Is autism a sensitivity to outliers?

Paul F. Harrison

Monash University, Australia

Abstract

Autistic people are unusually sensitive to the unusual. Furthermore, their own behavior lacks spontaneous outliers. Autistic people stick to strict routines and their speech is flat and monotonous, or alternatively their behavior and speech is *too* variable but still lacks outliers. This absence may be what gives rise to an impression of "oddness". In this paper, a model of sensitivity to outliers in terms of the alpha-stable family of distributions is developed. In particular the Gaussian distribution, which is sensitive to outliers, is contrasted with the Cauchy distribution, which is not. This is used to conjecture the origins of sensory hyper-sensitivities, hearing impairments in noisy environments, pedanticism, difficulty understanding metaphors, "weak central coherence", unusual interests, impairments in understanding emotional cues, and communicative deficits. A simple random walk model of behavior is also developed, with autistic people characterized by an absence of spontaneous outlier steps.

Key words: autism, mathematical psychology, statistical probability, classification (cognitive process)
PsycINFO classification: 3250

1 Introduction

This paper proposes the theory that autism is the result of a sensitivity to outliers. Autistic people are unusually sensitive to the unusual, and are either unduly fascinated or distressed by it. Furthermore, their own behavior lacks spontaneous outliers. Autistic people stick to strict routines and their speech is flat and monotonous, or alternatively their behavior and speech is *too* variable but still lacks outliers. A mathematical definition of this sensitivity to outliers will allow the consequences of this theory to be elaborated precisely.

Email address: `pfh@logarithmic.net` (Paul F. Harrison).

The Gaussian distribution is an example of a probability distribution that is sensitive to outliers. The Gaussian distribution is used for a very wide range of applications, its use being justified by the Central Limit Theorem. This states that the sum of a set of independent identically distributed random variables of finite variance will always tend to the Gaussian distribution as the number of variables added together goes to infinity.

The problems with Gaussian statistics are well known. A single outlier may skew estimates of the mean and standard deviation by an arbitrarily large amount, so that a few outliers can completely alter a result.

There exists a generalization of the Central Limit Theorem, called the Generalized Central Limit Theorem. This states that the sum of a set of independent identically distributed random variables will always tend to a distribution in the Lévy alpha-stable family of distributions as the number of variables added together goes to infinity. The Lévy alpha-stable family of distributions includes the Gaussian distribution, but also includes more robust distributions with heavier tails and non-finite variance. Use of this family of distributions has in the past been limited by their mathematical complexity, and the lack of closed-form expressions for all but a few members of the family. With the availability of increased computing power, these distributions are increasing in popularity (see e.g. Nikias and Shao, 1995).

In this paper, discussion will be limited to the symmetric alpha-stable distributions (there also exist skewed alpha-stable distributions, with one tail larger than the other). The symmetric alpha-stable distributions form a spectrum of distributions parameterized by the "characteristic exponent", α, which can range from 0 to 2.

The $\alpha = 2$ distribution is the Gaussian distribution. The $\alpha = 1$ distribution is the Cauchy distribution. These are the only known closed-form symmetric alpha-stable distributions. Under the usual parameterization of alpha-stable distributions, the "standard form" of the Gaussian distribution is (Nikias and Shao, 1995, p.32):

$$\frac{1}{2\sqrt{\pi}}e^{-x^2/4} \tag{1}$$

and the standard form of the Cauchy distribution is:

$$\frac{1}{\pi}\frac{1}{1+x^2} \tag{2}$$

Standard form distributions may be shifted and scaled in order to fit data. The scale parameter is called the "dispersion", denoted γ. The location parameter
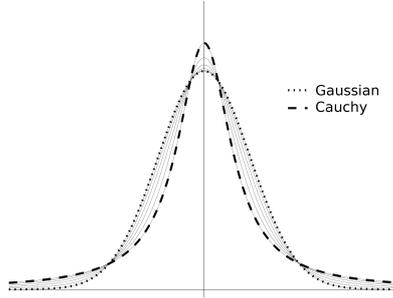
Fig. 1. Symmetric alpha-stable distributions for $1 \leq \alpha \leq 2$.

is denoted δ. δ and γ are analogous to the mean and standard deviation of the Gaussian distribution respectively. If s(x) is a standard form distribution, then its scaled and shifted version is:

$$\frac{\text{s}\left(\frac{x-\delta}{\gamma}\right)}{\gamma} \tag{3}$$

Consider the problem of fitting a distribution to a sequence of observations, that is, estimating γ and δ. Fitting a Gaussian distribution to Gaussian data, or a Cauchy distribution to Cauchy data will cause no problems. Furthermore, fitting a Cauchy distribution to Gaussian data causes no problems: the estimate may converge fractionally slower because the Cauchy distribution doesn't give proper weight to extreme values, but will converge. However, attempting to fit a Gaussian distribution to a Cauchy noise source fails. As more and more data points are considered, the estimated standard deviation tends to infinity, and the estimated mean does not converge to any single value.

Distributions with lower values of α have progressively heavier tails. With the exception of the Gaussian distribution, the tails of alpha-stable distributions fall off algebraically, as $x^{-\alpha-1}$. The tails of the Gaussian distribution fall off extremely rapidly, as e^{-x^2}. Standard form symmetric alpha-stable distributions for α ranging between 1 and 2 are shown in Figure 1.

It is theorized in this paper that autistic people model the world using alpha-stable distributions with high values of α, whereas normal people model the world with alpha-stable distributions having lower values of α. The Gaussian distribution represents the extreme end of the autistic spectrum. The Cauchy distribution will be used to represent normality. Comparison of these two known closed-form symmetric alpha-stable distributions will be used to
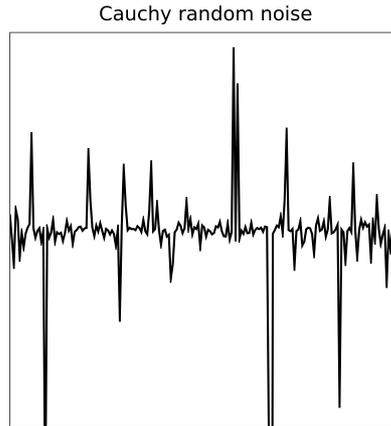
Fig. 2. Cauchy random noise.

attempt to explain the various features of autism.

