. The relevant issue here was to what extent this devaluation would affect lever pressing when the lever was again available. If the animals had learned that there was a positive causal relation between lever pressing and the receipt of food pellets, then lever pressing should be influenced by this causal knowledge and the current desirability of the outcome. Because the food pellets were no longer appetizing, the animals decreased their pressing of the lever.

Current challenges for associative theories

Several more recent findings pose new challenges for the adequacy of associative learning theory to explain causal understanding. Let's examine a recent study about the difference between observation and intervention in nonhuman animals.

Even when the same between-event contingencies are arranged, people make different causal inferences depending on whether they merely observe the occurrence of an effect or they know that someone or something else has intervened to produce that effect (e.g., Waldmann & Hagmayer, 2005). Blaisdell, Kosuke, and Waldmann (2005) were interested in whether or not rats could also exhibit a similar tendency. These researchers presented rats with a light followed by a tone and the same light followed by sucrose. The light should be a potential cause and the tone and the sucrose should be potential effects. Would the animals consider the light as the cause of both the tone and the sucrose?

To answer this question, after the above training, a lever was inserted into the chamber. In the Intervention group, the tone was presented each time the rats pressed the lever, whereas in the Observation group, the tone's presentation was not contingent on lever pressing, although the tone was presented the same number of times in each group. Therefore, one of the effects of the light, the tone, was only "observed" in the Observation group, whereas it was "intervened" in the Intervention group. If rats were to consider the light to be a common cause of the tone and the sucrose, then the presentation of the tone in the Observation group should lead the animals to infer that the light must have occurred and to expect sucrose as well. On the other hand, if rats in the Intervention group attributed the tone's presentation to their own lever pressing behavior, then they should not attribute the tone to the presence of the light, because it had been caused by their own behavior. Thus, rats in the Intervention group should not infer that the light had occurred as well, so that they should not expect any other effects of the light—specifically, the sucrose—to be present. This is exactly what Blaisdell et al. (2005) observed: when the tone appeared, rats in the Intervention group did not look for sucrose, whereas rats in the Observation group did.

Thus, it seems that nonhumans' behavior can evidence complex causal reasoning not based on the mere extraction of contingency information about given events. Here, rats may show that they are not only capable of forward learning, from cause to effect, but that they are also able to observe effects and to diagnose whether the cause should have occurred or not. Indeed, in the Blaisdell et al. (2005) study, we might be seeing hints of diagnostic abilities that are embedded in the notion of causal explanation (in the terms considered by Anne Reboul in this conference). The rats' behavior suggests that these diagnostic

capacities may not be uniquely human; indeed, these abilities may emerge from bidirectional associations that have been the focus of learning theorists for over 100 years. Numerous studies show that, during the course of learning, humans and animals do not acquire just forward associations, but also backward—or bidirectional—associations between paired events (Arcediano, Escobar, & Miller, 2005; Asch & Ebenholtz, 1962; Frank & Wasserman, 2005). These bidirectional associations might help to explain the diagnostic abilities that both human and animals exhibit.

Conclusions

Parallels between Pavlovian conditioning and human causal judgment, research on instrumental conditioning, and recent work on the distinction between observed and intervened effects, all suggest that causal knowledge lies at the root of both human and animal behavior. We do not deny that humans' causal understanding is far more advanced than animals'; but, that advancement is likely to be premised on the basic rules of causal association that were proposed centuries ago by David Hume. Whether that advancement is simply a further elaboration of these rudimentary rules or something qualitatively different is a live empirical question.

References

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. Quarterly Journal of Experimental Psychology, 33B, 109–122.

Arcediano, F., Escobar, M., & Miller, R. R. (2005). Bidirectional associations in humans and rats. Journal of Experimental Psychology: Animal Behavior Processes, 31,301-318.

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. Proceedings of the American Philosophical Society, 106, 135-163.

Blaisdell, A. P., Kosuke, S., & Waldmann, M. (2005). Seeing versus doing: Two modes of assessing causal models by rats. Proceedings of the 12th Annual International Conference On Comparative Cognition

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. Quarterly Journal of Experimental Psychology, 49B, 60-80.

Fales, E., & Wasserman, E. A. (1992). Causal knowledge: What can psychology teach philosophers? Journal of Mind and Behavior, 13, 1-27.

Frank, A. J., & Wasserman, E. A. (2005). Associative symmetry in the pigeon after successive matching-to-sample training. Journal of the experimental Analysis of Behavior, 84, 147-165.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. Psychonomic Science, 4, 123-124.

Hume, D. (1964). Treatise of human nature (edited by L. A. Selby-Bigge). London: Oxford University Press. (Original work published 1739)

Kaufman, M. A., & Bolles, R. C. (1981). A nonassociative aspect of overshadowing. Bulletin of the Psychonomic Society, 18, 318-320.

Kelley, H. H. (1973). The processes of causal attribution. American Psychologist, 28, 107-128.

Killeen, P. R. (1981). Learning as causal inference. In M. L. Commons & J. A. Nevins (Eds.), Quantitative analyses of behavior (Vol.1): Discriminative properties of reinforcement schedules (pp. 89-112). Cambridge, MA: Ballinger.

Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. Journal of Experimental Psychology: General, 125, 370-386.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appleton-Century-Crofts.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), The psychology of learning and motivation, (Vol. 21, pp. 229-261). San Diego, CA: Academic Press.

Shapiro, K. L., Jacobs, W. J., & LoLordo, V. M. (1980). Stimulus-reinforcer interactions in Pavlovian conditioning of pigeons: Implications for selective associations. Animal Learning and Behavior, 8, 586-594.

Skinner, B. F. (1971). Beyond Freedom and Dignity. New York: Knopf.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. Learning & Motivation, 25, 127-151.
Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. Journal of Experimental Psychology, 76, 171-180.
Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two models of accessing causal knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 216-227.
Wasserman, E. A. (1990). Attribution of causality to common and distinctive elements of compound stimuli. Psychological Science, 1, 298-302.
Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: The role of within-compound associations. Quarterly Journal of Experimental Psychology, 51B, 121-138.

Discussion

Humean and human causal cognition
Anne Reboul
Nov 16, 2005 14:15 UT

Thanks to Leyre and Edward for a very interesting contribution. Let me begin with what I think we uncontroversially agree about: I (being a staunch admirer of Hume) have no doubt that a big part of human causal cognition is Humean association. I also think that one can go a great way with Humean association, and I fully believe that humans are in no way special in the animal world as far as Humean association goes. Finally, I have no doubt that Humean association is one basis of both human and nonhuman causal cognition. Thus, I entirely agree with both Leyre and Edward on all these points and I suspect most people would. There are two points on which we may not agree, the first one not touched in Leyre and Edward's contribution, the second one which they discuss. The first one is the possibility that human causal cognition, in addition to Humean association, is informed by specific domain knowledge in, e.g., naive physics, naive psychology, naive biology. Supposing that only humans are able of diagnostic causal knowledge (the next point), it may be thought that part of the input for such diagnostic causal inferences comes from such domain specific knowledge, leaving aside to what extent it is, or not, innately specified. The second question, very well discussed in Leyre and Edward's paper is to what extent animals may be thought to be able of retrospective causal inference and to what extent such inferences are tied to associative mechanisms. In other words, is association bidirectional? As Leyre and Edward point out, there are some well-described associative mechanisms which seem to entail a sort of retrospective cognition, for instance retrospective revaluation. They give a very nice and detailed description and indeed I used retrospective revaluation in a non-published talk to illustrate how similar such associative phenomena are to human causal cognition as it is expressed through language (some causal constructions seem to miror such associative mechanisms). The question is, is it what I meant by diagnostic causal cognition? In retrospective revaluation, two putative causes are presented simultaneously, followed by an effect (to put it in fully causal terms). Then, only one of them is presented, either followed or not by the effect. This changes the predictive value of the other non-presented cue. Is it retrospective causal cognition? I would tend to think that it is retrospective learning, i.e. that present data will change a previous association. However,though I don't doubt that it is a sophisticated process (it has somehow to rely on a A-notB type of reasoning, because the second cue is not explicitly given), I'm not sure I would call it a diagnostic inference (not a claim made by Leyre and Edward, by the way). Another indication of retrospective learning given in the paper is the fact that "learning depends on the relevance of the potential cause to the potential effect", giving as example food aversion. However, I'm slightly wary of food aversion which seems a misfit among associative mechanisms in that the effect is not necessarily contiguous in time with the cause (more than two hours can elapse between food ingestion and sickness, whereas other associative mechanisms only work in a very short window of time) and in that once established the association seems impossible to undo. Thus, food aversion might well be a very special kind of associative mechanism, whose adaptive significance is pretty obvious and that might well mean that food is very limited in the types of causal association it can go into.

Domain-specific knowledge
Leyre Castro
Nov 23, 2005 19:36 UT

Anne Reboul suggests the possibility of domain-specific knowledge influencing human causality judgment. We concur that this might be the case. Such knowledge could be expressed in a greater facility to form certain associations than others. Our point was that animal behavior may also be influenced by domain-specific knowledge. This is exactly what Garcia and Koelling's (1966) experiments intended to illustrate. When the potential cause was combined with a relevant effect, learning proceeded particularly quickly. Anne suggests that taste aversion learning might reflect a special kind of associative mechanism. One might instead propose that taste aversion learning more generally shows that organisms are better prepared to associate stimuli that are presented together when a plausible causal link can be inferred between them.

(Just to clarify a fine point, we do not consider stimulus relevance or belongingness studies to be examples of either retrospective or diagnostic learning.)

Blaisdell et al.'s experiment: a few questions
Anne Reboul
Nov 16, 2005 14:16 UT

Finally regarding the Blaisdell et al. experiment, I'm not quite clear why it should show anything more than retrospective learning, whereby rats in the Intervention group would consider their intervention as a more relevant potential cause than light. A potential misunderstanding here is that I'm not clear as to whether in the second part of the experiment the light was on all the time (for both groups) and coupled with tone in a random fashion in the observation group, but coupled with lever pressing in the intervention group. Neither am I clear as to whether sucrose was given in the second phase of the experiment. I'd like further details on that experiment.

A few more details about the Blaisdell et al. procedure
Aaron Blaisdell
Nov 16, 2005 17:20 UT

In the Blaisdell et al. procedure, all rats received the same training which consisted of trials on which the light was forward paired with the tone (light-->tone) and trials on which the light was forward paired with sucrose. Both trial types were interspersed within each training session, and all event durations were 10s.

In the testing phase that followed, levers were extended into the conditioning chamber, and rats were allocated to one of two test conditions. In the Observation condition, rats were merely presented with the tone (matched to the number of presentations in the Intervention group). (Note: lever presses had no consequence in this condition.) In the Intervention condition, the tone was presented each time the rat happened to press the lever. Neither the light nor sucrose were ever presented during testing, and the rats were not trained to press the lever in any manner. The levers were merely made available. Fortunately for the purpose of the study, all of the rats did press the lever of their own accord at least a few times during the test phase.

Measures of the amount of nose poking (i.e., anticipation of sucrose) into the feeding niche during the tone at test revealed more nose poking in the Observation condition than in the Intervention condition.

Blasidell et al.'s procedure
Aaron Blaisdell
Nov 17, 2005 3:28 UT

I tried posting this before, but it never appeared on the website, so I'm posting again.

The procedure used by Blaisdell et al. is as follows:

In the training phase all rats receive the same treatment which consists of trials with a 10-s light forward paired with a 10-s tone interspersed among other trials with the 10-s light forward paired with 10-s of sucrose. We used parameters to get sensory preconditioning and not conditioned inhibition (i.e., the tone and sucrose should both be treated was effects of the light if the rats are constructing a causal model relating these events.)

