The uses of didactic and cognitive training technologies in the behavioral treatment of Autism Spectrum Disorders.
Abstract:

The behavioral interventions currently available for autism are labour-intensive and expensive; technology offers the potential to provide highly adaptive and home-deliverable interventions at a much lower cost. In this paper, we review 28 technology-based interventions in autism, including: DVDs and CD-delivered software that can be used on any home computer, small-scale, commercially available ‘robot companions’ and immersive virtual reality environments. We conclude that technology has considerable potential to reduce some of the crippling costs of interventions in this sector, but that convincing clinical proof of their effectiveness is lacking. There is an urgent need for more involvement from clinicians in the design of improved clinical trials, as well as in aspects of the hardware and software design.

Keywords: Autism, Technology, Behavioral Interventions
Technology and Autism

The uses of didactic and cognitive training technologies in the behavioral treatment of Autism Spectrum Disorders.

Autism Spectrum Disorders and Asperger Syndrome

Autism spectrum disorders (ASD) are neurodevelopmental disorders characterized by deficits in social interaction and communication, as well as stereotyped or repetitive behaviours (APA, 1994). It is a condition marked by extreme heterogeneity (Tager-Flusberg & Joseph, 2003). Asperger syndrome (AS, henceforth incorporated within ASD unless stated) is a diagnosis given to “high-functioning” adults with no language or general cognitive delay, but who nevertheless show communicative or social abnormalities similar to those found in other conditions on the spectrum (Ozonoff et al., 1991). ASD is defined behaviourally (APA, 1994); we are almost certain by now that ASD encompasses a range of mechanistically discrete conditions, although details of this mechanistic heterogeneity remain extremely poorly understood (Herbert, 2005; Herbert & Anderson, 2008; Ronald et al., 2006).

The cost of autism care in the UK alone has recently been estimated at £28 billion ($43.5 billion) per year, with the lifetime cost of caring for a single individual with autism and a learning disability estimated at £4.7 million ($7.3 million) (Foundation for People with Learning Disabilities, 2007). A significant proportion of this is personnel costs, in what is a highly labor-intensive, largely palliative industry (although see eg Dawson et al., 2011).


“you’re so patient, but a computer is even more patient, computers are great in comparison to you. When I was five I could already write and do sums, but nobody noticed because [...] I was scared of people, because I was unable to speak”

(Sellin, 1993)

Insofar as the bewildering heterogeneity of neurological and histochemical abnormalities associated with ASD can combine to cause a single, autistic cognitive phenotype, this is found in the tendency of individuals with ASD to prefer systematizable, rule-based situations to unpredictable situations where empathy is required (Baron-Cohen, 2002; Cohen, 2007). A common favourite is the use of technology, which requires attention to detail, an ability to derive and implement abstract rules, and which is affect-free per se (el Kaliouby et al., 2006; Picard, 2009).

The past two decades have seen an explosion of activity from individuals with ASD using technology as a medium to interact online, with an Autism Liberation Front active on Second Life (ii), countless online autism community groups (eg iii and iv) and online video diaries from individuals with autism that have reached almost a million views(v).

The fact that so many individuals with ASD enjoy technology and technology-mediated interaction makes it a perfect forum for providing interventions; people always perform better in situations in which they feel secure and motivated (Dweck, 1986). Furthermore, there are a number of other practical advantages (Bishop, 2003;
Murray et al., 2005). The heterogeneity of symptom severity in people with ASD can pose practical difficulties for care providers (Myers et al., 2007); technology is perfectly suited to provide individually tailored interventions that are suitable for a wide variety of abilities (Bishop, 2003), with some technologies being developed (eg cheap imitative robots, gaze-contingent environments) that potentially even severely impaired individuals can use, supervised only by a parent or care-giver. Technological interventions also allow the user to work at different speeds and locations, and never lose patience with the endless repetition that many people with ASD desire (Wilkinson et al., 2008). Finally, they are potentially considerably less expensive than traditional sources of care provision, which are labor intensive and consequently extremely costly (Myers et al., 2007).