2   Sensory hyper-sensitivity

Autistic people commonly report unusually high sensitivity to certain stimuli (O'Neill and Jones, 1997; Jones et al., 2003). This sensory hyper-sensitivity will be discussed first, as it is the most direct application of the theory.

Suppose that a person models the world using a Gaussian distribution. Such a person will overestimate the intensity of a stream of sensation containing outliers, as such outliers will excessively inflate a Gaussian estimate of its overall magnitude (i.e. the standard deviation). Examples of such sensations are crackling noises ("impulsive noise", as in Figure 2) and the feel of a prickly material on skin. Abrupt once-off noises such as balloons popping—outliers of loudness in the current aural environment—will also be unduly startling. Noises that are unusually loud as compared too all other noises a person has have encountered will also cause excessive distress.

Consider this description by the mother of an autistic child (Kanner, 1943, p.235):

> He had many fears, almost always connected with mechanical noise (meat grinders, vacuum cleaners, street cars, trains, etc.). Usually he winds up with an obsessed interest in the things he was afraid of. Now he is afraid of the shrillness of a dog's barking.

Mechanical sounds are often impulsive (clanking, clattering, etc). Vacuum cleaners do not produce impulsive type noise, but are unusually loud overall.
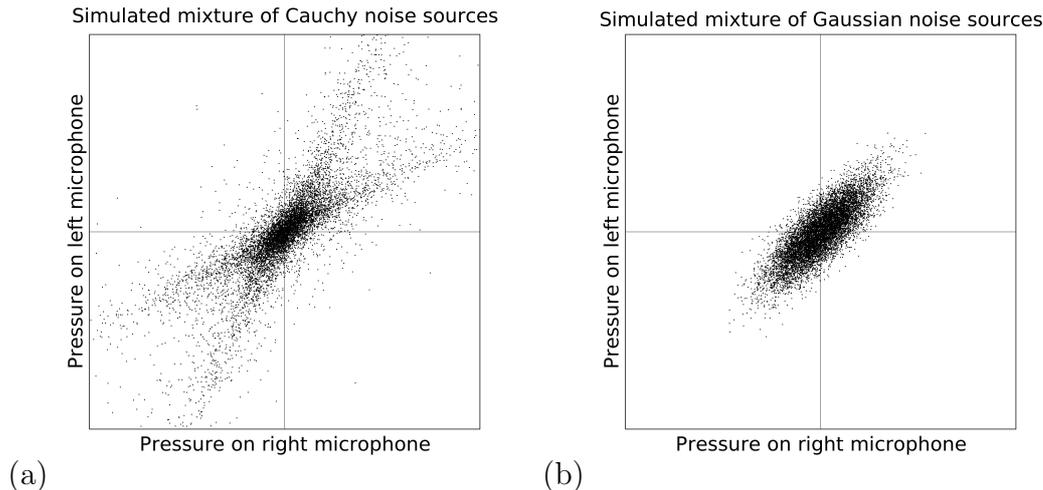
Fig. 3. Examples of the blind source separation problem (scatter plots).

A dog's bark is generally an outlier of loudness in the aural environment it occurs in.

Attwood (1998) describes sensitivity to three types of sound. The first type is abrupt sounds such as a dog barking, a phone ringing, or a balloon popping. Sensitivity to such sounds is as expected by the present theory. The second type is high-pitched continuous sounds such as those produced by electric motors. This second class is more puzzling, it might be that loud sounds of high pitch are simply unusual. The third type is complex or multiple sounds such as might be heard at a social gathering. This third type will be discussed in the next section.

To compensate for their sensitivity to outliers, an autistic person might decrease their overall sensitivity. That is, their expectation of the intensities of sensation they may encounter will be excessively high. Thus they may show a greater tolerance than normal for sensations that do not contain outliers.

A recent review of investigations into sensory dysfunction (Rogers and Ozonoff, 2005) does not list any systematic investigation of differences in response to sensations (such as sounds) having different characteristic exponent, or similar. This is something that may be worthwhile investigating.

3   Hearing deficits

Autistic people frequently show difficulty understanding speech where multiple people are talking at once (Attwood, 1998, p.83). In signal processing, disentangling multiple sound sources is known as the blind source separation problem (or more generally "Independent Components Analysis").

For example, one might wish disentangle two sound sources given input from two microphones. As shown in Figure 3, it is easy to determine the two sources if pressure waves from those sources follow a Cauchy distribution, but impossible if pressure waves follow a Gaussian distribution. Similarly, an attempt to separate sounds (assuming those sounds have $\alpha < 2$) using a Cauchy model will succeed where a Gaussian model will fail and a model with high α will have difficulty. A suitable algorithm for separation of alpha-stable noise sources is described by Sahmoudi et al. (2005).

This is analogous to a person disentangling sound sources by comparing what they are hearing in their left and right ears. It should be noted that in practice a person has more than two channels effectively available to them, as they can decompose what they are hearing into frequency bands, and can also make use of the slight time difference between a sound reaching each ear. However the difficulty remains even where more channels are available.

4   Classification

Let us now consider the effects of a sensitivity to outliers on higher-level cognitive functions. One important way people make sense of the world is by *classification*. A classification may be represented mathematically as a "mixture model". As will be seen, modifying the underlying probability distribution used to represent classes in such models can lead to quite different ways of classifying events and objects.

Suppose we have a set of observations $x_1 \ldots x_l$ of events or objects, each observation being a vector of real-valued properties $x_{i,1} \ldots x_{i,m}$. For example:

- Observations of faces and associated emotions: the position of the edges of the mouth, the amount of creasing around the eyes, the position of the eyebrows, and so on, plus how happy, angry, sad, etc. the person is.
- Observations of various properties of an object, and properties of a word associated with that object.

From this set of observations, we may wish to predict a currently unknown property of a thing given known properties of that thing. This may aid, for example, in:

- Understanding emotional cues: given a person's facial expression, what emotions are they experiencing?
- Language production: given the properties of this object, what word should I use to describe it?
- Language understanding: given the name of an object, what properties

should I expect it to have?