At test, levers are extended into the cage and rats are allocated to one of two test conditions. In the Observe condition, rats receive presentations of the tone. In this condition lever presses are recorded but have no stimulus consequence. In the Intervene condition, the tone is presented every time the rat presses the lever. Our dependent measure is nose poking into the feeding niche which provides a measure of the rat's expectation that sucrose will be delivered. We saw higher rates of nose poking in the Observe condition than in the Intervene condition. It is important to note that the rats did not receive any training on the lever. The first time the rats ever encountered the lever was at test, and we merely hoped that they would press the lever so that the tone would come on. Our fears were alleviated by the fact that all of the rats pressed the lever at least a few times during testing.

Thus, when the tone—one effect of the light—was presented alone, the rats acted as if they diagnostically predicted the light and its other effect—sucrose. When the tone was presented only upon a lever press, the rats acted as if they did not expect sucrose, presumably because the lever press intervention produced the "graph surgery" effect Pearl describes.

Delusion as an abnormal causal reasoning process. A search for a common ground in schizophrenia and dementia in older people

Sebastien Carnicella (Lecturer, University Institute of Technology, Strasbourg)
Philippe Oberling (Associate Professor of Neurophysiology, Faculté de Médecine de Strasbourg)
(Date of publication: 8 December 2005)

Abstract: Delusion is an abnormal causal reasoning, whose process seems rooted in the inability to compute probability estimates (contingency judgment). We will analyse the psychological and neurobiological substrates of such a defect in both human and non-human species in order to provide an unitary framework of delusion disorders.

Psychotic states with their classic symptoms of delusions, hallucinations, illogical thinking and formal thought disorders, encompass a broad range of psychiatric illness -the prototype being schizophrenia-, but also various neurological disorders and toxic-metabolic disturbances. Psychiatric symptoms are common in Alzheimer's disease (AD), with psychosis -as evidenced by hallucinations and delusions- present in approximately 50% of affected patients. This propensity toward psychosis coupled with the prevalence of AD in the population renders it second only to schizophrenia as a source of psychotic states. Various clinical and biological features are shared by both schizophrenic and demented older patients such as AD. Without entering into an extensive review of the similarities and the differences between both diseases (see e.g., White and Cummings, 1996), the following common characteristics should be noticed. As previously mentioned, positive psychotic symptoms are common in AD. Delusions occur more frequently than hallucinations. Alike in schizophrenia, the delusions of dementia tend to be unsystematized and loosely held, with the majority being of a simple persecutory type. Importantly, delusions in AD are not correlated with the severity of dementia and are an independent manifestation of brain dysfunction (Flynn et al., 1991). Contrastingly, visual hallucinations are more frequent than auditory ones in AD, the reverse pattern being observed in schizophrenia. Neuroimaging studies have evidenced a lateral/third-ventricular enlargement associated with hippocampus and temporal lobes atrophies in both diseases, along with a reduced glucose metabolism in frontal lobes. Roughly speaking, neurochemical studies have evidenced a disturbed dopamine/acetylcholine balance in both diseases, which seems to result from an overactivity of the central dopaminergic systems in schizophrenia, and from an under activity of the presynaptic cholinergic systems in AD. Finally, psychotic symptoms in both diseases are treated using mainly dopamine-blocking agents, even if the therapeutic efficacy of these agents in AD is modest compared with that seen in schizophrenia. Surprisingly with regard to the original description of schizophrenias as a group of Dementia Praecox (Bleuler, 1911/1966), such clinical analogies between schizophrenic patients and demented older ones have attracted only a limited amount of researches. Since several years, our group was interested by these similarities, but focusing mainly on delusion disorders (Oberling & Carnicella, in press). Delusion is defined as a false belief based on incorrect inferences and interpretation about external reality (DSM III-R). Despite their semantic contents (persecutory, grandiosity, bizarre, etc), delusion disorders seem to be characterized by the spreading activation and maintenance of erroneous causal relationships between events. More precisely, deluded patients exhibit a particular and somewhere specific pattern of hasty decision making, the so-called 'jump to conclusion' effect (Garety & Hemsley, 1997). With regard to non-deluded (non-deluded schizophrenic, depressed, obsessive-compulsive or anxious) patients, deluded patients need fewer information to reach a decision, a finding which is reliably observed throughout studies. This pattern of response does not seem to rely on affective bias, such as a poor motivation to engage into the tasks or the urge to finish them as quickly as possible (for a review, see e.g. Garety & Freeman, 1999). On one hand, deluded patients have a strong tendency to bind both together information that simply co-occurs randomly. This leads to spontaneously generated abnormal causal relationships (Chapman and Chapman, 1973), and is referred as the 'data-gathering bias' (Hemsley & Garety, 1997; Garety & Freeman, 1999; but see Kaney & Bentall, 1989). On the other hand, when they are facing reasoning tasks, deluded patients tend to proceed normally, except that they require fewer information while doing more errors, but with a level of conviction in the correctness of their choice being higher than the one of normal subjects (Linney et al., 1998). It should be clear from the abovementioned data that the psychological features of deluded disorders are not yet fully captured. It is thus tempting to try having a further insight into the psychological and biological mechanisms of delusion disorders. This is the purpose of the present article.

Using an associative word recall task, Nestor and his colleagues (1998) evidenced that schizophrenia was characterized by faulty modulation of associative links, but within an apparently spared lexicon. In their task, both the connectivity (associative strength) and the network size (number of associates) of the words varied in such a way that the list contained equal proportions of four types of words: 1/ high connectivity-small network size (HC-SN), 2/ low connectivity-small network size (LC-SN), 3/ high connectivity-large network size (HC-LN), and 4) low connectivity-large network size (LC-LN). The schizophrenic patients showed a particularly pronounced effect of the connectivity of the to-be-remembered words. Regardless of network size, recall improved substantially for words of high connectivity and declined dramatically for words of low connectivity. By contrast, the normal comparison subjects showed the usual best recall for words of HC-SN, followed by words of LC-SN, then by words of HC-LN, and finally by words of LC-LN. Connectionist simulation evidenced that modeling the schizophrenic response pattern led to aberrant activation and network instability that could account for thought intrusion and delusional thinking (Han et al., 2003). Whether the schizophrenic pattern is primarily driven by the dominance of strong associations over weak ones, or the occurrence of competitive distracting relationships that might interfere, remains to be determined. In line with previous studies using associative tasks (e.g. Chapman and Chapman, 1973; O'Carrol, 1995; Nestor et al., 1998), Han et al. (2003) hypothesized that the schizophrenic response pattern might be primarily driven by the relative absence of weak associations.

Using tasks that allow the independent assessment of memory for item and for associative information, Naveh-Benjamin (2000) evidenced in older adults, considerable difficulties in binding together unrelated components of an episode into a cohesive entity. Older adults are particularly deficient in tasks that require the binding of information to contextual elements, that is, background information that can disambiguate the meaning of a target event (Baddeley, 1982). More precisely, older adults show context activation/updating impairments, context maintenance starting to be altered in the oldest adults (age > 75 years), a phenomenon that is further exacerbated in adults suffering from early stage of AD (Braver et al., 2001; 2005). Deficit in the use of contextual information processing constitutes a striking analogy between dementia in older adults and schizophrenia. Bleuler (1911/1966) first observed the intrusion of dominant but contextually inappropriate associations in schizophrenic thought. Direct experimental evidence came from the rigorous work of Chapman and Chapman (1973) who demonstrated a pronounced schizophrenic bias for dominant meanings of homonyms (e.g., "pen" as a writing instrument) even when preceding sentential context called for secondary meanings (e.g., "pen" as an enclosed fence). In line with several studies (e.g. Rizzo et al., 1996; Oberling et al., 1999; Bazin et al., 2000), Cohen and his colleagues (Cohen & Servan-Schreiber, 1992; Cohen et al., 1999) provided compelling evidence that context updating and maintenance was altered in schizophrenic patients in a way similar to what was observed in older adults and AD (Braver et al., 2005).

Based on the abovementioned observations and keeping in mind that not every schizophrenic or older demented patients are deluded, we propose the following hypothesis to account for delusion disorders in these pathologies: The core neuropsychological alteration in both schizophrenia and dementia relies on a dysfunction of the associative processes, patients experiencing difficulties to bind both together information. The deficit of binding concerns primarily weak associations such as contextual ones (Cohen & Servan-Schreiber, 1992; Oberling et al., 1999; Braver et al., 2005). Thus, the bulk of information coming from patients' past, present and future cannot be organized into a coherent entity. This results to an incoherent internal speech which can lead to disorganized external behaviors (verbalization, action, etc.). The associative deficit appears suddenly in schizophrenia at a post-pubertal stage, when the frontal cortices are supposed to connect to other structures. It appears progressively in older adults when the brain start to engage into neurodegenerative mechanisms. At the neuro-anatomical level, such a disconnection induces a global hypo-activity of the frontal cortex which affects its most rostral part, namely the prefrontal cortex, a region which is involved in many cognitive functions such as reasoning, planning, or executive functions (Ramnani & Owen, 2004). From a neurochemical point of view, schizophrenia is characterized by an overactivity of the mesolimbique dopaminergic system, whereas dementia is characterized by a loss of the forebrain cholinergic neurons. These facts lead in both cases to a disturbed dopamine/acetylcholine (DA/Ach) balance, or more precisely stated, to an altered DA/Ach ratio. In both cases (schizophrenia and dementia) the DA/Ach ratio is higher than the one in normal subjects. To the extent that the DA/Ach ratio follows a simple Weber's law, increasing the level of DA or decreasing the level of Ach, should produce similar outcomes. Simply put, the central dopaminergic system is involved (among other functions) in the detection of newly formed associations (for a review, see e.g. Young et al.,

2005), whereas the forebrain cholinergic system supports attentional and mnesic processes (Everitt & Robbins, 1997). On this ground, increasing the DA/Ach ratio leads to the formation of new and strong, albeit irrelevant, associations, either by increasing the dopaminergic flow or by decreasing the cholinergic one. Associative processes must be impaired enough to generate abnormal behaviors that lead to the clinical diagnostic of schizophrenia or dementia. When those processes are less severely impaired, individuals are clinically classified as schizotypal-prone or pre-demented. This emphasizes that the associative processes can be intact (normal young and older individuals), mildly impaired (schizotypal traits and pre-dementia) or severely impaired (schizophrenia and dementia), therefore suggesting a progressive gradient of deterioration as the diseases are exacerbated. If one is ready to accept the associative deficit as the core neuropsychological process in both schizophrenia and dementia, it remains to consider that not all schizophrenic and older demented patients suffer from delusion disorders. We here suggest that the demented (praecox or late-onset) pattern results from an initial lost a weak associations (such as evidenced by the contextual deficit observed in these pathologies), followed latter on, as the diseases are more salient from a clinical point of view, by a net dominance of strong associations over weak ones (Escobar et al., 2002), leading thus to disorganized internal speech and external behaviors. In this framework, delusions would result from the intrusion of competitive distracting associations that will interfere with the primarily established and aberrant strong associations. We acknowledged that our hypothesis constitutes a simple, not to say a simplistic, view of delusion disorders. Nevertheless, it provides the tremendous advantage of being experimentally testable.