Overview of Technology Methods

Technology is used in increasingly varied pedagogical contexts, both within the classroom and as a tool for helping us to understand better user motivation (du Boulay, in press; Arroyo & Woolf, 2005; Yannakakis and Hallam, 2006). The technology-enhanced interventions reviewed in this paper use various methods:

The most easily available (eg Golan & Baron-Cohen, 2006; Grynszpan et al., 2008; Tanaka et al., 2010) are CD-delivered software applications that, once developed, cost only several cents per unit to produce, and which can be used by anybody who has a computer at home. Software can automatically email records of a user’s participation after each session (e.g. Tanaka et al., 2010). This allows clinicians to check each
user’s compliance and progress, and adjust the nature of the training set if necessary. Other online features, such as a high-scores website (Tanaka et al., 2010) and online games can also be used to encourage motivation and participation. These software applications have huge potential scalability. For example, the FastForWord program, an auditory perceptual training software package for children with specific language impairment (www.scilearn.com) has been used by over a million students in over 40 countries (Merzenich et al., 1996; Tallal et al., 1996; Temple et al., 2003; although other authors question its effectiveness – Bishop et al., 2006).

Other common methods include Virtual Reality (VR) environments, which range from 3-D environments run using a joystick and a home computer (eg Mitchell et al., 2007) through to fully immersive environments with a headset display, body tracker and three-dimensional hand controls. Large-scale commercial development by computer game developers has recently seen the cost of these items plummet, with headset-mounted displays and motion sensors available for only several hundred dollars per unit (vii).

In robotics, the costs vary considerably depending on the complexity of the application. Of the robots used in the interventions described here, one (from Billard et al., 2007) is commercially available at a cost of $3,000 per unit (viii).

For the purposes of this review, we consider only technology-mediated interactions in which the use of the technology is intended to reduce the requirement for trained clinicians or teachers. Thus, techniques such as video modelling (reviewed in Di
Gennaro Reed et al., 2011) and neurofeedback (eg Kouijzer et al., 2009) are not included in this review.

**Overview of Article Structure**

The 28 studies described in this paper present a number of uses of technology. In part 1, we discuss *didactic technologies* – ie domain-specific tutors designed to target particular deficiencies in emotion and face recognition (1a), language development (1b) and other social skills (1c).

In part 2, we discuss a range of *other technologies* that are more flexible in their potential uses. The technologies we review include virtual reality (2a), gaze-contingent interfaces (2b), robotics (2c), digital play environments (2d) and ‘emotional hearing aids’ (2e).

For the reader’s ease, we have summarised the studies featured in each section in a separate table. We also start each section with a brief, one-paragraph section summary, to allow the reader to skip on if he or she wants.

**Part 1 – Didactic Technologies**

[insert summary table 1 about here]

1a – Emotion and face recognition tutors.

**Summary:**
A number of interventions have provided home-deliverable software or DVD packages that use photos, video clips and text descriptions to teach face recognition and the recognition of emotion from faces. The majority have found that subjects improve within the trained environment, but that these improvements do not generalise to performance of the same task in a ‘real-world’ environment. The reasons for this are discussed. One recent study (Golan et al., 2009) provided DVD packages to younger children (aged 4 to 7) and reported surprisingly strong generalisation in a number of areas.

Discriminating one face from another is a subtle, and highly specialised, perceptual discrimination task (Tarr & Gauthier, 2000). Abnormalities in identifying faces and in recognising the emotions contained on faces have been variously identified not just in autism (Dawson et al., 2005; see also Jemel et al., 2006) but also schizophrenia (Feinberg et al., 1986), depression (Gur et al., 1992), various agnosias including prosopagnosia (Damasio et al., 1982), Down syndrome (Annaz et al., 2009) and even Williams syndrome (Annaz et al., 2009)

Being able to “read” faces and glean information from them is hugely important as a tool for learning during development (Karmiloff-Smith, 1992). For example, gaze following during infancy (ie the tendency to look where other people are looking) has been shown to correlate with subsequent language development (Brooks & Meltzoff, 2005); similar relationships have been observed for learning in other areas (Johnson, 2010).

The interventions reviewed here mostly take the form of home-deliverable tutoring
systems, based on a varying combination of photographic stimuli, movie clip, voice clips and written descriptions. In several cases (e.g. Faja et al., 2008) these are accompanied by explicit instructions on how to recognise faces and emotions.

For example, in the Mind Reading software reported in Golan & Baron-Cohen (2006), 412 emotions are divided into 24 emotion groups and 6 'difficulty' levels. Each emotion is defined and demonstrated using silent films of faces, voice recordings and written examples of situations that evoke this emotion. The user has the choice either of browsing through the emotion library, of taking structured lessons, or of playing one of five different educational games.