One way to achieve this would be to split the observations into classes. Each class would have a certain center point u_i $(1 \leq i \leq n)$ and a certain amount of spread about that center in each dimension, specified by a probability density function $f_{i,j}$ (for simplicity, the possibility that properties may be correlated within a class will be ignored). From this we may calculate the probability density function of each class:

$$\mathrm{F}_i(y) = \prod_{j=1}^{m} \mathrm{f}_{i,j}(y_j - u_{i,j}) \tag{4}$$

Given an overall likelihood of each class occurring p_i, the probability density functions of each class may be added together to give an overall probability density function:

$$\mathrm{F}(y) = \sum_{i=1}^{n} p_i \mathrm{F}_i(y) \tag{5}$$

This is called a mixture model.

From this, we may calculate the likelihood that a given point y belongs to a certain class:

$$\mathrm{P}(y \in \text{Class } i) = \frac{p_i \mathrm{F}_i(y)}{\mathrm{F}(y)} \tag{6}$$

If only some of the properties of the observation are known, only those dimensions would be used in these calculations. Once it has been determined which classes the present observation might be a member of, these classes can be used to predict the unknown properties of that observation.[1]

The next two sections will discuss how the properties of a mixture model change if a different underlying distribution is used.

4.1 Two classes, one property

Let us first consider the case of two classes being used to classify observations with a single observed property $(n = 2, m = 1)$. Examples of this case

---

[1] A possible simplification to this model would be to omit the aggregation of observations into classes, and to instead say that each individual observation represents the center point of a "class", each class having the same size and shape of distribution about its center. Such a model would be more like memory than classification.
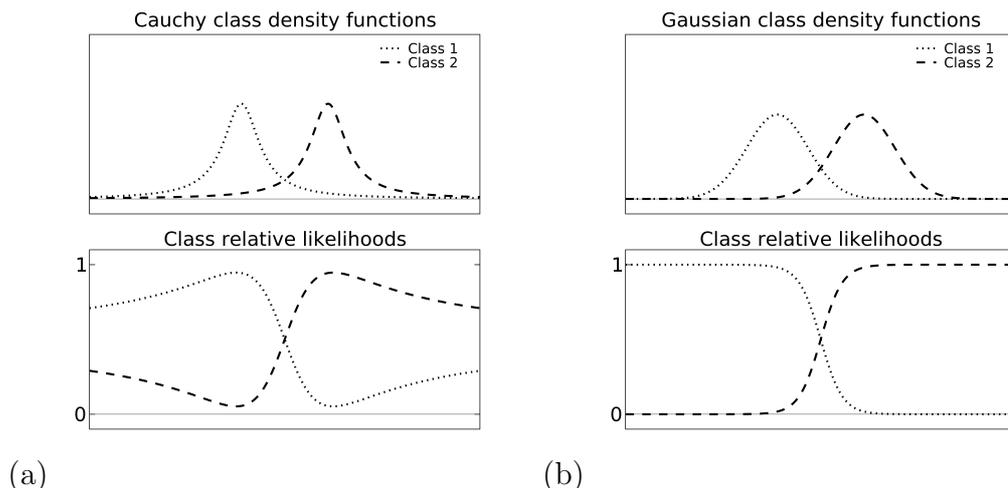
Fig. 4. Mixture models with two classes of equal size.

are shown in Figure 4. These simple examples already show a key difference between classification in the Cauchy and Gaussian cases. When Gaussian distributions are used, the class relative likelihood curves follow a sigma curve. [2] Such curves are well suited to binary, if-then, style type reasoning. Outlier data is very strongly assigned to one or the other of the classes. When Cauchy distributions are used, however, outlier observations are not so strongly assigned to one or the other of the classes. The strongest assignment occurs near the center point of each of the classes. This is a common-sense approach to classification: if something is unlike anything you have seen before, it is best to keep an open mind about it.

Common sense is a trait that autistic people lack. They do not realize that the usual rules may not apply in unusual situations. This is consistent with the use of a Gaussian classifier.

A teacher writes of students with Asperger syndrome (Gill, 2003, p.198):

> They cannot take into account extenuating or changing circumstances. A rule is a rule, no matter what! They operate in black-and-white terms, whereas we are so often in the gray.

On the other hand, this autistic tendency to strong classification will aid in rote memory skills. If only one class is relevant, other classes will not interfere with recall of the properties of that class. A class here would be a memorized fact, or perhaps a link in a chain that can be recalled sequentially. (Such a sequence may also drive the enactment of a fixed routine, where at every step in the routine there is only one possible next step.) This skill would only be
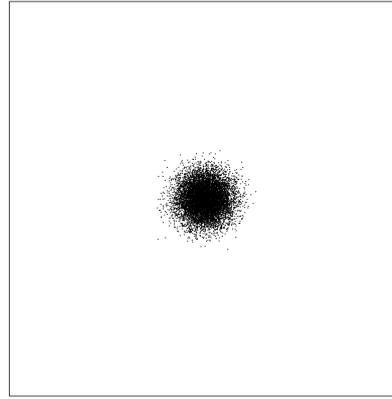
---

[2] This curve is in fact the logistic function commonly used in Artificial Neural Networks.

Random points with independant Cauchy
distribution in the x and y dimensions
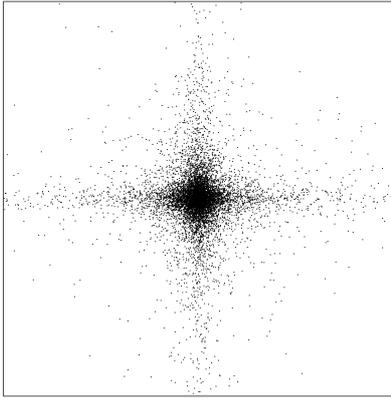
Random points with independant Gaussian
distribution in the x and y dimensions

(a)   (b)

Fig. 5. Random points sampled from (a) a Cauchy based class, (b) a Gaussian based class.

present if the person tends to use classes that do not overlap to a large degree (have small dispersion).