References

American Psychiatric Association (1987): Diagnostic and statistical manual of mental disorders, 3rd rev. A.P.A.: Washington DC.
Baddeley AD (1982) Domains of recollection. Psychol Rev 89:708–29.
Bazin N, Hardy-Bayle MC, Perruchet P, Feline A (2000) Context dependent information processing in patients with schizophrenia. Schizophr Res 45:93–101
Bleuler E (1966): Dementia praecox or the group of schizophrenias. International Universities Press: New York. (originally published in 1911).
Braver TS, Barch DM, Keys BA et al. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. J Exp Psychol Gen 130: 746–63.
Braver TS, Satpute AJ, Rush BK et al (2005). Context processing and context maintenance in healthy aging and early stage dementia of the Alzheimer's type. Psychol Aging 20: 33-46.
Chapman LJ, Chapman JP (1973).Disorder thought in schizophrenia. Prentice Hall. Englewood Cliffs, NJ.
Cohen JD, Barch DM, Carter C, Servan-Schreiber D. (1999). Context-processing deficits in schizophrJ Ab Psychol 108: 120–33.
Cohen JD, Servan-Schreiber D (1992) Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol Rev 99:45–77.
Escobar M, Oberling P, Miller RR (2002). Associative deficit accounts of disrupted latent inhibition and blocking in schizophrenia. Neurosci Biobehav Rev 26: 203-16.
Everitt BJ, Robbins TW (1997). Central cholinergic systems and cognition. Annu Rev Psychol 48: 649-84.
Flynn FG, Cummings JL, Gornbein J. (1991). Delusions in dementia syndromes: investigation in behavioral and neuropsychological correlates. J Neuropsychiatry Clin Neurosci 3: 364-70.
Garety PA, Freeman D (1999). Cognitive approaches to delusions: a critical review of theories and evidence. Br J Clin Psychol 38: 113-54.
Garety PA, Hemsley DR (1997): Delusions: Investigations into the psychology of delusional reasoning. Psychology Press: Hove.

Han SD, Nestor PG, Shenton ME et al. (2003). Associative memory in chronic schizophrenia: A computational model. Schizophrenia Res 61: 255-63.
Kaney S, Bentall RP (1989). Persecutory delusions and attributional style. Br J Med Psychol 62: 191-8.
Linney YM, Peters ER, Ayton P (1998). Reasoning biases in delusion-prone individuals. Br J Clin Psychol 37: 285-302.
Naveh-Benjamin M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. J Exp Psychol Learn Mem Cog 26: 1170-87.
Nestor PG, Akdag SJ, O'Donnell BF et al. (1998). Word recall in schizophrenia: a connectionist model. Am J Psychiatry 155: 1685-90.

Oberling P., Carnicella S (in press). Toward a rodent model of delusion disorders with construct validity. In: Danion J.M., Jouvent R. Perturbation et récupération des functions cognitives. Maison des Sciences de l'Homme, Paris.
Oberling P, Gosselin O, Miller RR (1999). Latent inhibition in animals as a model of acute schizophrenia: a reanalysis. In: Haugh M, Whalen RE. Animal model of human emotion and cognition. American Psychological Association: Washington. pp 87-102.
O'Caroll RE (1995). Associative learning in acutely ill and recovered schizophrenic patients.
Schizophrenia Res
15: 299-301.
Ramnani N, Owen AM (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat Rev Neurosci 5: 184-94.
Rizzo L, Danion J-M, Grange D et al. (1996) Impairment of memory for spatial context in schizophrenia. Neuropsychology 10:376–84.
White KE, Cummings JL (1996). Schizophrenia and Alzheimer's disease: clinical and pathophysiologic analogies. Compr Psychiatry 37: 188-95.
Young AM, Moran PM, Joseph MH (2005). The role of dopamine in conditioning and latent inhibition: what, when, where and how ? Neurosci Biobehav Rev 29: 963-76.

Discussion

Association or self-monitoring or both?
Anne Reboul
Dec 14, 2005 8:16 UT

I've found Sebastien's and Philippe's paper very interesting. Though my ignorance makes me utterly unable to comment on it, the idea that a common disturbance of dopamine/acetylcholine balance is responsible for delusions in both schizophrenia and AD through a disturbance of associative links strikes me as an exciting hypothesis. Regarding the cognitive side of the hypothesis, the idea that delusion stems from a disturbance of associative weight leading to abnormal causal relationships seems very reasonable in delusions of persecution. A first question is whether it can account for other types of delusion. Another question is whether this hypothesis is proposed as an alternative to the Frith hypothesis, according to which delusions of control in schizophrenic patients are due to deficits in self-monitoring. If not, how do the two hypotheses relate to each other and in which way are they compatible?

Causality vs. Explanation. Objective Relations vs. Subjective Interests.

Denis Hilton (Professor of Social Psychology, Universite de Toulouse)

(Date of publication: 16 January 2006)

Abstract: I will argue that a strong distinction needs to be made between causal attribution and causal explanation. Whereas causal attribution is a cognitivee process that involves referring an event to its source, whether it be a painting to its author, or an event to its origin, explanation is a three-place predicate describing a social interaction whereby someone explains something to someone else. As such causal explanation must obey the rules of conversation: A good explanation must be probably true, informative given an interlocutor?s state of knowledge, relevant to her interests, and expressed clearly. Whereas causality is objective, explanation is subjective (or intersubjective) in nature.

I will argue that a strong distinction needs to be made between causal attribution and causal explanation (cf. Hilton, 1990). For example, attributing the 9/11 attacks to someone is not the same as explaining them to him. Whereas causal attribution is a cognitive process that involves referring an event to its source, whether it be a painting to its author, or an event to its origin, explanation is a three-place predicate describing a social interaction whereby someone explains something to someone else. As such causal explanation must obey the rules of conversation (Grice, 1975); a good explanation must be probably true, informative given an interlocutor's state of knowledge, relevant to her interests, and expressed clearly. Whereas causality is objective, explanation is subjective (or intersubjective) in nature. In the accounts of causal attribution and explanation that I have given elsewhere (Hilton; 1990; Hilton & Slugoski, 1986), I distinguish two phases in the explanation process: a first phase of counterfactual reasoning, and a second phase of contrastive explanation that follows conversational rules. I wish to argue below that the first phase of counterfactual reasoning is objective because the quality of a counterfactual depends on its being an exact description of the objective world. Later I will expand the notion of contrastive explanation by showing how the contrasts of interests are constrained by the implicit purposes behind the causal question posed with respect to a given event.

Causality in the objects: Counterfactuals and causal models

Recent work on causal reasoning in artificial intelligence, philosophy and psychology has substantiated the importance of counterfactual reasoning in causal attribution (see the contribution by Sloman to this seminar, and his 2005 book for a review). This work has done much to clarify and extend the counterfactual analyses of causation given by philosophers such as Hart & Honoré (1985) and Mackie (1980). Nevertheless we should remember that, to echo Austin (1962), causal models are essentially constative in nature: they are more-or-less accurate descriptions of reality, and their quality resides in the predictions they support. So if I believe that the cock's crowing is not the cause of the sun's rising, I can creep out and tape the cock's beak to stop him crowing at dawn – and then observe what happens. Reality will be the judge of whether my causal hypothesis, which can be expressed as the indicative conditional If I stop the cock crowing, the sun will still rise, is correct or not. More generally, we judge the correctness of constative conditionals by their accuracy as descriptions of the nature of the world according to the Ramsey test (Evans, Handley & Over, 2003); that is the probability of the consequent given the antecedent. So if, in a given world of discourse (e.g. France), the indicative conditional If a restaurant has two stars (p) then the food will be excellent(q) will be judged good if the consequent (q) is probable given the antecedent (p). Counterfactual conditionals are also constatives in that they are assertions about the nature of reality, even if they describe events that have not actually happened. They are still judged good if they describe what would have happened if a factor that actually was the case had been "undone". For example, if we agree that it is unlikely that the USA would not have entered the Second World War if Japan had not attacked Pearl Harbour, then we would agree with the counterfactual conditional if Japan had not attacked Pearl Harbour, then the USA would not have entered the war. Causal models can of course influence what constative conditionals (whether expressed as indicatives or counterfactuals) are judged probable or not. For example, many people might be surprised by my assertion If Joan of Arc had lost the Battle of Orleans, then the world would be speaking French. How silly! Or how come? After all, it is normally the conquerors in battle that get to impose their language. Well, to

support the counterfactual conditional, my argument would be to provide you with facts that motivate a new causal model: England at this time was ruled by French-speaking kings of Norman origin. Had they won the Battle of Orleans, England and France (with its larger population) would have been amalgamated into one kingdom, whose language of administration and education would have been French. This kingdom would have dominated Europe and then colonised the world. To the extent that you buy this causal model, then you may accept the Joan of Arc counterfactual conditional as being rational and true rather than silly and idiotic. Causal models thus support indicative and counterfactual conditionals, and all function as descriptions of reality that can be true or untrue. They are thus objective in nature in that their quality depends on their truth-value; that is, their degree of approximation to that reality. While causal models and counterfactual conditionals certainly support explanations, explanations also have an essentially subjective component, having to do with the knowledge and interests of causal inquirers. Causal models and constative conditionals are thus necessary but not sufficient for understanding causal explanation processes.

The pragmatics of explanation

In terms of linguistic theories (Levinson, 1983), causal models and constative (indicative and counterfactual) conditionals are "semantic" (they have truth-values), whereas explanations are "pragmatic" (they have utility values). Explanations are sensitive to context and serve practical interests in conversation. First, experimental work shows that in interpersonal explanations, we follow Grice's maxims of conversation to change our explanations to complement what the other doesn't know. For example, in line with the maxims of quantity, people tend to focus on personal factors when explaining an act of juvenile delinquency to someone who knows of relevant situational factors but not personal ones, but vice-versa if the interlocutor is perceived to know of relevant personal factors but not situational ones (Slugoski et al., 1993). In line with Grice's maxim of relation, people will give an explanation that may seem most relevant to their interlocutor. For example, Norenzayan & Schwarz (1999) found that when experimental participants believed that they were giving explanations of real-life mass murders to a psychologist, they referred more to personal factors than when they believed they were giving explanations to a sociologist, to whom they were more likely to give situational explanations. It seems unlikely that these changes in explanations reflect changes in the participants' underlying beliefs about the "facts of the case" or causal models of the relevant behaviours (e.g. delinquency, mass murder). Intrapersonal explanation is also pragmatic in that the inquirer's satisfaction with an explanation will change with her knowledge-state. Counterfactual reasoning often reveals a plethora of conditions that are necessary for an outcome to occur, yet we tend to identify one or at most two factors in causal explanations. Given the plethora of alternatives, the problem then becomes to select the most relevant and informative explanation, and this will often be the condition that is perceived as abnormal in the circumstances (Hart & Honoré, 1985; Hilton & Slugoski, 1986). For example, many "naïve" observers will have attributed the Concorde's crash in 2000 to the presence of débris on the runway during takeoff, as this certainly seems abnormal, and without its presence the accident would not have occurred. However, "expert" observers in aeronautics consider the presence of débris on runways to be "normal" in that from time to time débris inevitably gets onto runways, and therefore aircraft have to be designed to withstand this. The relevant contrast case for aeronautics experts is the "ideal design" which would have enabled Concorde to withstand the shock of the débris, which they compare to the actual design of Concorde to find the weak point (in this case, the unprotected fuel tanks) which they identify as "the" cause (cf. Hesslow, 1988). The difference in the explanations favoured by naïve and expert judges in this case have nothing to do with the objective "facts of the case" (the perceptions of the crash itself and the sequence of events during take-off), but rather in their subjective interpretation (e.g. differences in what constitute "normal" comparison cases). The privileged status of intentions in human causal explanation A challenge for theories of causal explanation is to understand why free deliberate human actions make such attractive and powerful explanations. For example, few people appear to explain the Concorde's crash by tracing causality through the débris through to the antecedent abnormal condition that was responsible for it (faulty maintenance on the tail of a Continental Airlines jet that took off just before Concorde). Yet experimental data obtained on similar scenarios (Hilton, McClure & Slugoski, 2005) indicates that it seems highly likely that people would have traced causality through to an antecedent condition if had been a deliberate human act of sabotage (e.g. someone placed the débris on the runway expressly to cause the accident). Information gain accounts in unfolding causal chains Counterfactual reasoning appears to be unable to explain this difference, as we would agree in either case that if a) the debris had not fallen off the tail of the preceding airliner, or b) the