A number of studies in this area have been reasonably well controlled and had decent sample sizes (Golan & Baron-Cohen, 2006 – trained N=32; Tanaka et al., 2010 – N=42), and some have also tested how far learning effects generalise to other areas. For example, Golan and Baron-Cohen (2006) include in the pre- and post-testing battery some of the faces and emotions included in the training set, some faces and emotions not included in the intervention, and some clips from feature films.

Almost every study has found that performance of subjects with ASD within the trained environment improves over time. It is striking, however, that three studies (Golan & Baron-Cohen, 2006; Faja et al., 2008; Tanaka et al., 2010) have reported that these improvements fail to generalise to face and emotion recognition tasks outside the specific group of faces and emotions that was included in the training set.
(although Faja et al., 2008 and Tanaka et al., 2010 report some relatively minor changes in face processing strategies). In other words, they have failed tests of distal generalisation (Porayska-Pomsta et al., 2011). The authors are to be commended for the clarity with which they have designed their experiments in order to address this question.

The one study that has reported considerable far-transfer of training effects is Golan et al., 2009 (see also Baron-Cohen et al., 2007), who developed an animated TV series called The Transporters (www.thetransporters.com) in which the faces of human actors are morphed onto trains. (This was in order to encourage and motivate children with autism, who are known to prefer mechanical objects that move in a repetitive and predictable manner). Fifteen 5-minute episodes each focus on a different key emotion or mental state. In a small-scale (trained N=20) evaluation of this package, Golan et al. (2009) reported that 4 to 7-year-old children who had watched at least 3 x 5-minute episodes a day for 4 weeks showed strikingly strong improvements in emotion recognition, even using faces and emotions that were not part of the trained set. One reservation here, though, is that the control groups only took part in pre- and post-testing; since the trained group also received a booklet encouraging parents to spend one-on-one time with their children discussing the emotions contained in each episode, the failure to replicate this in the control group represents a serious limitation of this study.

1b – Language tutors.
Summary: a number of small studies have presented explicit word- or spelling-tutoring software packages for children with autism, featuring computerised tutors and phonological/orthographic feedback. Software packages are potentially ideal for teaching subjects with autism, in that they are well positioned to provide the sorts of frequent repetition that many subjects with autism frequently request. Suitable for one-on-one tutoring, they are also adaptable to a range of abilities and learning speeds. Again, though, most of the studies in this area remain under-powered and poorly controlled.

Language development is one of the core areas of deficiency in ASD (APA, 1994; Happé & Ronald, 2008). There is, though, considerable heterogeneity with regard to the exact nature and severity of the language impairments reported (Kjelgaard & Tager-Flusberg, 2001; Rice et al., 2005). For an excellent, specialized recent review of interventions in this area, we recommend Pennington (2010).

We include three studies that present vocabulary tutors, in which word teaching drills are variously accompanied by object animations (Schlosser & Blischak, 2004) and a computer-animated face that speaks the words as they are presented (Massaro & Bosseler, 2006). Moore and Calvert (2000) found in a small-scale study (trained N=7) that retention of taught words was better for computer-presented words than for those taught in a behavioural treatment drill. Massaro & Bosseler (2006) found, in another small study (trained N=5), that words presented concurrently with a computer-animated face that spoke the words for the child were retained better than words presented alone on screen. Schlosser & Blischak (2004) looked at whether adding synthetic speech output and orthographic feedback improved the efficacy of their
spelling tutor; they found that it did, but did not perform statistical tests on their small sample (trained N=4).

Hetzroni & Tannous (2004) developed a tutor based on structured, turn-based play that involved pressing buttons to demonstrate comprehension of recorded spoken language in mock-ups of “real-world” situations such as ‘having breakfast’. They tested it on 5 children with ASD that had a tendency towards echolalia – repeating spoken words or parts of words, that is a normal developmental phase but which is often excessive and maladaptive in ASD (Wing & Gould, 1979). Even after a relatively small amount of time playing the games (18 10-25-minute sessions) video-coding of the childrens’ behaviour in a free play environment demonstrated a reduced tendency toward echolalia. Again, though, sample sizes are extremely small and the study was uncontrolled, which makes it impossible to draw firm conclusions.