4.2 Many classes, many properties

Let us now consider the two-dimensional case (and higher dimensional cases in general). If we maintain the condition that properties of a class are not correlated with one another,[3] then Cauchy distributed classes take on a cross shape, while Gaussian classes are circular or elliptical (Figure 5). As before, the Cauchy classes have very much heavier tails than the Gaussian classes.

These differences combine to give very different judgments of the relative likelihoods of different points belonging to each of a set of classes. This is illustrated in Figure 6. As in the one dimensional case, the Gaussian classifier has far less areas of ambiguity, where multiple classes may be relevant. This is especially the case for outlier observations. In the Gaussian case outliers are assigned very strongly to their nearest class, whereas in the Cauchy case the further out one goes the less strongly a point will be assigned to a specific class. A consequence of such a lack of ambiguity would be pedantic use of language.

However a further difference is present in the two dimensional case that was not present in the one dimensional case. This derives from the cross shape of the Cauchy classes (Figure 5a), the effect of which can be seen in Figure 6a. A normal person, using Cauchy classes, will be able to see an analogy to an

---

[3] If this does not hold, it may be possible to linearly re-parameterize the observation space so that it does. This is always possible in the Gaussian case, but only sometimes possible in other cases.

Cauchy class relative likelihoods   Gaussian class relative likelihoods

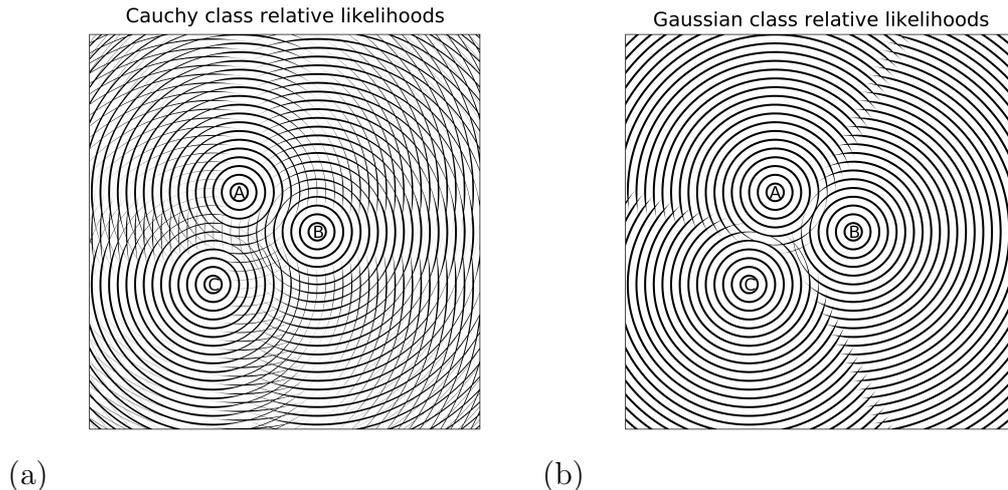(a)   (b)

Fig. 6. Relative likelihood of each point in a space of possible observations belonging to each of a set of classes. The thickness of each circular ripple at each point indicates the likelihood of that point belonging to the class at the center of the ripple.

existing class even if there is a large mismatch in one of the parameters. An autistic person will not, requiring all parameters to match reasonably well before assigning an observation to a class.

This difference provides an explanation of the often noted difficulty autistic people have with metaphors and spontaneous pretend play. For example, a toy car is similar to a real car in many respects, but differs greatly in size. A normal child would see the similarities, and suppose the toy to be an instance of class "car". An autistic child would see it simply as an unusually shaped object. Similarly, a normal child might initially confuse a doll with a real person but an autistic child would not.

This difference becomes particularly important in the case of contextual properties, properties not of an object itself but the kinds of situations in which it may occur. For example, realizing that one sees a certain teacher only at school and not elsewhere provides a small amount of extra information with which to recognize that teacher. However, it may be the case that one *usually* sees the teacher at school but *sometimes* sees them elsewhere. A Gaussian based model may make one of two types of error in this situation:

- Set the dispersion of the distribution too low for the contextual property and thus *fail to generalize* the class to other situations where it would be appropriate.
- Set the dispersion too high for the contextual property and thus *ignore contextual clues*.

Consider also classification where spatial or temporal location is included in the properties of an object or event, such that nearby objects or events are

grouped together. Cauchy classification groupings will have fuzzy spatiotemporal boundaries, whereas in the Gaussian case the grouping will be quite sharp edged. This will affect what other objects or events are considered relevant when predicting unknown properties of an object or event. Again the autistic use of context will be different to normal.

"James", an autistic author on the internet, writes (Jones et al., 2003):

> Most people have a mind like a flashlight, with an area of high focus, and a larger area of partial awareness; my mind is more like a laser pointer, that highlights only a single small dot.

These contextual considerations are highly reminiscent of the "weak central coherence" theory of autism (Frith and Happé, 1994).

5   Specialized interests

One of the key features of autism is an odd pattern of interests. Such interests tend to be both unusual in their choice of subject matter and of sustained unusual intensity. How might this be explained?

Consider the problem of adjusting an existing mixture model to account for some new data points. If the data points are sufficiently surprising then it may be necessary to create a whole new class to accommodate them. Even if a new class is not required, the new data points may require adjustment of certain model parameters in order to be assimilated—such data points might be called "influential". It seems reasonable to suppose that a person will seek out and pay special attention to influential observations (see e.g. Relevance Theory, Wilson and Sperber, 2004).