saboteur had not placed the débris on the runway, the accident would not have occurred. Probabilistic models of the kind proposed by Spellman (1997) would probably do better on these examples, as data suggests that in these kinds of unfolding causal chains, distal causes appear to increase the probability of outcomes more when they are free deliberate actions than when they are accidental, "natural" causes (Hilton et al., 2005). The probabilistic criterion, that causes are those conditions which most increase the probability of outcomes appears to work well in these unfolding causal chains, where the successive events follow normally and foreseeably from each other in a directed causal sequence. There is a logic of mechanism in their sequence, and they cannot be switched around and still produce the effect. They can even be thought of as sequenced "recipes" for bringing about effects, either in the sense that a saboteur may intentionally concoct a plan to bring about an aircraft crash, or simply in the sense that an unintended piece of debris on runway is a "recipe for disaster". Once the first cause has acted, the chain of causation only has to run normally on for the outcome (e.g. accident) to happen (e.g. the debris pierces the fuel tank, leading to fuel to escape, leading to a catastrophic fire as in the case of the Concorde crash in July 2000). Since free deliberate actions lead to a greater increase in the probability of the outcome they have high information-value. And since the remaining parts of the chain now become predictable, they become redundant details that can be safely omitted from an economical explanation. Focusing on intentional factors in explanations can here be explained in terms of Gricean constraints, such as informativeness.

Social utility vs. information gain accounts in opportunity chains

However, probabilistic analyses fail in what we term opportunity chains where the action of a first cause creates an opportunity for the second cause to act. For example, a natural cause (the refraction of light through broken glass) may cause some shrub to smoulder, allowing a second cause to create a bush fire. This first cause creates an opportunity for a second cause to act, which could either be an intentional action (someone pours petrol on the flames) or a natural cause (e.g. wind springs up). What is interesting here is that people rate the second factor as more strongly causal if it is a deliberate action even if it does not increase the probability of the fire any more than does the natural event. The same applies to experimental variations in the first event: someone deliberately igniting the flames is judged to have increased the probability of a forest fire just as much as does the refraction of light, yet people consider the human action to be more causal (Hilton et al., 2005). It may therefore be that another criterion is at work that leads judges to favour intentional actions as explanations. In particular, following Tetlock's (2002; but see also Smith, 1789/2002) "intuitive prosecutor" analogy, we hypothesised that people may focus on human actions in explanation that allow social control of negative outcomes through the menace of sanctions. We therefore asked participants to evaluate the factors in the event chain in terms of how much social control and preventability they allowed. For example, for the distal cause in the forest fire scenario, the question read "How much do you think society can control and prevent the occurrence of events like a youth setting fire to the shrub in the future (e.g., by warnings, punishments, etc). Judgments were made on a scale from 0% to 100%, with anchor points of 'Not at all' at 0% and 'Completely' at 100%. Mediational analyses showed that social controllability judgments correlated with causality judgments, and mediated the tendency for human actions to be preferred to natural events as causes.

Pragmatic accounts of the preference for human actions as explanations

The finding that perceptions of social controllability mediate perceptions of causality suggests that social utility – and not informativeness – may explain why human actions are preferred as explanations. Future research should address the question of whether social utility would explain the preference for human actions in unfolding causal chains – for example, would the perceived social controllability and preventability of sabotage attempts predict our preference for them as explanations over natural events? Perhaps not – as faulty aircraft maintenance also seems socially controllable through the menace of sanctions, yet this nevertheless does not seem to make it a strong candidate as a cause. Whatever the answer to this question, the selection of explanations will probably depend on subjective characteristics of the hearer rather than objective characteristics of the world.

Conclusions

In this essay, I have attempted to relate recent work in cognitive science on causal modelling and conditional reasoning to earlier work in ordinary language philosophy and social psychology on the

pragmatics of explanation. My conclusion is that causal explanation requires both kinds of analysis. For example, recent cognitive science approaches have addressed the question of how we identify actual causes in pre-emptive causal chains, where an unfolding causal chain, which would have produced the outcome, is interrupted by a new causal chain which takes over and actually produces the effect. An example would be of a fire, which, if left to itself, would inevitably burn down a house. However, another fire reaches the house first, thus pre-empting the first fire from running its course and burning down the house (Halpern and Pearl in press, a; b; see also Mandel 2003). Causal modelling approaches thus seem to be very useful for resolving the attribution question that I raised at the beginning of this essay, in the sense of tracing an outcome to its origins. However, while the AI and cognitive approaches may solve the causality problem, they do not to our knowledge give fully satisfactory answers to the explanation problem, nor tell us how to attribute responsibility. For example, given that we accept that Fire B is the actual cause, what is our preferred attribution in the sense of picking out the factor that is to be held "responsible" for the damage it caused to the neighbour's house? Would it matter if we knew that Fire B came about because someone deliberately fanned some smouldering flames, rather than because a breeze sprang up and brought them to life? Our experimental data suggest that it would, and that a pragmatic analysis in terms of the causal inquirer's knowledge-state and interests is necessary to understand how and why. Indeed, our finding that social controllability predicts causal attribution indicates that it is not just individual interests, but group interests that may drive this process. We would expect the above questions to be posed even if we knew that another concurrent fire would have had exactly the same result, just as we would not dismiss a trial for murder simply because we knew the victim would have been later murdered by someone else anyway. Finally, it seems to me that an attention to pragmatics helps understand what is specifically human about causal explanation. Human beings may share the same kinds of associative learning processes with lower animals such as rats (Shanks & Dickinson, 1988) and even bees. Insofar as they are uniquely capable of representing counterfactual events (that were predicted by context but did not happen) or of engaging in conditional reasoning (Gardenfors, 2003), then they may dispose of cognitive processes that distinguish them qualitatively from other animals. But we will do well to remember that people are also social cognitive animals that communicate through language, who can recognize the intentions and belief-states of others and engage in co-operative communication. They will also seek to control the behaviour of others through expressing anger, holding them responsible for their actions, and identifying the causes and consequences of their behaviour. This social nature will determine human explanation processes in ways that cannot be captured by a purely cognitive perspective that treats people as isolated individuals who use language simply to construct and test representations of their environments.

References

Austin, J.L. (1962). How to do things with words. Oxford, England: Clarendon Press.

Evans, J.St.B.T., Handley, S.H., & Over, D. (2003). Conditionals and conditional probability. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 321-335.

Gärdenfors, P. (2003). How homo became sapiens: On the evolution of thinking. Oxford: Oxford University Press.

Grice, H.P. (1975) "Logic and conversation", in P. Cole and J.L. Morgan (eds) Syntax and Semantics 3: Speech Acts (pp. 41-58), New York: Academic Press.

Halpern, J. and Pearl, J. (in press, a) "Causes and explanations. A structural model approach. Part 1: Causes". British Journal for the Philosophy of Science.

Halpern, J. and Pearl, J. (in press, b) "Causes and explanations. A structural model approach. Part 11: Explanations". British Journal for the Philosophy of Science.

Hart, H.L.A. and Honoré, A.M., (1959/1985) Causation in the law (2nd ed.), Oxford University Press.

Hesslow, G. (1988) "The problem of causal selection", in D. Hilton (ed.) Contemporary science and natural explanation: Commonsense conceptions of causality (pp. 33-65), Brighton: Harvester Press.

Hilton, D.J. (1990) "Conversational processes and causal explanation", Psychological Bulletin 107: 65-81.

Hilton, D.J. McClure, J.L. & Slugoski, B.R. (2005). The course of events: Counterfactuals, causal sequences and explanation. In D. Mandel, D.J. Hilton & P. Catellani (Eds.). The psychology of counterfactual thinkng. London: The Psychology Press. (pp.44-73).

Hilton, D.J. and Slugoski, B.R., (1986) "Knowledge-based causal attribution: The abnormal conditions focus model", Psychological Review 93: 75-88.

Levinson, S.C. (1983). Pragmatics. Cambridge: Cambridge University Press.
Mackie, J.L. (1980) The cement of the universe: A study of causation, Oxford: Clarendon Press.
Mandel, D.R. (2003) “Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning”, Journal of Experimental Psychology: General 132: 419-434.
Norenzayan, A. & Schwarz, N. (1999). Telling what they want to know: Participants tailor causal attributions to researchers’ interests. European Journal of Social Psychology, 29, 1011-1020.
Shanks, D.R. and Dickinson, A. (1988). The role of selective attribution in causality judgment. In D.J. Hilton (Ed.). Contemporary Science and Natural Explanation: Commonsense conceptions of causality. Brighton: Harvester Press.
Sloman, S.A. (2005). Causal models: How people think about reality and its alternatives. Oxford: Oxford University Press.
Slugoski, B.R., Lalljee, M.G. Lamb, R. and Ginsburg, J. (1993). Attribution in conversational context: Effect of mutual knowledge on explanation-giving. i>European Journal of Social Psychology, 23, 219-238. Smith, A. (1789/2002, ed. K. Haakonssen) The theory of moral sentiments, Cambridge: Cambridge University Press.
Spellman, B. (1997) “Crediting causality”, Journal of Experimental Psychology: General 126: 323- 348.
Tetlock, P.E. (2002) “Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians and prosecutors”, Psychological Review 109: 451-471.

Discussion

Setting fire to mono-causation
Robert Stonjek
Jan 18, 2006 11:20 UT

The attribution of causation to a single initial source is an assumption that requires closer examination. Let’s consider once again the fire example.

A fire is deliberately lit by individual A. The fire subsequently burns the house down, killing everybody sleeping inside. With no other information given, can the damage to the house and death of the family be attributed to person A?

Let’s consider some additional detail. The fire is lit in a wood heater, a practice done by this family throughout the winter. On the previous 50 days, the fire was lit in much the same way, and with the consent of the whole family, and no subsequent damage occurred on those previous occasions.

But person B placed wet clothes above the fire to dry them, and it was the catching on fire of the clothes that caused the fire that caused the damage. But clothes had been dried in this way on every rainy day through winter, even though authorities warn that great caution should be exercised when placing clothes near fires.

But person C was responsible for removing the clothes from around the wood heater before going to bed, and it was after everybody went to bed that the secondary fire started – C had forgotten. But C had forgotten on other occasions with nothing more than a singed sock or two resulting.

The fire’s door seal was faulty. The manufacturer of the wood heater had recalled the heater doors for replacement of the door seal but no-one in the family was aware of this. Even so, regular maintenance would have seen the door seal renewed by about now anyway, and that was the responsibility of person D. It was the escaping hot gas through the faulty door seal that caused the additional heat on this occasion that caught the clothes on fire that burnt the house down killing persons A through to D.

Let’s look at the causation:
If **A** had not lit the fire or

If **B** had not placed clothes above the wood heater to dry or
If **C** had remembered to remove the clothes at bedtime or
If **D** had been vigilant with the maintenance or
If **any of them** had noticed the recall notice or If the **manufacturer** had not sold a faulty product,
The fire would not have gotten started in the clothes, burning the house down and killing all inside. Two such fires (house destroyed, family wiped out) occurred in our region last winter.