1c – Social skills tutors.

Summary: a variety of methods have been used to train social skills in subjects with ASD. The tasks vary from false belief tasks to social situations such as choosing where to sit in a crowded canteen. Computer-generated human faces have been used, and virtual environments, as well as more traditional software packages. As with the emotion and face recognition tutors, no study has conclusively shown that within-task improvements can generalise to a ‘real-world’ environment.

In a pioneering study, Swettenham (1996) designed animated computer games (based on text and drawn pictures of different social situations) aimed at teaching children with ASD the Sally-Anne false belief task. (The Sally-Anne task is classically used to
assess Theory of Mind, the understanding that others have mental states discrete from their own (Baron-Cohen et al., 1985). Swettenham (1996) measured the performance of children within the trained task, and then (in another good experimental design) tested how far the trained improvements had generalised using a graded variety of computerised and ‘real-world’ false belief tasks. They also found that children with autism were able to improve within the training paradigm, but that these improvements failed to transfer to more ‘real-world’ measures.

A number of other studies have used different methods to train social skills in ASD (see Table 2). For example, Mitchell et al. (2007) used virtual environments (3-D computer representations of real environments, which users could manipulate their way around with a joystick) to train subjects at choosing a place to sit in a crowded canteen. Unfortunately, despite the huge effort put into developing the software, the evaluation study they report is small and inadequately controlled, thus rendering the effectiveness of the intervention impossible to evaluate. Similarly, Grynszpan et al. (2008) present an excellently designed piece of software targeting the comprehension of dialogue containing pragmatic subtleties – but the measure they use to test whether trained improvements have generalised is merely a different version of the training software. Studies designed in this way make it impossible for the reader to judge whether, and how far, trained improvements have generalised outside the world of the training paradigm.
19 intervention packages have been reviewed so far. It is striking that the four studies that have satisfactorily tested for distal generalisation – ie generalisation to non-overlapping but similar tasks outside the trained set - all have reported negative results (Golan & Baron-Cohen, 2006; Faja et al., 2008; Tanaka et al., 2010, Swettenham, 1996). (Golan et al. (2009) have reported positive results here, but this finding requires replication by a larger, better controlled study.) Most of the other studies included here are under-powered and inadequately designed.

Why a failure to generalise? This failure of training improvements to generalise outside the training set probably reflects the well-documented difficulties of subjects with ASD in generalising and extrapolating (Cohen, 2007; Golan & Baron-Cohen, 2006; Plaisted, 2001). An inclination to focus on small details at the expense of the big picture (Happe & Frith, 2006) combined with a reduced ability to recognize the similarities between stimuli (Cohen, 2007) might lead participants to learn to solve the trained task using particular, small-scale features of the training stimuli – features that may be absent in a ‘real-world’ environment. This is a problem that may be particularly acute, for example in the use of simplified computer-generated faces to teach emotion recognition. The fact that technology can offer reduced and simplified learning environments represents a potentially vital tool for learning, but issues of
generalisability – i.e. how to scale the learning environment up to a ‘real-world’ complexity - remain to be addressed.

Convergent evidence increasingly suggests that autism is not simply a problem of subjects lacking specific pieces of declarative knowledge (e.g. ‘not knowing’ what emotions are). For example, recent eye-tracking studies (e.g. Speer et al., 2007) have shown that, in simplified, ‘low-load’ environments (such as when viewing a still image of a single human face against a black background) subjects with ASD orient in much the way that anyone else would, but that differences only emerge in more complicated, ‘high-load’ environments (such as when viewing a movie clip in which several actors are talking and interacting simultaneously). Perhaps one failure of the approaches reviewed so far is the tacit emphasis they place on teaching individuals with ASD things they do not know, rather than helping them to use knowledge they in fact already have (Prizant et al., 2005) (see Section 2a and conclusion).

The other important point is that most of these interventions have been aimed at older children and adults with ASD (the Golan et al. (2009) study targeted 4- to 7-year-olds, and found much stronger results). Given that we are starting to understand more about abnormalities early on in the development of autism (Garon et al., 2009; Elsabbagh & Johnson, 2010) interventions targeted at earlier ages may in future be shown to be more effective, because the brain is more plastic early in development (Johnson, 2010; Dawson, 2008).
Part 2 – Other Technologies

In this section, we include a selection of different technologically mediated interventions for subjects with autism. Although we have grouped this discussion by the technology used, many of these technologies (including virtual reality and robotics) have the potential to be used in a variety of ways. Where possible we have tried to reflect this in our discussion.