Let us use the information content of a data point given the current mixture model as a measure of the surprise it causes. The information content of an event is the negative logarithm of its probability, given the current model (Shannon, 1948). Thus, the less likely an event, the more information content it has, and the more surprising it is. Let us further define the influence of a data point to be the derivative of the surprise. These definitions allow us to say that the most likely (Maximum Likelihood) model is the model for which the total surprise caused by the known data is minimized. Also, for each class center point in this Maximum Likelihood model the total of the influence on each property will be zero. [4]

_____
[4] There may be local minima, maxima, and saddle points for which this is true also, and this also does not take into account estimation of the dispersion of the distributions.
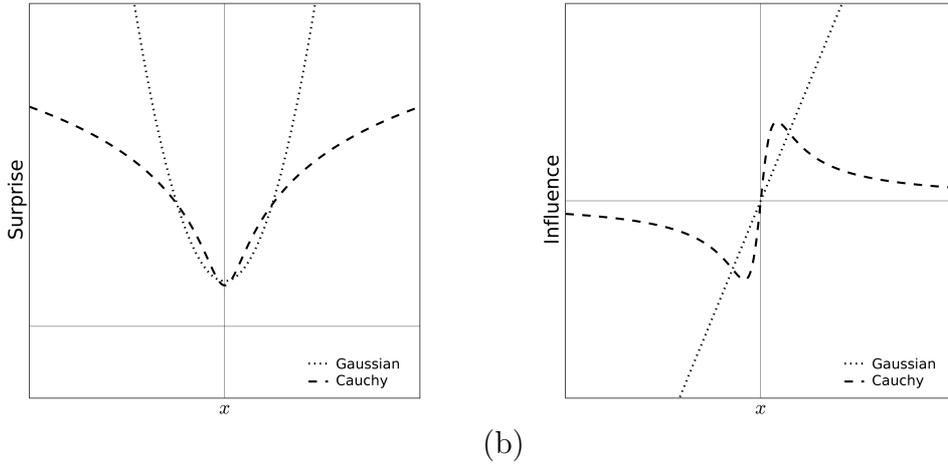
Fig. 7. (a) surprise curves, (b) influence curves.

The information and the influence functions for the standard Gaussian distribution are:

$$\mathrm{I}(x) = \frac{x^2}{4} + \log(2\sqrt{\pi}) \tag{7}$$

$$\mathrm{I}'(x) = \frac{1}{2}x \tag{8}$$

Note that minimizing the information in a Gaussian model is equivalent to finding a least squares fit.[5]

The information and influence functions of the standard Cauchy distribution are:

$$\mathrm{I}(x) = \log(1 + x^2) + \log \pi \tag{9}$$

$$\mathrm{I}'(x) = \frac{2x}{1 + x^2} \tag{10}$$

Both surprise and influence differ markedly between the Gaussian and Cauchy distributions, as illustrated in Figure 7. Normal people will be surprised by outliers, but not intensely so, and will actually be influenced less by outliers than by data points that differ only a little from expected—a common-sense

---

[5] Thus the many machine-learning algorithms that use a least squares criterion might be considered "autistic" under the present theory. This applies also to gradient descent algorithms that update model parameters using a linear influence function, such as certain Artificial Neural Network training algorithms.

attitude. Autistic people remain both intensely surprised and intensely influenced by outliers. Therefore their focus of interest will be on the unusual.

This still leaves the question of why autistic people tend to specialize on one particular unusual topic. One reason for this, as will be argued in the next section, may be that autistic people will not explore occasional areas far outside their focus of interest as normal people would. Thus their interests may widen only very gradually over time.

Another reason might be that if all but a few observations of some class of occurrence could be explained by a class with low dispersion, an autistic person might be driven to try to find some explanation for those outliers. One might say that a few surprising observations *falsify* an otherwise good theory. If the outliers can be explained, the bulk of observations will also be massively better explained. An autistic child might then act like a little scientist, seeking an explanation for those outliers. Such an explanation might involve the creation of a new class with distinct properties, or a re-parameterization of the observation space. Such an explanation would be far less useful to a normal person with Cauchy surprise curves.

To summarize, the odd interests of autistic people may be explained by an initial tendency to be interested in outliers, a deficit of exploratory behavior, and possibly a drive to find an explanation for outliers within a subject of interest.

6   Deficits of exploratory behavior

Lawson (2003, p.183) writes:

> Transition is an amazing part of everyday life. It is so common that most individuals seem to hardly notice it. ... Transition and autism are like enemies. They are foreign to one another because they represent opposing abilities. Transition says "it's time to move on," and it assumes that one is ready, able, and willing. Autism says "I have to stay here because here is all I know."

It is proposed that a random walk is a reasonable though simple model of exploratory behavior. This walk might be through a physical space, a space of possible solutions to some problem, or a space of possible foci of attention.

If one adds a concept of terrain, such that the person is more likely to walk downhill than uphill, or in a possibility space will rarely take a step into an unlikely or uninteresting area, one has the Metropolis-Hastings algorithm for
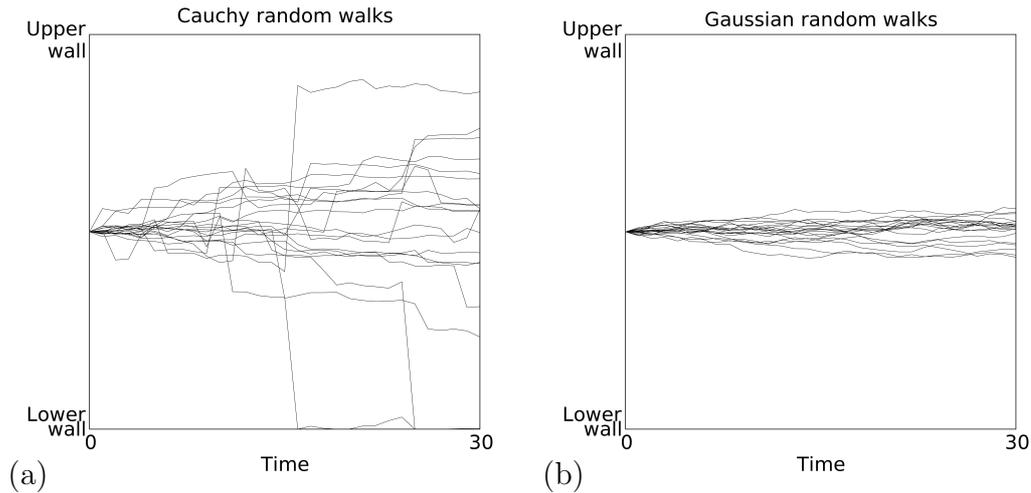
Fig. 8. Examples of random walks. Each walk contains 30 steps.

correctly sampling from a probability distribution (see Robert and Casella, 2004). A variant of this algorithm, Simulated Annealing, allows the location of the global minimum of a function that contains local minima.