But what if B had worn his socks another day, or thrown them out? The fire may not have started on that night but on some other night or not at all.

All too often, in real life, the causal source is spread over a number of events which all contribute to the outcome, all are essential, but where no single or initial causative agent can possibly be identified. (lighting the wood fire was not the cause of the house fire).

Kind Regards. Robert Karl Stonjek

The problem of causal selection
Denis Hilton
Jan 18, 2006 23:01 UT

Thanks to Robert Karl Stonjek for this very illustrative example. It Illustrates the point that counterfactuals can identify conditions that are all necessary for the outcome (A's lighting the fire, B's placing clothes over the wood heater to dry, C's failure remove the clothes at bedtime, lack of maintenance, failure to notice the recall notice, and the manufacturer's sale of a faulty product). It is true that the house fire would not have happened if any of these conditions were undone. In this sense, the cause "philosophically speaking" (J.S. Mill) will be the ensemble of necessary conditions defined this way, and is important to understanding how the event happened. But I do not think that people would always cite the whole causal story in an explanation.

In this example, the outcome is due to a number of (in)actions that violate norms – hanging the clothes despite the authorities' warnings, failure to clear the hanging clothes, the manufacturing fault, lack of vigilance over maintenance, failure to hear about the manufacturer's recall. All these are thus abnormal conditions that are very strong candidates as causes in the law (Hart & Honoré, 1985) as they are in ordinary human causal ascription (Hilton & Slugoski, 1986).

But to cite all these abnormal conditions in an explanation might seem a bit long-winded. In ordinary conversation, we tend to select just one or two factors to mention as "the" cause. I think that this selection will tend to follow pragmatic interests. These could be informational: suppose a causal inquirer knows the family's routines, and wants to know what happened this time that caused the fire. Here the explanation would focus on factors that distinguish this tragic night from other, "normal" nights (e.g. forgetting to take the clothes away, and lack of maintenance). In contrast, an inquirer who knows that these wood stoves are faulty, but knows nothing about the family in question, would need to know what distinguished this family's case from the "normal" case where the fire did not happen.

Finally, how would we go about attributing responsibility in this case? First, the initiating action of lighting the fire was deliberate, but did not have the purpose of burning the house down. For example, had someone knowingly given the family these defective burners with the intention of causing a house fire, I think we would have no hesitation in tracing the house fire back to this "source". When I talk of "deliberate" actions in my position paper, I mean ones which have the aim of producing the outcome in question.

However social roles may be important. For example, the Regional Safety Officer may take the view that it is in people's nature to be forgetful from time to time (just as aeronautics experts know there will be débris on runways from time to time). So he would presumably focus on the manufacturer's obligation to produce a fail-safe design, and designate his failure to do so as "the" cause.

Re: The problem of causal selection
Robert Stonjek
Jan 21, 2006 23:30 UT

In law (locally, here in Australia), any coroner considering these conditions would rule that *the* cause to be “an unfortunate combination of events and/or conditions” and would not assign any one event/condition as *the* cause. In litigation, the manufacturer would be found to be only a contributing element in the events and so only partially liable. ‘Mono-causation’ seems to be a more attractive model of events in the USA (my perception).

It is true that people familiar with the family in my story would cite only a single cause, but they would not all cite the same cause. Those unhappy with that brand of wood heater would quickly blame the manufacturer; those who thought that one of the family members was unreliable may attribute the events to them.

A recent example of multiple causation is the unfortunate destruction of the space shuttle on re-entry. The contributing factors are complex. The attributed cause was the foam coming loose from the fuel cell and striking the shuttle’s wing. But foam had come loose before and foam had struck and damaged the shuttle before.

Attempts had been made to fix the problem but NASA’s budget had been cut and they didn’t have the resources to fully address the problem. Their options were to refuse to fly and put themselves out of business, or hope that the latest attempt to better secure the tiles would work – a wing and a prayer?

But simply damaging the shuttle is not sufficient for the disaster to occur. The crew were aware that there was damage and had visually checked it through the shuttle’s window. But they were told not to worry, partly because there was nothing that could be done – the shuttle’s orbit was too low for a rescue involving the space station and the crew were not equipped for space walks and even if the were they had no tools or spares such a repair.

This makes for a very complicated causation path, as many of the problems can be traced back to budget cuts but no budget cut was directly responsible for the events of that particular flight.

We have, then, two levels of causation (in both examples). We have the conditions present when the event in question took place (and the causative agents of those events) and the actual sequence of events that results in the disaster.

If the clothes had been removed from the heater or the foam had missed the wing on that occasion then the manufacturer’s faulty product or the budget cuts would still be present in identical form.

Without considering causative conditions alongside causative events, any mono-causation is just a matter of opinion which will be decided not on the actual causative agent, but on the competency of each side debating the issues. Kind regards. Robert Karl Stonjek

Re: The problem of causal selection
Robert Stonjek
Jan 21, 2006 23:30 UT

In law (locally, here in Australia), any coroner considering these conditions would rule that *the* cause to be “an unfortunate combination of events and/or conditions” and would not assign any one event/condition as *the* cause. In litigation, the manufacturer would be found to be only a contributing element in the events and so only partially liable. ‘Mono-causation’ seems to be a more attractive model of events in the USA (my perception).

It is true that people familiar with the family in my story would cite only a single cause, but they would not all cite the same cause. Those unhappy with that brand of wood heater would quickly blame the manufacturer; those who thought that one of the family members was unreliable may attribute the events to them.

A recent example of multiple causation is the unfortunate destruction of the space shuttle on re-entry. The contributing factors are complex. The attributed cause was the foam coming loose from the fuel cell and striking the shuttle's wing. But foam had come loose before and foam had struck and damaged the shuttle before.

Attempts had been made to fix the problem but NASA's budget had been cut and they didn't have the resources to fully address the problem. Their options were to refuse to fly and put themselves out of business, or hope that the latest attempt to better secure the tiles would work – a wing and a prayer?

But simply damaging the shuttle is not sufficient for the disaster to occur. The crew were aware that there was damage and had visually checked it through the shuttle's window. But they were told not to worry, partly because there was nothing that could be done – the shuttle's orbit was too low for a rescue involving the space station and the crew were not equipped for space walks and even if the were they had no tools or spares such a repair.

This makes for a very complicated causation path, as many of the problems can be traced back to budget cuts but no budget cut was directly responsible for the events of that particular flight.

We have, then, two levels of causation (in both examples). We have the conditions present when the event in question took place (and the causative agents of those events) and the actual sequence of events that results in the disaster.

If the clothes had been removed from the heater or the foam had missed the wing on that occasion then the manufacturer's faulty product or the budget cuts would still be present in identical form.

Without considering causative conditions alongside causative events, any mono-causation is just a matter of opinion which will be decided not on the actual causative agent, but on the competency of each side debating the issues. Kind regards. Robert Karl Stonjek

The Possible Influence of Perception of Causal Events on the Development of "if P then Q" Conditionals and Causal Reasoning

Peter Ford F. Dominey (Researcher, CNRS - Institut des Sciences Cognitives, Lyon)
(Date of publication: 15 February 2006)

Abstract: This note will address the development of a generalized 'if P then Q' schema based on accumulated real-world experience with the perception of causal events, in a constructionist context.

Background

The perhaps non-controversial context in which this work is set holds that a number of high-level social communicative and cognitive functions are based on a generalized mechanism for mapping structure in one domain onto structure in another domain (Dominey in Press). The initial domain of this method of inquiry has been language, developed by the "construction grammar" (CxG) school (e.g. Fillmore 1988, Goldberg 1995). In CxG, grammatical constructions are mappings from sentence form to meaning, and language is considered to be made up of a structured inventory of these mappings. We have recently implemented a neurophysiologically grounded model of sentence processing that learns to perform these mappings (Dominey et al. 2003, Dominey & Boucher 2005). To summarize with an example, a ditransitive sentence form such as "X was Y by Z to Q" maps onto a predicate-argument representation of a ditransitive event of the form Y(Z, X, Q) where Y, Z, X and Q in the sentence correspond to a ditransitive verb, an agent, object and recipient noun, respectively, and Y, Z, X and Q correspond to their respective lexical semantic meanings. For a variety of construction types, we demonstrated that a generalized structure mapping system could then map sentences to meanings, in order to develop abstract representations of the form to meaning mappings, i.e. grammatical constructions (Dominey 2002; Dominey et al. 2003). To validate the psychological reality of this approach, using a human-robot interaction set-up, we asked naive human subjects to perform transitive actions with simple objects in the field of view of a video camera, and at the same time to narrate these events, in order to provide (sentence, event) pairs to teach the system based on the meaning descriptions and their grammatical forms. Predicate-argument meanings were extracted from the video sequence based on a decomposition of the events into discrete sequences of physical contacts between the different entities. Indeed, this provided concrete evidence that event descriptions including assignment of agent, object and recipient roles could proceed in a fully mechanistic manner based on extraction of low level perceptual primitives. We thus demonstrated in a grounded robotic system, that by extracting these predicate-argument meaning representations from actual video scenes of causal transitive actions, and pairing these representations with the corresponding human narration of these events, the system could learn a miniature event language and the associated set of grammatical constructions (Dominey & Boucher 2005)

Objective

The objective of the current exercise is to argue that this kind of form-meaning mapping capability can apply in the domain of conditional sentences and causally related events. Specifically we consider the mapping between an "if X then Y" form and a corresponding meaning that can be extracted from the physical world. The corollary objective is that the resulting "if-then" construction can then be applied to arbitrary non-physically causal relations.

Method

The question immediately arises, what is the nature of this "causal" meaning and where does it come from. Up to this point we have considered event predicates like give(John, Ball, Mary) corresponding to John gave the ball to Mary. In order to capture the required relations for the current purposes, we must now take into consideration the states and state changes that occur as the result of events/action. Before and after an event such as give, the world is in two distinct states, and the event can be considered as a transition between these states. This corresponds to a triplet (enabling-state, action, resulting-state). From a "good old AI" perspective, this is an alternative manner of characterizing a production rule, in which the left hand side is the enabling state, and the right hand side is the resulting state change after the action of the rule is applied. This triplet can be broken into two pairs (enabling-state, action) and (action, resulting-

state). In our example, these two pairs correspond to (has(John, ball), give(John, ball, Mary)) and (give(John, ball, Mary), has(Mary, ball)). Anticipating the mapping of language onto these meaning pairs yields the "enabling if-then" "If John has the ball then he can give it to Mary" and the "resulting if-then" "If John gives the ball to Mary then she will have it".

In a scenario in which an infant (or a robot) observes numerous transitive and ditransitive physical events, it will be exposed not only to these events, but to the enabling initial conditions/states and resulting final conditions/states as well. We can thus propose that through the kind of mechanistic perception of events, a structured set of ordered pairs of the form (enabling-state, action) and (action, resulting-state) will develop. When paired with sentences such as "If you push that then it will fall" or "if you push that, it will fall" the structure-mapping mechanism will begin to create "If X then Y" conditional constructions.