Section 2a – Virtual reality

‘… I hear too much and see too much but my sense organs are ok – it’s just that inside it all sets off – words, sentences, ideas get torn apart and jumbled up…’

Sellin, 1993

Summary: Recent work suggests that one important deficit in autism is the ability to coordinate neural firing between spatially discrete brain areas (Rubenstein & Merzenich, 2003; Just et al., 2004; Wass, 2010); behaviourally, this may lead to subjects getting overwhelmed easily, and having problems with tasks that require the rapid or time-sensitive integration of information from spatially discrete areas of the brain (Just et al., 2004; Belmonte et al., 2004). Current technologies appear perfectly positioned to provide therapies that target this area – especially virtual reality, which offers the ability to integrate multi-modal information in a controlled and user-defined way, and to control the number of background distractors in an otherwise realistic training task. We review the few studies that exist here, and conclude that it is an area with untapped potential.
An increasing number of researchers (see Wass, 2010) are viewing ASD as a condition in which abnormal neural connectivity patterns (e.g. Rippon et al., 2007), possibly accompanied by excitatory/inhibitory neurotransmitter imbalances (Rubenstein & Merzenich, 2003) lead to a ‘noisier’ brain – i.e. a brain that is less able to filter out signals from noise. In a recent computational modelling paper on connectivity in autism, Cohen (2007) suggests that, as well as coming early in development, interventions should be “intensive […] and geared toward focusing their attention on forming relevant associations and on eliminating irrelevant associations.” This is a theme also found in many behavioural interventions (Myers et al., 2007) - programs such as TEACCH (Mesibov & Shea, 2010) and ABA (Howard et al., 2005) place a heavy emphasis on structure and repetition, and on a background environment that children do not find overwhelming.

Virtual reality (VR) offers countless opportunities here. The amount of background distraction can be adaptively varied depending on how well the subject is performing the task (cf Mitchell et al., 2007), in a way that is virtually impossible in a real-world environment. Also, VR offers the possibility of gradually integrating multi-modal information (proprioceptive, visual, auditory and tactile) at a pace that is defined by the user.

Sadly, the opportunities offered by VR in this area remain largely unexploited. One study designed a virtual-reality intervention that they described as a version of sensory integration therapy (SIT, Jung et al., 2006), although it is limited in scope. Furthermore, the fact that the study contains no pre- and post-tests other than the
trained set, and no untrained control group, make the success of the intervention hard
to evaluate.

Some studies (in addition to those that used VR for social skills coaching - described
in Section 1c) have used VR to train particular skills – to train children with ASD at
understanding imaginative transformation (Herrera et al., 2008), or to provide explicit
didactic training for real-life situations (e.g. shopping (Lanyi & Tillinger, 2004) or
crossing the road (Strickland et al., 2007)). The capacity of virtual reality
environments to provide gradual transitions between simplified and more complex
environments offers a significant advantage here, relative to the interventions we
reviewed in section 1. This is an area with considerable potential for future
development.

Section 2b – Gaze-contingent paradigms

Summary: Gaze-contingent paradigms use eye-tracking to create environments in
which objects react contingently depending on where the subject is looking. We
discuss two such interventions. The first, developed by Trepagnier and colleagues, is
a gaze-contingent environment that has been developed to encourage people with
autism to engage in social behaviors such as meeting eye gaze and following social
cues. The second, by Wass and colleagues, is a battery of gaze-contingent computer
games for infants that train early executive (ie ‘top-down’) attentional control.

In comparison to other methods of computerised attention training (eg Rueda et al.,
2005) that exploit the more traditional ‘point-and-click’ interface, gaze-contingency
offers the opportunity for more immersive, immediate and sensitive cognitive training environments. As oculomotor control develops early (Johnson, 1990) and is a highly robust aspect of human motor control, they also offer a method that can be applied to a wider range of subjects, including the very young and the heavily disabled.