Suppose that a person takes Gaussian distributed steps. Some examples of the random walks that might result are shown in Figure 8b. In general, the area explored only increases as $\sqrt{t}$ over time. Contrast this with the random walks that result from Cauchy distributed steps, examples of which are shown in Figure 8a. The Cauchy case is characterized by spans of relative inaction punctuated by abrupt transitions. These abrupt transitions are absent in the Gaussian case. In general, a Cauchy random walk covers an area that expands linearly over time (see Mandelbrot, 1983, p.368). Thus this model would seem to predict that a normal person will explore a space of possibilities more quickly than would an autistic person. An autistic person would be more *perseverative.*

In an autistic person, a symptom of such a deficit of exploration would be slowness in generalizing a specific successful strategy. Further, attempts to force large changes may cause distress, as the person will assign such large changes very little likelihood of success.

Courchesne (Courchesne et al., 1994) presents an argument that a difficulty in changing the focus of attention can be used to explain difficulties autistic people have in shared attention. They are unable to keep up with the large leaps of attention normal people make in the course of normal interaction. This may also derail autistic children from the normal social development track.

Alternatively, an autistic person might compensate by making larger though still Gaussian distributed steps (Figure 9). The price of such a choice would be an inability to make fine adjustments. A person who took this route might
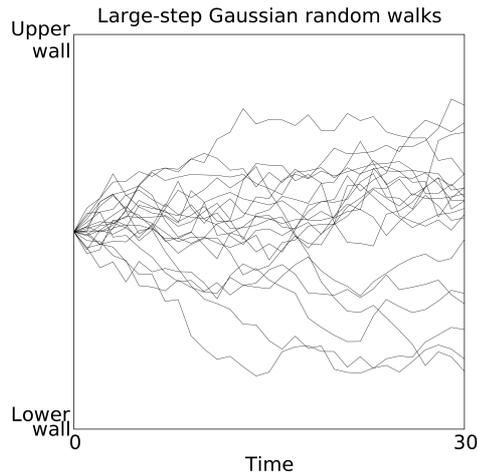
14

Fig. 9. Gaussian random walk with larger step size. While covering the same ground as the Cauchy random walk, there are no abrupt leaps in these walks.

be clumsy, or might have difficulty fixing their attention on a single topic. In social terms, the person would be unable to maintain their focus on the same object as the person they are interacting with, again making social interaction difficult.

Both of these alternatives may be seen in the one child under different circumstances. Asperger (1944, p.60) writes of Ernst:

> He was 'very precise': certain things always had to be in the same place, and certain events always had to happen in the same manner, or he would make a big scene. There was an interesting contradiction here: in certain matters he was particularly messy and could not get used to things being done in an orderly fashion, but in others he was pedantic to the point of obsession.

This apparent contradiction poses no problem to the present theory. The rigid, pedantic quality has to do with the characteristic exponent α (tail shape) of the step or model distribution, the messiness or precision to do with its dispersion γ. It is possible to be pedantically messy, to tolerate a certain amount of mess but no more. α is fixed, but γ may vary to suit circumstances, and may even be something a person could learn to control.

These predicted differences between autistic and normal behavior provide a means of rather directly testing the theory presented in this paper. Any such tests should concentrate on internally driven behavior, as with externally driven behavior there is always the risk of provoking an extreme reaction by presenting something unexpected. Some aspects of behavior that might be investigated are:

15

- Duration, pitch, and volume of syllables in speech.
- Length of phrases, sentences, and paragraphs in writing.
- Movement of the pencil while drawing.
- Patterns of eye movement when presented with a uniformly textured stimulus image.

The following quote from a typical case study in Asperger Syndrome (Manjiviona, 2003, p.66) displays several of the features that it is proposed may be quantifiable in terms of alpha-stable distributions (emphasis has been added):

> He used very precise language and tended to be very literal in his understanding. He had *difficulty modulating the volume of his voice* and talked loudly, *often in a monotone.* He used repetitive language and had difficulty sustaining a reciprocal conversation with others as he focused on delivering lengthy monologues on favorite topics, oblivious to the responses of his listener. He was capable of making eye contact but often had a *flat, deadpan, expressionless look, with little variation in eye gaze.*

Asperger's (1944, p.42) description of Fritz is consistent with an autistic person who has compensated by using a larger step size (emphasis has been added):

> When somebody was talking to him he did not enter into the sort of eye contact which would normally be fundamental to conversation. *He darted short 'peripheral' looks and glanced at both people and objects only fleetingly.* It was 'as if he wasn't there'. The same impression could be gained of his voice, which was high and thin and sounded far away. The normal speech melody, the natural flow of speech, was missing. Most of the time, he spoke very slowly, dragging out certain words for an exceptionally long time. *He also showed increased modulation so that his speech was often sing-song.*

Here, the presence of exceptionally long words might be taken as evidence against the hypothesis, but might alternatively reflect a steady enunciation of each syllable in long words. The description of the tone as sing-song indicates the presence of large variation in tone, but also perhaps that such variation flowed as in song instead of making occasional leaps. That is, as in Figure 9, and not as in Figure 8a.

An absence of eye contact is often described in the literature as a deficit specifically limited to person-to-person contact but, interestingly, Asperger's description of gaze makes no such restriction. Asperger's description of an overall absence of firmly fixed gaze is instead more consistent with a Gaussian random walk (Asperger, 1944, p.69):

> Therefore, the gaze goes past the other person or, at most, touches them incidentally in passing. However, autistic children do not look with a firmly fixed gaze at anything, but rather, seem to perceive mainly with their pe-

ripheral field of vision. Thus, it is occasionally revealed that they have perceived and processed a surprisingly large amount of the world around them.

Tantam's description of a 19-year-old with Asperger syndrome, Robert, appears to corroborate this (Tantam, 1991, p.151):

> His voice is also monotonous and he does not use expressive gestures. His gaze is roving, barely resting on one's eyes when one talks to him, but there is no obvious avoidance.

A further effect emerges when the random walk is over non-flat terrain, such that it is more likely to take steps downhill than uphill. Such a walk may be used to find global minimum of a function, such as the total surprise metric described earlier. This is called Simulated Annealing. Szu (1987) found that use of Cauchy distributed steps in Simulated Annealing gave the algorithm the ability to escape from local minima *exponentially* faster than would be possible with Gaussian distributed steps. Thus an autistic person may more easily get trapped at a local minimum than a normal person.