Appeal to Behavioral Development

The concept of form to meaning mapping has a long history in CxG, and we have recently demonstrated how such mappings can be performed (Dominey 2002, Dominey et al. 2003, Dominey & Boucher 2005). The question is whether this approach can extend to conditional constructions, and in particular the "if then" construction that involves mapping onto an (action, resulting-state) representation. For our purposes it would be nice if we could assume that indeed these (action, resulting-state) representations have some psychological reality. One likely avenue would be the link between action and goals in goal directed behavior. Developmental research indicates that agentive experience contributes to the creation of goal directed action representations in infancy, as during the first year of life the production of goal directed actions become progressively frequent, precise and refined (Sommerville et al. 2005). This implies the development of relations linking actions to goals such as in the (action, goal) pair (take(me, ball), have(me, ball)). Another avenue for building up (action, resulting state) representations would be from more basic physical regularities corresponding to observations that if one pushes an object it moves, if one drops an object it falls, etc. Numerous studies of infants' perception of physical events (e.g. Spelke 1991, Baillargeon 1994) demonstrate that early in the first year of life infants develop representations including those that would correspond to "if you push the block past the edge of the table, it will fall off", as revealed by infants predictive anticipation of such consequent events.

It is not controversial then to assume that during the first year of life, infants are developing and interpreting the world around them with representations of the form (physical event, resulting state) in a form of predictive relation.

Linking Causal Event Representations to Conditional Constructions

The proposition of a construction-based approach to linking "if-then" forms with predictive event structure is a rather obvious and not new. Indeed, Dancygier (1998) has taken an approach in which conditionals are considered in the framework of constructions (as defined by Fillmore 1988, Goldberg 1995 and others) that pair form with meaning. Dancygier notes that the analysis of conditionals has a long and varied history, with logical, truth conditional analyses vs analyses based on the form of the conditional, with the resulting possibility that many of the accounts do not even share a common view of what a conditional is. Dancygier thus proposes that a construction based analysis that pairs form with interpreted meaning can help. She further notes that in a move to clarify the situation one can distinguish between content, epistemic and speech act domains for interpretation of conditionals. In the content domain the clauses are linked causally, while in the epistemic and speech act domains, they are linked by more abstract relations. Indeed in the epistemic and speech act domains the presence of "if" indicates an instruction for the hearer to treat the assumption in its scope as not being asserted in the usual way. Our position will be that by grounding the development of if-then constructions in the content domain, the system provides itself with a basis for extrapolating to the epistemic and speech act domains. But the current exercise will be limited to the content domain.

Establishing the link

Here we will work through the process. As stated, we assume that the system will be capable of parsing event structure from the visual input, as we have previously demonstrated for physical events, as well as

for the enabling and resulting states. We previously categorized the events in terms of contact and its parameters such that, for example, take(Agent, object, source) is recognized as a contact between that agent and the object, co-motion of the agent and object, and the ending of a contact between the object and source. Remaining within this domain of actions that include touch, push, take, and give, the state variable that will perhaps be most salient is that of "possession". Indeed, goal directed behavior is often "possession oriented". So how is possession recognized? Interestingly, the notion of possession can be represented without the requirement for any new perceptual primitives, as it is captured by the primitive "contact." In this context, the actions give and take both produce clear changes in the possession status of different agents and objects. In the internal mechanistic representation extracted from the video sequence, there will thus be the contact (possession) status of elements in the world before a given event, then a representation of the event, followed by a representation of the resulting state. In this context, we can consider the generation of sentence meaning pairs such as:

1. "John has the block", contact(John, block)
2. "John gives the block to Mary", give(John, Block, Mary)
3. "Now Mary has the block", contact(Mary, block)

The observations give(John, Block, Mary) and contact(Mary, block) will enter into a predecessor – successor relation that will co-occur with statistical regularity. Indeed, give(Ag, Obj, Rec) and contact(Rec, Obj) will co-occur with high regularity, allowing a statistical learning mechanism to link them together in a relationship of temporal succession such that the subsequent presence of the predecessor element will yield an internal prediction of the arrival of the successor element. Again, the human developmental manifestation of this type of phenomena is observed in the studies of Spelke, Baillargeon and numerous others indicating that infants will anticipate the perceptual outcome of physical events.

Now the groundwork is laid for establishing the form to meaning mapping between "if then" utterances and these temporally associated elements. The phrasal constituents "John gives the block to Mary" and "Mary has the block" can be inserted into the "if then" structure to yield "If John gives Mary the block then she will have the block." Related (and perhaps more basic) forms could be "If you push that it moves" or "If you drop that it will fall," both of which are similarly grounded in the corresponding perceptual event temporal sequences.

The result of this binding of "If p then q" construction to phrases p and q which enter into temporal succession relations is that part of the semantics of this construction is the explicit coding of this temporal succession relation between p and q. That is, the "If p then q" encodes that the prediction relation holds between p and q.

Generalization

Knowledge about prediction/succession relations of the world can acquire representational status in a developing system by at least two distinct pathways. First, via observation of physical events, the system can acquire sufficient data to extract a prediction relation between two events that reliably co-occur in a repeating temporal sequence, as outlined above. Further, as outlined above, this succession or prediction relation can become linked to a grammatical "if p then q" construction where p and q correspond to the respective predecessor and successor elements. This provides the basis for the second pathway for acquiring knowledge of prediction relations – via use of the "if p then q" construction. As formalized by Goldberg (1995), a new construction is created when the intended meaning cannot be derived purely from the combination of the constituents. Thus the "if p then q" construction encodes the added information corresponding to the prediction or succession relation that holds between p and q. Thus, this construction can be used to extend or project this relation onto constituents p and q which have not otherwise been observed to enter into this relation.

Summary

Temporal succession or predictability relations that are robustly present in the physical world can be extracted by essentially mechanistic perceptual systems (e.g. as an extension to that implemented in Dominey & Boucher 2005). Via a form to meaning mapping mechanism, a grammatical "if p then q" form

can become associated with the succession/prediction relation. Once this mapping is established it can be generalized, and applied, via use of sentences built for the grammatical construction, to link previously unrelated pairs of constituents via this succession/prediction relation. This provides a basis for establishing the “content” domain (see Dancygier 1998) for interpreting if-then conditionals in terms of concrete causal relations. Future research should then examine whether this provides a stepping stone towards the use and interpretation of such conditionals in the epistemic and speech act domains.

References

Baillargeon, R. (1994). "How do infants learn about the physical world?" Current Directions in Psychological Science, 3, 133-140.
Dancygier, B. (1998) Conditionals and Predication: Time, Knowledge and Causation in Conditional Constructions, Cambridge University Press.
Dominey, P.F. (2002) "Conceptual Grounding in Simulation Studies of Language Acquisition", Evolution of Communication (2000), 4 (1) 57-85.
Dominey, P.F. (2005) "Toward a construction-based account of shared intentions in social cognition", Comment on Tomasello et al. "Understanding and sharing intentions", Behavioral and Brain Sciences (2005) 28:5, 696.
Dominey, P.F. (In Press) "Towards a Construction-Based Framework for Development of Language, Event Perception and Social Cognition: Insights from Grounded Robotics and Simulation"
Dominey, P.F., Boucher J.D. (2005) "Learning To Talk About Events From Narrated Video in the Construction Grammar Framework", Artificial Intelligence, 167 (2005) 31–61.
Dominey, P.F., Hoen M., Blanc J.M., Lelekov-Boissard, T. (2003) "Neurological basis of language and sequential cognition: Evidence from Simulation, Aphasia and ERP Studies", Brain and Language, 86(2):207-25.
Fillmore, C.J. (1988) The mechanisms of “Construction Grammar,” Berkeley Linguistics Society, 14, 35-55.
Goldberg, A.E. (1995) Constructions: A Construction Grammar Approach to Argument Structure, University of Chicago Press.
Spelke, E.S. (1991). "Physical Knowledge in Infancy: Reflections on Piaget's theory". In S. Carey & R. Gelman (Eds.) The Epigenesis of Mind: Essays on Biology and Cognition (pp. 133-169). Hillsdale, NJ: Erlbaum.
Sommerville, J.A., Woodward A.L., Needham, A. (2005) "Action experience alters 3-month-old infants’ perception of others’ actions". Cognition 96, B1-B11.

Discussion

Real Infants need more than observation
Robert Stonjek
Feb 25, 2006 12:01 UT

Infants in the process of the development of their understanding of causation in the procession of events, such as observing the passing of a ball from one child to another, are also exposed to numerous counterintuitive and complex sequences that are not interpretable through simple observation only.

Johnny, for instance, may drop the ball, lob the ball to Mary, or toss it directly up into the air via a hand action imperceptible (or confusing) to the observing infant. Johnny may offer the ball to Mary, who previously had no interest in the ball, then withdraw the ball to howls of protest from Mary who now takes an interest in the ball and demands sharing.

From the numerous possible real-world scenarios, the infant must select only the “Johnny gives the ball to Mary” event as a template for understanding possessional transactions? Unlikely!

The important additional step is the emulation of the behaviour, or some part of it, by the infant ahead of learning. In fact an infant at first learns almost exclusively by experience before observation and experiment and finally from observation alone (which never runs to completion, even in adults.)

Until the age of around four years old, the child references all events back to the self exclusively. Only after the development of ‘theory of mind’ can a child see events from the third person perspective. A good example of this is the so called ‘Sally-Anne’ test.

Sally and Anne are in a room together, observed by the infant. Sally places an object in a box and leaves the room. Anne removes the object and places it in her pocket. Sally returns and the child is asked where she will look to find the object. Children under (on average) 4.5 years old will say “in Anne’s pocket”, older children can see the problem from Sally’s perspective and so answer “in the box”.

Thus there is substantial non-trivial information processing occurring in even the youngest infant’s brain when considering even seemingly obvious transactions, and only after repeated exposure and self experimentation do they learn even the simplest if *p* then *q* transactions.

Kind regards. Robert Karl Stonjek

Expressing Causality in Natural Language. A Pragmatic Perspective

Jacques Moeschler (Professor, Unversité de Genève)

(Date of publication: 25 March 2006)

Abstract: This paper is devoted to the semantics and the pragmatics of causal constructions in natural language, mainly causative constructions, causative predicates and causal discourses with and without connective. I will argue that the representation of causality in discourse (the order effect-cause) is connected to the semantics of causal constructions, whose necessary argument is the patient (of the effect) and whose predicate denotes the resulting state of the causal event.

Causality is not a linguistic concept, but language is the best only tool human beings possess to express causality between events or, more generally, between states of affairs. In this contribution, I would like to show that the fine study of the linguistic means to express causality is of great interest for the study of human cognition. The main purpose of my paper is to illustrate two very interesting facts about linguistic causation.

The first fact concerns the syntactic means natural languages possess to express causality. As an anticipation of one of my claims, I will argue that the general pattern linguistic constructions offer to express causality is not the explicit description of an event and its agent, but the explicit description of the resulting state and its patient.

This observation is very surprising, because an a priori and intuitive description of causality would predict the opposite: causality is inferable from events and agents. The second point is at a first sight very strange, and until now unexplained: when we have to report causal relations between events or states, the linguistic order is the consequence-cause order, and not the cause-consequence order. Any attempt to connect the cognitive properties of causal relations and linguistic reports of causal relations would predict, for reason of processing, that the natural order between events (cause-consequence) would be the linguistically preferred order (Ahn & Noseck 1998). What is surprising is that natural languages all possess at least one causal connective (engl. ‘because’, fr. ‘parce que’, it. ‘perche’, etc.) that imposes the consequence-cause order:

(1) Axel fell because Abi pushed him.

(2) Axel est tombé parce que Abi l’a poussé.

What is surprising is that if we change the order of the events in (1) or (2), the causal reading is still accessible, but the appropriate interpretation is one in which the falling event causes the pushing event (for instance to put the patient out of danger), the push-fall causal reading (called ‘inferential order’) being more difficult to access:

(3) Abi pushed Axel because he fell.

(4) Abi a poussé Axel parce qu’il est tombé.