Trepagnier et al. (2004) describe the development of a gaze-contingent environment they developed to reward subjects with autism for looking other people in the eyes and for engaging in dyadic and triadic joint attention. [There are some papers here that we can’t got hold of that we have requested from Reviewer 1 if (s)he has a copy, otherwise we will have to re-write – see Note to reviewers.]

Wass et al. (submitted) have adopted a different approach. Using a commercially available Tobii eyetracker they developed a battery of computerised attention and working memory training games suitable for infants. These were aimed at improving the early development of executive attentional control (ie the control we have over what we attend to and what we ignore). In an active-controlled, medium-scale trial using typically developing 11-month-old infants (N=42, 5 lab visits per participant), they found that four training sessions (avg total training time: 77 mins) led to significant improvements at a number of measures of the top-down control of attention (Wass et al., submitted). It is possible (although unproven) that early abnormalities in aspects of executive attentional control may play a causal role in impairing subsequent learning in social situations (Zwaigenbaum et al., 2005; Landry & Bryson, 2004; Holmboe et al., 2010), so one potential application of this work is for infants at risk of developing autism.

Section 2c – ‘Robot friends’.
Summary: Robotics presents another promising medium for autism interventions. As companions, they offer the opportunity for highly predictable, repetitious interactions. Some small-scale studies have demonstrated that severely impaired children with autism will learn to attend to, and imitate, a robot companion. This offers a potential stepping-stone toward human-to-human interaction. There is a considerable amount of work ongoing in this field.

[insert figures 4 and 5 about here]

Another area of ASD research that shows considerable potential growth is the use of ‘robot friends’ for children with autism, an idea that dates back to 1976 (Weir & Emanuel, 1976). Interaction with robots tends to be predictable and repetitious; these are the sorts of interactions that some subjects with autism prefer (Thakkar et al., 2008). The early studies that have been conducted in this field tend to take the form of imitation interactions – i.e. robots that copy the movements of their ‘human friend’. This form of interaction is intended as a stepping-stone to turn-taking, and other aspects of human-to-human interaction.

Early versions of the AuRoRA project (Dautenhahn, 2000 - [www.aurora-project.com](http://www.aurora-project.com)) used a robot with a rectangular body and four wheels, exploiting the preference of many children with ASD for mechanical objects (cf the Transporters series – Golan et al., 2009). More recent versions (Billard et al., 2007) feature Robota, a humanoid doll that uses infra-red signals to mirror human body movements, is able to learn simple motor sequences, and has bi-directional communication capability (i.e. can invite imitation and imitate herself).
Duquette et al. (2008) provide one well-designed (although small N) experimental study on the effect of exposure to a robot (vs a human experimenter) on imitation rates (e.g. imitation of facial expression) and shared attention. For a matched period of 22 sessions, two 5-year-old children diagnosed with ‘low-functioning’ autism engaged in structured play with a robot that involved imitation of facial expressions, body movements and familiar actions, and two children took part in an identical play session with a human experimenter. Post-training, they found that the robot-trained children engaged in more shared attention and imitation of facial expression with their ‘robot friend’ than the human-trained children, although other measures such as imitation of body movements were not improved (Duquette et al., 2008). Other examples of robots for ASD include Keepon (Kozima et al., 2007) and work from Feil-Seifer & Mataric, (2008).

One further study comes from Tartaro & Cassell (2008) who presented six high-functioning ASD children aged 7-11 with a virtual peer (i.e. computer-animated children, presented life-size on a screen). This virtual peer was controlled using a ‘Wizard of Oz’ methodology (a human operator hid behind a screen and controlled how the virtual peer reacted). Video-coded post hoc analysis of the interactions with the virtual (compared to a ‘real-life’ peer) showed that, by the end of the session, children had learnt to take turns in joint story-telling faster with the virtual peer than with the ‘real’ peer.

*Section 2d* – *Digital Play Environments.*
Other projects use different media to explore interactive play in ASD. The Reactive Colours project (Keay-Bright, 2006, http://www.reactickles.org/gallery) is a digital play environment in which children with autism stand in front of a screen and explore different sorts of touch interface. Touching the screen evokes different combinations of auditory and visual cues – in one activity, dragging your finger across the screen creates a trail of bubbles and an ethereal sound; in another, hitting the screen results in cymbal crashes and huge splodges of colour.