7   Communicative and social deficits

In the present theory, communicative and social deficits are not the result of some single dysfunction. Instead, these deficits represent the cumulative product of all of the differences discussed above. Any difference is a barrier to communication and social interaction, and autistic people are different in many ways.

The differences of behavior discussed in the previous section—in autistic people's eye gaze, patterns of movement, and manner of speech—provide an immediate disadvantage by giving an impression of "oddness". Furthermore, differences in autistic people's preferred pattern of attention make it hard for them to follow the attentional lead of normal people, making shared attention difficult. For example, when talking they may stay stuck on a topic whilst others have moved on to discussing something else, or alternatively they may have difficulty staying focused on any single topic.

It was seen earlier that use of a different characteristic exponent leads to a different assignment of data points to classes. Optimal class center points may also be different for different characteristic exponents. Thus an autistic person will not be using the same basic conceptual labels as a normal person. One of the most noticeable manifestations of this is the autistic tendency to be pedantic in their use of words, that is, to interpret words in terms of sharp-edged formal definitions (Figure 6b) instead of their usual common-sense

17

and sometimes metaphorical usages (Figure 6a). Thus to properly understand normal minds, an autistic person would have to build a separate theory as to the concepts that normal people use. To make matters worse, a mixture of several Gaussian classes would be required to approximate each Cauchy class. The converse also holds: it would be inappropriate for a normal person interacting with an autistic person to assume that their mind works in the same way (see e.g. Gill, 2003).

At a more basic level, there may be differences in the classification of facial expressions (and other cues) and their association with emotions. Autistic people sometimes have difficulty identifying emotions (Asperger, 1944; Tantam, 1991), and this could be explained by such a difference in classification. For example a normal person might recognize a large grin as happiness, even if the eyes were not crinkled appropriately or the shoulder muscles not relaxed, but an autistic person might only recognize happiness if all parts of the expression fit their model of a happy face. Even if this were not a problem for recognition of true emotions (the facial expressions for which are adopted reflexively and unconsciously), it could pose problems when classifying pretend emotions (deliberate imitation of what the person considers the key features of a facial expression to be). For example, a care-giver frowning in order to scold an autistic child, but also trying to hide their amusement at the child's behavior, might give signals that were confusing to the child. An autistic person might also have difficulty adopting the facial expressions a normal person would associate with different emotions.

The following description of Fritz by Asperger (1944, p.46) is evidence of this possible distinction between recognition of true and pretend emotion:

> Now, a word about the boy's relation to people. At first glance, it seemed as if these did not exist or existed only in a negative sense, in mischief and aggression. This, however, was not quite true. Again, accidentally, on rare occasions, he showed that he knew intuitively, and indeed unfailingly which person really meant well by him, and would even reciprocate at times. For instance, he would declare that he loved his teacher on the ward, and now and then hugged a nurse in a rare wave of affection.

And more generally (Asperger, 1944, p.49):

> These children often show a surprising sensitivity to the personality of the teacher. However difficult they are even under optimal conditions, they can be guided and taught, but only by those who give them true understanding and genuine affection...

Frith (2003, p.111) also notes that autistic children are better at recognizing the universal facial expressions, which are expressed reflexively and do not need to be learned, than culturally specific emotions, which must be learned.

Autistic recognition of emotion may also be less robust overall. By a fortunate coincidence, an accidental computational model of such a deficit already exists. Cohen et al. (2003) describe a system for classifying (posed) facial expressions in video recordings. Their system extracts a variety of facial parameters, and uses a classifier essentially the same as the mixture models described in this paper to find the class of emotion with the best fit to these parameters. The emotions the system attempts to detect are anger, disgust, fear, happiness, sadness, and surprise. Cohen et al. compare the results for classifiers of three different forms: classes where each parameter is assumed to be independent of the others and follow a Gaussian distribution, classes where each parameter is assumed to be independent of the others and follow a Cauchy distribution, and a "Gaussian Tree-Augmented Naive Bayes" (TAN) model that makes use of correlations between parameters.

Overall the TAN model was reported to be best at classifying faces. In the absence of a corresponding Cauchy TAN model, this is not relevant to the current discussion. More interesting is the comparison between the Cauchy and Gaussian models. Where the system was trained and then tested on a single person, the Cauchy model only did very marginally better than the Gaussian model. However, when the system was trained on a database of people, then tested on a different person again, the Cauchy model did significantly better than the Gaussian model (the use of a larger database, however, narrowed this difference).

The interpretation of this as a model of autistic emotion recognition is interesting. The "autistic" Gaussian model had special difficulty *generalizing* emotional expressions from one person to another. The Cauchy model did especially well when generalizing from small amounts of data.

8   Comparison to other theories

There are a number of existing theories as to the causes of the various features of autism.

One prominent theory is "mind-blindness". The "mind-blindness" theory explains deficits of socialization, communication, and pretend play in terms of a lack of understanding that others may know, want, feel, or believe things that one does not and vice versa (Baron-Cohen et al., 1985). The present theory provides alternative explanations for many of the phenomena mind-blindness seeks to explain (as discussed in the previous section). Most notably the present theory says that shared attention is impaired because an autistic person has difficulty making the occasional large shifts of attention necessary in order to follow the attention pattern of a normal person, not because autis-

tic people are unable to read attention cues such as gaze. The lack of pretend play is in the present theory explained by an absence of metaphorical-type classification of objects (Section 4.2).

Another theory, that explains certain features of autism that mind-blindness does not explain, is that autistic people have a deficit (or perhaps just difference) in "executive function" (Ozonoff et al., 1991). Executive function is the ability to maintain an appropriate mind-set in order to attain some goal. The specific deficits of executive function autistic people are described as having are rigid and inflexible problem-solving strategies, and a tendency to perseveration. Autistic people sometimes actually perform better than controls in terms of maintenance of problem solving set (Ozonoff et al., 1991; Griffith et al., 1999), an example of perseveration as a useful trait. Under the present theory, this rigidity and perseveration is an instance of autistic people's overall tendency not to make occasional large leaps, a tendency also present in such things as speech intonation and eye gaze. Thus the present theory is compatible with the executive function theory, but views this deficit/difference as part of a larger pattern, and not as a specific dysfunction of an "executive functions module" in the brain. The present theory is potentially valuable in offering a mathematical model of executive function deficits/differences.