What is still more puzzling is the following fact: if we suppress the causal connective, as in (5) and (6), only the consequence-cause order in (5) triggers the causal reading, whereas (6) is typical of what linguists call narratives, that is, discourses where the linguistic order parallels the chronological order of events:

(5) Axel fell. Abi pushed him.

(6) Abi pushed Axel. He fell.

We have now all the basic linguistic material allowing discussing the causality puzzle: (i) why is (5) the natural order for conveying a causal report of events and (ii) why does the causal connective (‘because’)

impose the consequence-cause order?

I will give two main arguments for the cognitive motivation of the consequence-cause order in causal reports: the first refers to the semantics of causal constructions and the second to the linguistic distribution of connectives. Finally, I will give an experimental argument in favour of the consequence-cause order.

A general semantic pattern for causal constructions

The syntax and semantics of causality has been extensively and very precisely analysed (Kaynes 1975, Levin & Rappaport, Jackendoff 1990, Pustejovsky 1995 among others). Linguists recognize three syntactic means to express causality: (i) causative constructions, (ii) ergative constructions and (iii) inaccusative constructions. The first strategy is used either when no causative verb (like 'kill' = cause to die, 'sink' = cause to be sunk, 'bake' = cause to become baked, etc.) is accessible in the lexicon. Languages differ widely on this point: for instance, French has a very poor causative verbal lexicon, whereas English can easily give a causative meaning to transitive verbs, as in (7):

(7) Bill raced Aga Khan's horse at the Prix d'Amérique (= made the horse run)

(8) *Bill a couru le cheval de l'Aga Khan au Prix d'Amérique (≠ a fait courir le cheval)

Causative constructions use a predicate operator ('make' in English, 'faire' in French) which trigger the so-called causative reading, as in (9):

(9) Mary made the children eat the soup.

(10) Marie a fait manger la soupe aux enfants.

What seems to be an universal property of causative construction in natural language is that the use of the causative operator in addition to a causative verb conveys the implicature (= the pragmatically inferred meaning) that the means to cause the event is not an ordinary one.

Compare (11) and (12):

(11) Bill stopped the car.

(12) Bill made the car stop.

The implicature in (12) is that Bill stopped the car by a special means, the handbrake for instance. The second means to express causality in natural language consists in using a causative verb in a transitive construction. This type of construction, called ergative (the ergative case is the mark of the agent in ergative languages, such as Basque), has the special property to entail a resulting state, described by the causative predicate:

(13) Bill stops the car → the car is stopped

(14) Mary opens the door → the door is open

(15) Autumn yellows the leaves → the leaves are yellow

Some of these verbs can have an intransitive use, in which the agent (ergative case in ergative language, nominative case in nominative-accusative languages) is deleted, but the resulting state is still entailed:

(16) The car stops → the car is stopped

(17) The door opens → the door is open

(18) The leaves yellows → the leaves are yellow

If we now try to find the common properties between these three strategies for expressing causality, what we have to target are the basic parameters expressed by these constructions. Causative constructions express the agent and the caused event or state, containing respectively an agent (optionally a theme, i.e. the object of the event) and a patient:

(19) Mary (AGENT) made the children (AGENT) eat (EVENT) the soap (THEME)

(20) The wind (AGENT) made the leaves (PATIENT) fall (EVENT)

In the ergative (transitive) constructions, the agent, the causal event and the patient are expressed, but the event entails the resulting state (the meaning of a causative verb is to entail the resulting state)

(21) John (AGENT) broke (EVENT) the branch (PATIENT)

Finally, the inaccusative (intransitive) construction only expresses the patient (in the subject position) and the event, which also entails the result-state:

(22) The branch (PATIENT) broke (EVENT)

Now, the answer to the previous question is easier to give: what is common to all these constructions is the event predicate and the patient. And the event entails the resulting state, that is, what is meant by the causative verb or the causative construction. So the conjecture I would like to propose is the following: what creates a causative meaning in a sentence is not the agent or the event functions, but the resulting state and the patient ones. In other terms, the explicit description of a patient and of a result state is what makes the construction of a causal meaning possible. If the conjecture is correct, then we have a first argument to explain why causal discourses have the consequence-cause order: the consequence of the causal-event is either an event or a state, whose argument is a patient. The practical implication of the argument is the following: the utterance of (23) and (24), which describe states, are good candidates for being followed by an explanation, whose pragmatic function is to state the causal event:

(23) Axel is sick.

(24) John is dead.

(25) Axel is sick. He ate too much chocolate.

(26) John is dead. He was killed in a road accident.

So, the semantics of causative constructions can be given by the following schema: the resulting state is part of a complex structure containing the event predicate and the agent, none of them being mandatory. Figure 1 represents the general pattern, and Figure 2 the realisation of the pattern for (22):

Figure 1: semantic structure of a causative sentence

Figure 2: semantic structure of ‘the branch broke’

The distribution of causal, inferential and temporal connectives

The second argument I would like to present is linked to the distribution of connectives in French, i.e. causal connective ‘parce que’ (‘because’), inferential connective ‘donc’ (‘so’) and temporal connective ‘et’ (‘and’). One relevant distributional fact is the order of causal relations with connectives. Causal connectives as ‘parce que’ (‘because’) generally introduce the cause. The combination of event and state (as cause or consequence) gives four possible pairs of utterances, given in (27), all interpreted as causal in the consequence-cause order. (28) shows the reverse order (cause-consequence) implying the inferential use of ‘parce que’ (i.e., the cause-consequence order). (29) presents cause-consequence sets of utterance with an inferential connective (‘donc’, ‘so’), (30) consequence-cause sets of the reverse, (31) and (32) the same structure with a temporal connective (‘et’, ‘and’).

(27) Canonical (consequence-cause) series with ‘parce que’: causal readings

a. CAUSE (EVENT, STATE) : Marie est malade parce qu’elle a trop mangé

‘Mary is sick because she ate too much'

b. CAUSE (EVENT, EVENT) : Jean est tombé parce que Marie l’a poussé

‘John fell because Mary pushed him‘

c. CAUSE (STATE, STATE) : Marie ne peut pas boire d’alcool parce qu’elle est mineure

‘Mary cannot drink alcohol because she is a minor‘

d. CAUSE (STATE, EVENT) : Le médecin soigne Axel parce qu’il est malade

‘The doctor is treating Axel because he is sick’

(28) Non-canonical (cause-consequence) series with ‘parce que’: inferential readings

a. Marie a trop mangé, parce qu’elle est malade

‘Mary ate too much, because she is sick’

b. Marie a poussé Jean, parce qu’il est tombé

‘Mary pushed John, because he fell’

c. Marie est mineure, parce qu’elle ne peut pas boire d’alcool

‘Mary is a minor, because she cannot drink alcohol’

d. Axel est malade, parce que le médecin le soigne ‘Axel is sick, because the doctor is treating him’

(29)Canonical (cause-consequence) series with ‘donc’: inferential and causal readings

a. # Marie a trop mangé, donc elle est malade (# means a change in reading)

‘Mary ate too much, so she is sick’

b. # Marie a poussé Jean, donc il est tombé ‘Mary pushed John, so he fell‘

c. Marie est mineure, donc elle ne peut pas boire d’alcool

‘Mary is a minor, so she cannot drink alcohol’

d.Axel est malade, donc le médecin le soigne

‘Axel is sick, so the doctor is treating him’

What happens in (29) with the cause-consequence order of ‘donc’ is that the causal readings are restricted the states as causes, event-causes giving rise to inferential reading.

The test to trigger the inferential reading is the following: if the consequence can be false, then the reading is inferential.

(30)Non-canonical (consequence-cause) serie with ‘donc’: inferential readings

a. Marie est malade, donc elle a trop mangé

‘Mary is sick, so she ate too much’

b. Jean est tombé, donc Marie l’a poussé

‘John fell, so Mary pushed him’

c. Marie ne peut pas boire d’alcool, donc elle est mineure

‘Mary cannot drink alcohol, so she is a minor’

d. Le médecin soigne Axel, donc il est malade

‘The doctor is treating Axel, so he is sick’

(31)Canonical (cause-consequence) series with ‘et’ : causal and inferential readings

a. Marie a trop mangé et elle est malade

‘Mary ate too much and she is sick’

b. Marie a poussé Jean et il est tombé

‘Mary pushed John and he fell’

c. # Marie est mineure et elle ne peut pas boire d’alcool

‘Mary is a minor and she cannot drink alcohol’

d. ? Axel est malade et le médecin le soigne

“Axel is sick and the doctor is treating him”

(32)Non-canonical (consequence-cause) series with “et”: inferential reading impossible

a. ?? Marie est malade, et elle a trop mangé

“Mary is sick, and she ate too much”

b. ?? Jean est tombé, et Marie l’a poussé

‘John fell, and Mary pushed him’

c. ?? Marie ne peut pas boire d’alcool, et elle est mineure

“Mary cannot drink alcohol, and she is a minor”

d. ?? Le médecin soigne Axel, et il est malade

“The doctor treats Axel, and he is sick”

The following table gives a summary of these distributions:

Table 1 : causal and inferential readings of ‘parce que’, ‘donc’, ‘et’

The picture is rather interesting: causal readings cannot be obtained in the cause-consequence with donc when the cause is an event and with et when the cause is a state: parce que is the only connective allowing the causal reading whatever the cause is (a state or an event). But the cost to ensure causal relation is the consequence-cause reading. This order triggers a necessary constraint: the consequence-cause with donc always yields an inferential reading, and no readings are accessible with et. So, the causal readings with the cause-consequence order are obtained either by donc (the cause is a state) or by et (the cause is an event), the consequence of this distribution being that parce que is the optimal candidate for causal readings, whatever the cause is. We have thus a second argument for the consequence-cause order: this order is the only one compatible with any type of aspectual classes (state or event) and corresponds to the canonical order of causal relations with a connective (parce que).

Experimental data on causal discourses

The connective argument is not strong enough: it should be tested with other languages than French to have a greater strength. What I would like to do now is to add an third argument, based on experimental data. I will give a very brief survey of a first interesting conclusion obtained from two experiments. The first experiment was a simple elicitation task: we show a series of event sentences composed of 8 syllables and ask 38 students from University of Lyon 2 to complete the stimuli either by a cause (20 students) or a consequence (18 students). 36 stimuli were offered to each subject. From these 36 initial propositions, 10 pairs of propositions have been selected, 5 pairs of highly associated propositions (more than 50% of given responses) with the consequence-cause and the cause-consequence order, 5 weakly associated (less than 35% of given responses) with both orders. The first condition to be tested was the strength of association, and the second condition the order of utterances (cause-consequence and consequence-cause).Table 2 (Moeschler et al. 2006) gives the series of inputs for the second experiment:

strength of association	proposition 1	proposition 2 (consequence)	% answers	proposition 2 (cause)	% answers
strong	Paul a pris ses médicaments,	il va guérir.	50	il était malade.	94
	Le gendarme a beaucoup couru,	il est essoufflé.	85	il poursuivait quelqu'un.	94
	Jérôme a arrosé les plantes,	elles poussent mieux.	50	elles avaient besoin d'eau.	55,5
	Jean s'est acheté des lunettes,	il voit mieux.	70	il avait des problèmes de vue.	50
	Le vase de cristal est tombé,	il s'est cassé.	70	quelqu'un l'a fait tomber.	50
weak	Marie s'est tordu la cheville,	elle doit se soigner.	20	elle faisait du sport.	16,6
	La barque a heurté le rocher,	elle a coulé.	35	il y avait du courant.	16,6
	Marie a lu sans ses lunettes,	elle n'a rien vu.	15	elle voit bien de près.	22,2
	Le chien a attrapé des puces,	on va l'emmener chez le vétérinaire.	20	il s'est roulé dans l'herbe.	16,6
	Véronique s'est lavé les mains,	elle va passer à table.	25	elle avait jardiné.	22,2

Table 2: ten selected propositions for the experimental

These 20 likely utterances were balanced by 20 unlikely utterances (in the order cause-consequence and consequence-cause), and all of these 40 utterances have been checked by control-utterances. 22 subjects passed the cause-consequence series, 27 subjects the consequence-cause series, and 22 the control-utterances. The design of the experiment was implemented with the E Prime software, and the subject, after having read the prompt, had to read the second proposition and type ‘e’ or ‘p’ for ‘likely’ or ‘unlikely’ . Reading time has been recorded, the analysis dealing with reading time. The results are the following: (i) with high associated pairs of propositions, no significant difference in reading time occurs; (ii) on the contrary, with weak associated pairs of propositions, a significant difference in reading time occurs: the causal reading (consequence-cause) is quicker than the cause-consequence order (164,80 ms against 308,84 ms). This result must be precisely analyzed, and our data have now to be completed by new experiments. At this stage, the precise interpretation is not certain, but data from the second experiment allows us to conclude that the consequence-cause order has some cognitive motivation. In effect, it seems that when no strong conceptual association between event predicates exist, the causal reading is more accessible than the inferential one. This is a crucial point for explaining the consequence-cause order in causal discourses, though it must be checked by other experiments, in particular about causal and inferential connectives. Can we for instance predict that the presence of ‘parce que’ will give better results than its absence, or that ‘parce que’ will trigger a better treatment of causal connection than ‘donc’?