The appeal of such a play environment are intuitively understandable (Keay-Bright, 2006). It is a highly predictable, beautifully realized environment that rewards the kinds of repetitive, perfectly contingent interactions that some people with autism have been reported to prefer (Escalona et al., 2002). Although there is no direct training component, such endeavors are important as assistive technologies (for a wider review of this area, see Mirenda, 2008).

Section 2e – ‘Emotional hearing aids.’

A number of technologies exist aimed at assisting and augmenting communication in individuals with autism (for a more thorough review, see Mirenda, 2008). Probably the most thoroughly researched work in this area comes from the field of affective computing (e.g. Picard, 1997; Picard et al., 2004; el Kaliouby & Robinson, 2004; el Kaliouby et al., 2006; see also www.affectiva.com) - ie teaching computers to
understand and sense human emotions (Picard, 1997; el Kaliouby et al., 2006; Picard, 2009).

One use of this technology in this area (Poh et al., 2010) involves measuring galvanic skin response (also known as electrodermal activity), which has for more than a century (e.g. Jung, 1906) been used as an index of emotional arousal. Sensors can be used in two ways – firstly, worn by an (e.g. non-linguistic) person with ASD, and used by parents’/caregivers to track that person’s arousal level (thus potentially anticipating and preventing the distressing ‘meltdown’ episodes that are so common to the caregivers of people with ASD) (Picard, 2009). Alternatively, the devices can provide a wearer with feedback on their own arousal levels (el Kaliouby et al., 2006).

Other work here involves cameras/webcams than can be trained to recognise emotion based on a combination of visual (facial) cues and top-down predictive emotional models. The wearable camera can either be pointed outwards, allowing a user to receive the computerized predictions about the emotional states of the person he or she is talking to, or pointed inwards (el Kaliouby et al., 2006) thus allowing the wearer to receive automatized predictions of what they are communicating about their own mental states.

This remains, however, a field in its infancy. We are some way from the point where computers are as accurate and reliable at recognising emotions as typical adults are (el Kaliouby et al., 2006; Picard, 2009). And to our knowledge, no studies yet exist in this area that systematically evaluate the effectiveness of these applications.
Conclusions

Despite several promising beginnings and considerable interest in this area (Pennington, 2010; Di Gennaro Reed et al., 2011), we are still some way from clinically verified technology-based interventions for autism.

Better inter-disciplinary cooperation will help us toward achieving this goal. To this end, we end by proposing a number of specific points:

For technology designers:

i) early, intensive interventions are preferable. A number of clinicians (Dawson, 2008; Myers et al., 2007) have concluded that early, intensive interventions are most likely to be successful because neural and behavioural plasticity is greater earlier in development (Johnson, 2010; Karmiloff-Smith, 1992; Heckman, 2006). Interventions targeting at-risk infants and very young toddlers should be the target of future work.

ii) clinical trials are essential to assessing the validity of a new intervention. Well-designed intervention trials (with a randomised, matched clinical control group including ‘technology placebos’ (Kaptchuk et al., 2000)), adequate sample sizes and a large enough training period) are expensive and time-consuming to run, but they are essential to addressing the utility of interventions. Without this evidence base, it is impossible to assess potential clinical applications and to select promising areas for follow-up work.
iii) autism is a complex condition that consists of more than a few, core cognitive deficits. A number of interventions have been aimed directly at remediating particular aspects of the autistic cognitive phenotype (eg computer programs aimed at teaching individuals with autism to recognise emotions). Although inconclusive, the evidence reviewed here suggests these approaches have been largely unsuccessful at demonstrating transfer (see sections 1a, 1b, 1c). Instead we have suggested that more imaginative approaches may be required, targeted at applying (rather than acquiring) knowledge and based on a better empathetic understanding of how the world ‘feels’ to people who have autism. One particular goal here may be that of helping people to cope with the experience of a phenomenologically ‘noisier’ brain that some individuals with autism experience (see Sections 2a and 2b).

For clinicians:

i) technologically mediated interventions can be high-intensity at relatively low cost. Many technology-mediated interventions are home-deliverable and therefore offer the opportunity for intensive and prolonged exposure (up to thousands of hours, potentially supervised only by parents) that is financially unachievable for clinician-mediated interventions. To this end, any use of technology that can be shown usefully to help strengthen even a single subcomponent of the autistic behavioural phenotype should be investigated, perhaps as a part of a wider intervention battery.

ii) different uses of technology are suitable for different learning goals and different populations. Language-based, point-and-click computerised training programs are clearly only suitable for less severely impaired individuals on the autism spectrum. Other technologies have been shown to be suitable for other populations (eg simple
imitative robots for more severely impaired individuals, and gaze-contingent environments for infants).