The present theory is also compatible with the "weak central coherence" theory of autism (Frith and Happé, 1994). Weak central coherence is a lack of understanding of overall structure of a collection of parts, and therefore also a tendency not to see individual parts in their global context. A way was described in Section 4.2 whereby weak central coherence could arise from sensitivity to outliers in classification. As with the executive function theory, the present theory may have value in offering a mathematical formalization of weak central coherence.

9 Conclusion

In this paper, it has been conjectured that sensitivity to outliers is sufficient to explain many reported features of autism: sensory hypersensitivities, hearing impairments in noisy environments, pedanticism, difficulty understanding metaphors, "weak central coherence", unusual interests, impairments in understanding emotional cues, and communicative deficits. Furthermore, an absence of outliers in behavior explains qualitative "oddness" of behaviors such as eye movement and gestures, odd attention patterns, and difficulty with shared attention.

One important element in the theory is that, while the characteristic exponent α that a person uses is fixed, the overall size of the distribution γ may be varied.

This raises the possibility of teaching autistic people a compensatory skill of consciously adjusting the overall size of distribution to match the situation. This might be used on the one hand to reduce a tendency for preservation of sameness, fixed routines, and circumscribed interests, or on the other hand to reduce attention deficit and anxiety. This is not a skill a normal person would necessarily need.

The theory is at the level of cognition, and says nothing about the proximate biological causes of autism. It is a simple enough change that one may suppose it could result from a single change to some parameter of neural growth or functioning. Such a change would have a pervasive effect on brain functioning, and would not be specific to any one mental faculty. As natural phenomena display variation following a variety of distributions, the ultimate cause of autism may be the evolutionary utility of a gene pool that includes people who model the world using a variety of different characteristic exponents.

Descriptions of autistic people in the literature have been shown consistent with the theory presented. However to validate the theory properly it would be necessary to directly measure the behavior of autistic people and normal controls, and to attempt to fit these measurements to the alpha-stable family of distributions. It is possible that in some cases the necessary data has already been collected in the course of other experiments, and merely needs to be subjected to this novel form of analysis.

References

Asperger, H., 1944. 'Autistic psychopathy' in childhood. In: Frith, U. (Ed.), Autism and Asperger Syndrome. Cambridge University Press.

Attwood, T., 1998. Asperger's Syndrome: A Guide for Parents and Professionals. Jessica Kingsley Publishers.

Baron-Cohen, S., Leslie, A., Frith, U., 1985. Does the autistic child have a "theory of mind"? Cognition 21, 37–46.

Cohen, I., Sebe, N., Garg, A., Chen, L. S., Huang, T. S., 2003. Facial expression recognition from video sequences: temporal and static modelling. Computer Vision and Image Understanding 91, 160–187.

Courchesne, E., Townsend, J., Akshoomoff, N. A., Saitoh, O., Yeung-Courchesne, R., Lincoln, A. J., James, H. W., Haas, R. H., Schreibman, L., Lau, L., 1994. Impairment in shifting attention in autistic and cerebellar patients. Behavioral Neuroscience 108 (5), 848–865.

Frith, U., 2003. Autism: explaining the enigma. Blackwell Publishing.

Frith, U., Happé, F., 1994. Autism: beyond "theory of mind". Cognition 50, 115–132.

Gill, V., 2003. Challenges faced by teachers working with students with As-

perger syndrome. In: Prior, M. (Ed.), Learning and Behavior Problems in Asperger Syndrome. The Guilford Press.

Griffith, E., Pennington, B., Wehner, E., Rogers, S., 1999. Executive function in young children with autism. Child Development 70 (4), 817–832.

Jones, R., Quigney, C., Huws, J., 2003. First-hand accounts of sensory perceptual experiences in autism: a qualitative analysis. Journal of Intellectual and Developmental Disability 28 (2), 112–121.

Kanner, L., 1943. Autistic disturbances of affective contact. Nervous Child 2, 217–250.

Lawson, W., 2003. Remembering school. In: Prior, M. (Ed.), Learning and Behavior Problems in Asperger Syndrome. The Guilford Press.

Mandelbrot, B., 1983. The Fractal Geometry of Nature. W. H. Freeman and Company.

Manjiviona, J., 2003. Assessment of specific learning difficulties. In: Prior, M. (Ed.), Learning and Behavior Problems in Asperger Syndrome. The Guilford Press.

Nikias, C. L., Shao, M., 1995. Signal Processing with Alpha-Stable Distributions and Applications. Wiley Inter-Science.

O'Neill, M., Jones, R., 1997. Sensory-perceptual abnormalities in autism: A case for more research? Journal of Autism and Developmental Disorders 27 (3), 283–293.

Ozonoff, S., Pennington, B., Rogers, S., 1991. Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. Journal of Child Psychology and Psychiatry 32 (7), 1081–1105.

Robert, C. P., Casella, G., 2004. Monte Carlo Statistical Methods. Springer.

Rogers, S. J., Ozonoff, S., 2005. Annotation: What do we know about sensory dysfunction in autism? a critical review of the empirical evidence. Journal of Child Psychology and Psychiatry 46 (3).

Sahmoudi, M., Abed-Meraim, K., Benidir, M., 2005. Blind separation of impulsive alpha-stable sources using minimum dispersion criterion. IEEE Signal Processing Letters 12 (4).

Shannon, C. E., 1948. A mathematical theory of communication. Bell Systems Technology Journal 27, 379–423.

Szu, H., 1987. Fast simulated annealing. Physics Letters A 122 (3), 420–425.

Tantam, D., 1991. Asperger syndrom in adulthood. In: Frith, U. (Ed.), Autism and Asperger Syndrome. Cambridge University Press.

Wilson, D., Sperber, D., 2004. Relevance theory. In: Ward, G., Horn, L. (Eds.), Handbook of Pragmatics. Blackwell Publishers.