Conclusion

In this paper, I tried to give a positive answer to the causality puzzle. I gave three argument in favor of the consequence-cause order: the semantic structure of causative constructions in French and English, the distribution of causal, inferential and temporal connectives and the result of experiments (reading time) on the cause-consequence and consequence-cause orders. The interesting point is that all these data converge, and give a rather precise outline of the type of causal model natural languages are shaped for.

References

Ahn W. & Nosek B. (1998). "Heuristics used in reasoning with multiple causes and effects". Proceedings of the 20th Annual Conference of the Cognitive Science Society, Mahwah (NJ), Erlbaum, 24-29.
Blochowiak et al. (2006). "Le projet causalité: analyses quantitatives et qualitatives d’un pré-test". Cahiers de linguistique française 27, to appear, clf.unige.ch.
Jackendoff R. (1990), Semantic Structures, Cambridge (MA), MIT Press.
Kayne R.S. (1975). French Syntax, Cambridge (MA), The MIT Press.
Levin B. & Rappaport Hovav M. (1995). Unaccusativity. At the Syntax-Lexical Interface, Cambridge (MA), MIT Press.
Moeschler, J. (2003). "Causality, lexicon, and discourse meaning". Rivista di lingusitica 15/2: pp. 277-303.
Moeschler J.Chevallier C., Castelain T., Van der Henst J.B.,Tapiero I.(2006). "Le raisonnement causal. De la pragmatique du discours à la pragmatique expérimentale". Cahiers de linguistique française 27, to appear, clf.unige.ch.
Pustejovky J. (1995). Generative Lexicon, Cambridge, The MIT Press.

Discussion

Neurological perspective
Chris Lofting
Apr 3, 2006 15:21 UT

A trait in the neurology appears to be the conservation of energy. In this process it is more efficient to transmit the general to then be followed by the particulars. This means that once the general has been transmitted, specialist elements applicable to the general are all that is needed to transmit and so 'update' the general, refine it into a 'complete' form. This methodology is energy conserving overall.

Thus the transmission of sensory data as AM (amplitude modulation) and its conversion into FM (frequency modulatation and so a PULSE form) implies a pulse train working from general to particular. That 'general' would be in the form of the 'whole' and so the beginning and ending of X prior to the differentiations of 'what did what to whom and when'.

Since the brain appears to be focused on self-referencing so these traits of the neurology will filter up into our basic cognitive processes and seed our grammars etc. (as does the basic neural dynamics of processing patterns of differentiating/integrating allow for the development of specialisations such as linguistics where the general dichotomy of differentiate/integrate is relabelled as noun/verb dynamics. These universals allow for LOCAL contexts to give us the variations of languages we witness in our species)

Chris.

local contextes?
Jacques Moeschler
Apr 9, 2006 14:12 UT

Chris's contribution is a very suggestive and general comment on neurophysiological processes. If I clearly understand Chris's position, it implies that language viariations and specific linguistic traits are caused by local contexts. The question is the follwing: is it possible to consider what linguists would define as a distributional generalisation (the consequence-cause order) as a local context?

Jacques Moeschler

Outdated scenes, acorns and oaks.
Teresa Bejarano
Apr 5, 2006 16:58 UT

Moeschler's paper is a truly though-provoking one. Sincerely, thank you: This is the main message. Here are the provoked thoughts. /// When a causal event is described, the causal event is necessarily previous to the description. In other words, when a causal event is being described, the causal event itself is an already outdated scene. At that moment, the current scene is made up of just the consequence or resulting state. Certainly, perceptive updating is the necessary and adaptive rule. Behaviour must be guided by current, non-outdated, contents and beliefs. However, we are able to rescue past beliefs (cf. the 'theory of mind') and outdated scenes. Causal connectives offer a good resource in order to accomplish the second task. Causal connectives as 'because' ('parce que', 'porque') -"the optimal candidates for causal readings"- give to the outdated scene a role in the description of the updated scene. I have put 'outdated scene' where Moeschler puts 'cause.' Is it a correct substitution? I cannot offer any linguistic argument. /// Anyway, let us focus on two examples of Jackendoff's. (1) * "An acorn grew into each oak" is not a grammatical sentence, but (2) "An oak arose from each acorn" is a totally grammatical one (Jackendoff, 2002, p. 85). Given that "Topic is outside the scope of quantifiers in the Comment" (ibidem, p. 416), neither "an acorn" in (1) nor "an oak" in (2) can serve as a Topic. In conclusion, 'an acorn' in (1) is not a Topic. But we can reach a contrary conclusion. Let us see it. In relation to the oak, the original acorn is an outdated scene. If an outdated content occupies the initial position in an utterance (i.e., the position that is the adequate one for Topics), then the outdated content will be required to be the Topic. This requirement –I suggest- arises because there is a similarity between Topics and outdated contents. Outdated contents must be updated, and, analogously, Topic must be modified by the Comment. In short, 'an acorn' in (1) receives simultaneously two contradictory requirements –be and not to be the Topic-. This is why (1) is not a grammatical sentence. With regard to these examples, we could say that the order 'outdated scene–updated scene' is not the preferred order. But we could also say that the order 'cause–consequence' is not the preferred one. As I openly admitted, I cannot offer any argument in favour of my substitution. /// But let us move away from that question, and let us focus on the inferential reading. "Marie a trop mangé, parce qu´elle est malade". The inferential reading can be interpreted as a metapragmatically causal reading. It would be a case of meta-speech. This is an old suggestion. "Je dis

(que Marie a trop mangé), parce qu' elle est malade". Then, the order 'consequence (my speech act)–cause (piece of evidence)' would invade the inferential reading.

are outdated scenes always causes?
Jacques Moeschler
Apr 9, 2006 14:14 UT

First of all, thanks a lot for this very interesting comment. I would like to contribute to the discussion by distinguishing to aspect of the issue: the first is the very suggestive idea to give an internal and temporal definition of what a cause is, that is an outdated scene. The second point is the parallelism of the updated - outdated scenes and the topic-comment order in a sentence. 1. Teresa's question "I have put 'outdated scenes' where Moeschler puts 'cause'. Is it a correct substitution?" receives a positive answer. Nevertheless, if the equivalence were complete, Teresa's argument could be used not only for causaul discourses, but also for temporal ones. Unfortunately, it is not the case, because the natural order for temporal discourses is temporal order. So causal discourses have an extra property, which is not temporal, but causal. This what makes the experimental dimension of our work interesting if not relevant: this "causal" property is just what we try to define precisely. One hypothesis is that cause is linked to strength, and not only to association. 2. I think the paralellisme between the consequence-cause order and the topic-comment order is very stimulating. The main argument we can give is that what has to be explained (by a causal relation) is what is known, and that the explanation itself should be a new information. Now can we reduce causal or explanation relation to informativeness? I would prefer suggesting a relevance-based account: an explanation is relevant in as much it produces a contextual effect (implication of a new assumption, strengthening or eradiction of an old information) balancing cognitive processing efforts, which means that the causal connection should be accessible. Here again, the information VS relevance explanation should be carefully tested, for example with experimental methods. Finally, I would say that I totally agree with the metapragmatic analysis of the inferential reading. Thanks a lot for the precise consequence (S's speech act)-cause (piece of evidence) analysis.

By the cause of proper linguistics
Robert Stonjek
Apr 11, 2006 2:13 UT

'Because' is not merely a "causal connective." The word 'because' is really be-cause, and 'cause' is the issue. The word 'Because', then, directly addresses the issue of causation.

The word was originally 'by-cause' as in "**by** what **cause**". The word was and is, quite simply, a declaration of causation eg he rested **by** the **cause** of tiredness – he rested **because** he was tired.

An early quote helps us to focus in on this linguistic causative probe: c1386 Frankl T Chaucer "By cause that he was hire Neighebour." - (because he hired a neighbour). c1486 Bk. St. Albans Diijb, "Theis be not enlured ... by cause that thay be so ponderowse." [Quotes from the Oxford English Dictionary]

When we ask after the causative agent of an event, we are also asking "**by** what **cause**?" The answer is "**by** the **cause** of" or "**by cause** of" or "**because** of".

To be strictly correct, 'causation' refers to a causative agent and the possible consequences of it. Becausation refers to the result of a causative agent and probes back from there in search of that agent. The entire causation issue concentrates on finding the causative agent of some current condition, but if we follow strict denotation based on the historical roots of the words used, we are actually asking about "by-causation" (**by** what **cause** did this happen) and so becausation is the correct word. To study the consequences of any causative agent we are studying 'causation.'

Well, it's too late to correct the semantics now. But we do still have the word 'because'. Let's at least preserve its original and proper meaning – after all, it is, or at least should be, one of the key words in understanding the issue of causation generally.

Kind Regards, Robert Karl Stonjek

Causal Maps and Bayes Nets. A cognitive and computational account of causal learning and therory formation

Alison Gopnik (Professor of Psychology, University of California, Berkeley)

(Date of publication: 5 December 2006)

Abstract: We have proposed that children's intuitive theories are 'Causal maps" , coherent representations of the causal relations among events. The Bayes net formalism is a computational way of representing and learning causal maps. Other animals appear to have an understanding that their own interventions can have causal effects,(as in operant conditioning) and independently can detect correlations among events in the world that signal causality (as in classical conditioning). However, they do not seem to put these two kinds of knowledge together to form causal maps.

Résumé : Nous proposons que les théories intuitives des enfants sont des "cartes causales", des représentations cohérentes des relations causales entre les évènements. Le formalisme en réseau de Bayes est une façon computationnelle de représenter et d'apprendre des cartes causales. Les autres animaux semblent comprendre que leurs propres interventions peuvent avoir des effets causaux (comme dans le conditionnement opérant) et indépendamment peuvent détecter des corrélations entre des évènements dans le monde qui indiquent la causalité (comme dans le conditionnement classique). Ils ne semblent pas cependant assembler ces deux sortes de connaissance pour former des cartes causales.