**iii) technology offers the potential for repetitive, restrictive, adaptive and user-controlled training paradigms.** Many clinician-mediated interventions emphasise the utility of restrictive, repetitive environments (Howard et al., 2005; Dawson et al., 2010). Immersive technologies such as virtual reality and gaze-contingent interfaces offer an opportunity to create exactly such restricted environments. Furthermore, they offer the opportunity to manipulate the parameters within an environment (eg presence of multi-sensory cues, complexity of setting etc) in a more subtly adaptive, repetitive and more finely user-controlled manner that is possible in more traditional clinician-mediated interventions. Thus, they provide potentially exciting tools for designing expansions of the clinical interventions currently in use.

**iv) simplifying learning environments is both a strength and a weakness.** If carefully designed, it is possible (albeit unproven (although see Duquette et al., 2008)) that technologically mediated interactions may be able to act as a stepping stone to human-to-human interaction - for example, severely impaired children with autism might learn (over hundreds or thousands of hours of exposure) to partake in certain highly predictable interactions with robots, skills that they are then able (in certain settings) to apply to human-to-human interactions (Duquette et al., 2008). However, we have also reviewed considerable evidence suggesting that simplifying a learning environment leads to problems of generalisability (ie applying that knowledge in ‘real-world’ settings). In a case such as recognising emotions from a computer-
generated vs real human face, it is possible that learnt skill acquisition will never be able to ‘bridge the gap’ between a computerized and ‘real-world’ environment.

Most of the interventions reviewed here (Section 1) have provided domain-specific didactic training aimed at improving particular ‘surface skills’ (eg face recognition). Instead, we have speculated that a more fruitful approach for future work might be that of aiming to provide less specific, more ‘domain-general’ cognitive training targeted at improving key learning skills (such as attention and memory).

Close inter-disciplinary cooperation is vital for future progress in this field. We hope that this article represents a step in that direction.
References:


American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders (4th ed.).


Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, T., Smith, T., Menzies, R., Foster, M.E., Alcorn, A., Wass, S., Bernadini, S., Avramidis, K., Keay-Bright, W.,


Footnotes

i “ich staune wessen geduld aber grosser ist deine oder die von einem computer die vom computer ist doch unauffrichtig ganz fast ehrlich dagegen die von dir tatsache ist ich konnte mit fast fuenf jahren auch schon schreiben und sogar rechnen aber es hat niemand gemerkt weil ich so chaotisch war aber das war ich einfach aus angst vor den menschen gerade weil ich unfahig war zu reden fiel mir das lesen so leicht”.
Translation by S.Wass.

ii http://slurl.com/secondlife/Porcupine/37/185/105/

iii http://www.autismandcomputing.org.uk/

iv http://www.autistics.org/

v http://www.youtube.com/watch?v=JnylM1hl2jc


vii http://www.stereo3d.com/hmd.htm#chart

viii http://www.didel.com

ix “…ich kann ein wenig zu viel hoeren and zu viel sehen aber die sinnesorgane sind o.k. einfach innen geht ein durcheinander leider los woerter saetze ideen werden so auseinandergerissen und zerissen die einfachsten dinge werden aus dem zusammenhang der wichtigen wirklichen einzelnen anderen aussenwelt gerissen…’.
Translation by S.Wass.
Figure 1: A screenshot the training software presented in Golan et al., 2006
Figure 2: A screenshot from the educational DVD presented in Golan et al., 2009

Figure 3: A screenshot from the software presented in Mitchell et al., 2007
Figure 4: The robot used for training in Duquette et al., 2008

Figure 5: The robot in its two different appearances – pretty doll (left) and plain (right). Two children playing a limb-to-limb imitation game involving legs, head and arms. In this work, Robota’s mirroring the motion is working in a ‘puppeteering mode’, i.e. as a remote controlled robot where the experimenter picks up subtle movements of the children and responds with contingent movements of Robota’s limbs and head.

Figure 5: Photos from Billard et al., 2007
Figure 6: A child playing with Reactive Colours (Keay-Bright, 2007)